Itron White Paper

Energy Forecasting

Defining Normal Weather for Energy and Peak Normalization

J. Stuart McMenamin, Ph.D

Managing Director, Itron Forecasting

Introduction

Weather normalization is a process that adjusts actual energy or peak outcomes to what would have happened under normal weather conditions. Normal weather is also used to develop the base forecast of future energy use. A variety of methods have evolved to define normal weather, but there are some fundamental principals that must be followed. The purpose of this paper is to define these fundamentals.

In financial forecasting for energy delivery companies, models are typically estimated with monthly billing data, which represents customer usage over a set of billing cycles. The weather variables used as explanatory variables in these models are usually defined to represent the weather that occurred over these cycles. The models (or coefficients from these models) are then used to develop a variety of estimates, including:

- Estimates of historical energy usage for the cycles with normal weather
- Estimates of historical calendar month energy usage with actual weather
- Estimates of historical calendar month energy usage with normal weather
- Estimates of historical unbilled energy with actual weather
- Estimates of historical peak demand with normal peak producing weather
- Forecasts of calendar month energy usage with normal weather
- Forecasts of peak demand with normal peak producing weather
- Forecasts of peak demand under extreme design conditions

The recommendations presented here are based on the fact that there is a nonlinear relationship between weather and energy usage. Because of the nonlinearities, it is important that actual and normal weather variables are constructed in a way that avoids aggregation bias.

Actual Weather Data is Chaotic

Actual weather data varies from year to year. But the one thing that is constant is that it is chaotic every year. There is a large annual cycle, but within this cycle, there are bursts of warm and cool weather. Peak loads are driven by strings of extremely hot days in summer and extremely cold days in winter. It is impossible to predict exactly when these patterns will occur, but they do occur every year. In some years, the extremes are more extreme than in other years. In some years, the extremes occur on weekends, and do not produce peak energy use.

Figure 1 provides an example of a single year of weather data. The plot shows maximum, minimum, and average temperatures for each day of the year. There are two popular ways to utilize this type of daily weather data to define normal weather patterns. One is appropriate for modeling normal peaks, and one is appropriate for energy normalization. The two methods are:

- Rank and Average
- Average by Date

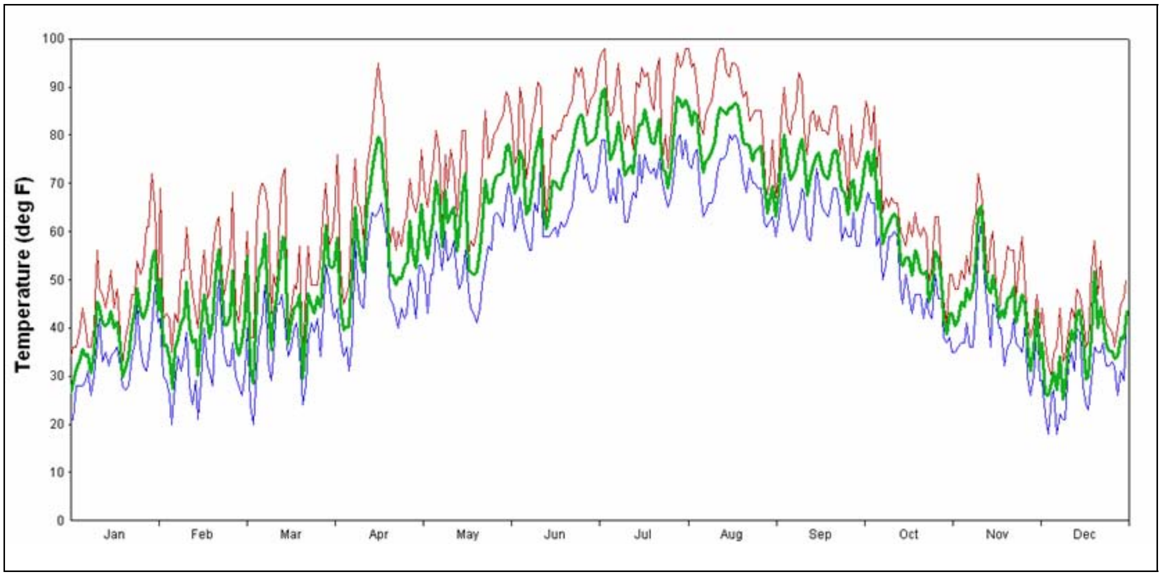

Figure 1: *Example of actual weather data*

Rank and Average Method

The rank and average method is used to develop a typical weather pattern that has normal extremes and normal average temperatures in each season or month. The method starts by taking each year of data and ranking the days from hottest to coldest. Figure 2 shows an example of weather data that is ranked for the year by average temperature. In this chart, the average values are ranked, the maximum values are ranked independently, and the minimum values are ranked independently. For each day, the average values are computed as the average over the 24 hourly values.

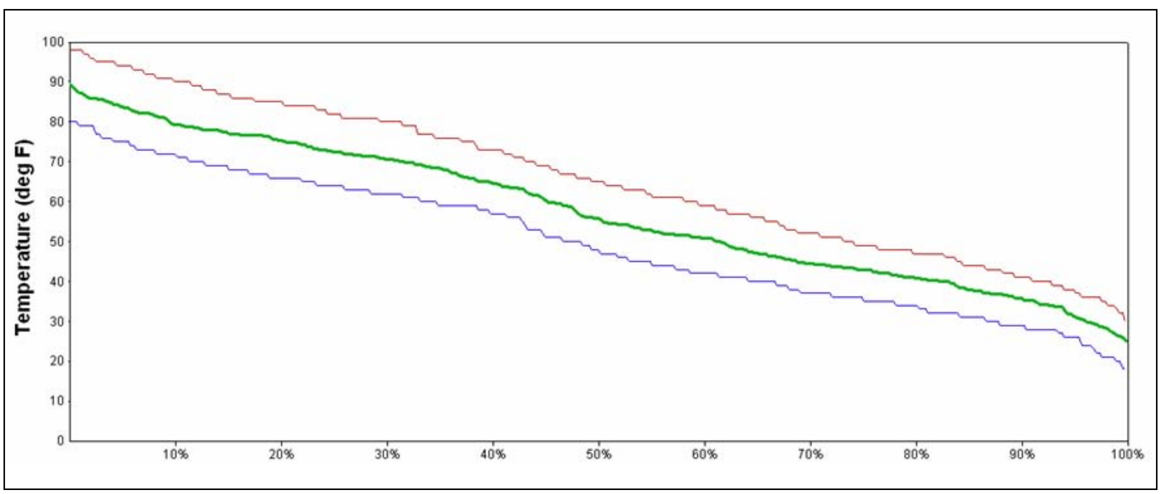

Figure 2: *Example of one year of data ranked by average temperature*

One approach ranks the data within each historical month and then averages across days in the month. For example, for July all of the hottest days in July are averaged, all of the second hottest days in July are averaged, and so on. Although this produces relatively extreme weather in each of the summer months and each of the winter months, it will typically understate extreme weather for the typical summer season and the typical winter season.

A second approach ranks the data within each historical season and then averages across days in the season. For example, the hottest summer days are averaged across years, the second hottest summer days are averaged across years, and so on. After this seasonal averaging is completed, it is necessary to assign the resulting values to months within the season.

Figure 3 provides an example of the results when temperatures are ranked within each month and the ranked values are then averaged across 30 years.

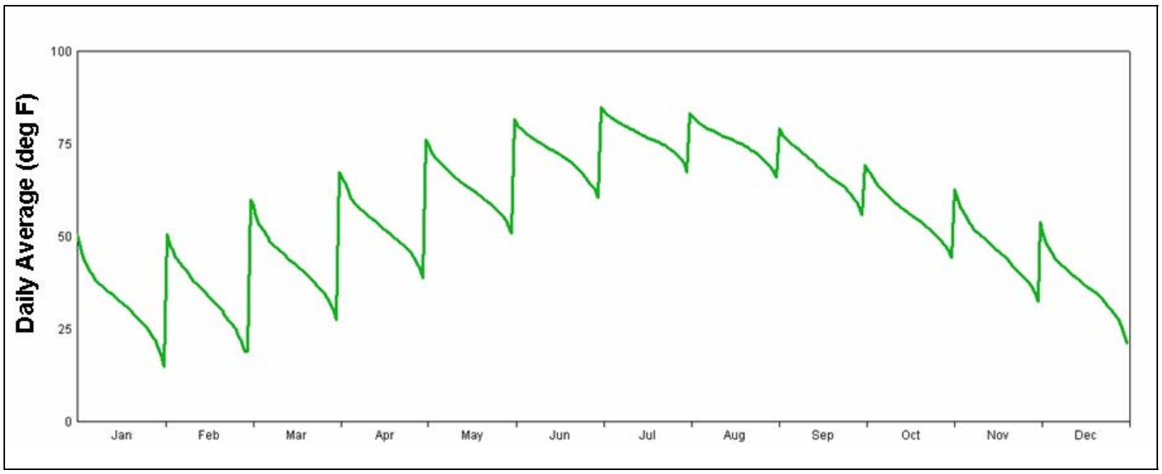

Figure 3: *Example of Rank and Average Temperatures*

Once weather data has been ranked and averaged, the last step is to assign the results to specific days in each year. There are several ways to do this assignment. For example, the extreme temperatures can be assigned to the dates that have the most extreme temperatures in a particular year. Or they can be assigned based on the dates that have the highest average temperature when temperatures are averaged across years by date. In either case, the pattern can be rotated so that the extreme temperatures (hot and cold) do not fall on weekends or near holidays.

Regardless of the assignment rule, at the end of the process the result will be a chaotic weather pattern, resembling the data presented in Figure 1. These types of results can be used for the following types of analysis:

- Building energy simulation models that require weather for a typical year.
- Hourly simulation of energy use by rate class for a typical year.
- Hourly simulation of system loads for a typical year.

There are alternative weather patterns available for these purposes, such as the typical meteorological year data (TMY and TMY2) developed by government agencies. These data provide a chaotic weather pattern developed by concatenating selected months of historical data into a single annual series. While these data represent average weather conditions fairly well, they do not necessarily represent peak producing weather, and may not be appropriate for simulations of monthly or annual peaks.

Weather data can be ranked within a year, season, or month and averaged across years to construct a chaotic weather pattern that embodies typical extreme values as well as typical average values. The results must be assigned to specific days, and there

is no theoretically correct way to do this assignment. This type of data is appropriate for simulating 8,760 hourly patterns used for building simulation models or hourly statistical models.

Average by Date

The second approach that is frequently used is to average weather values across years by date. For example, in the historical period the temperature statistics for all January 1 values are averaged, the temperature statistics for all January 2 values are averaged, and so on through to December 31.

Figure 4 shows an example of the results of this process. In the figure, the top line shows the average of temperature values and the bottom lines show averages of cooling degree day (CDD) and heating degree day (CDD) values based on a 65 degree trigger temperature.

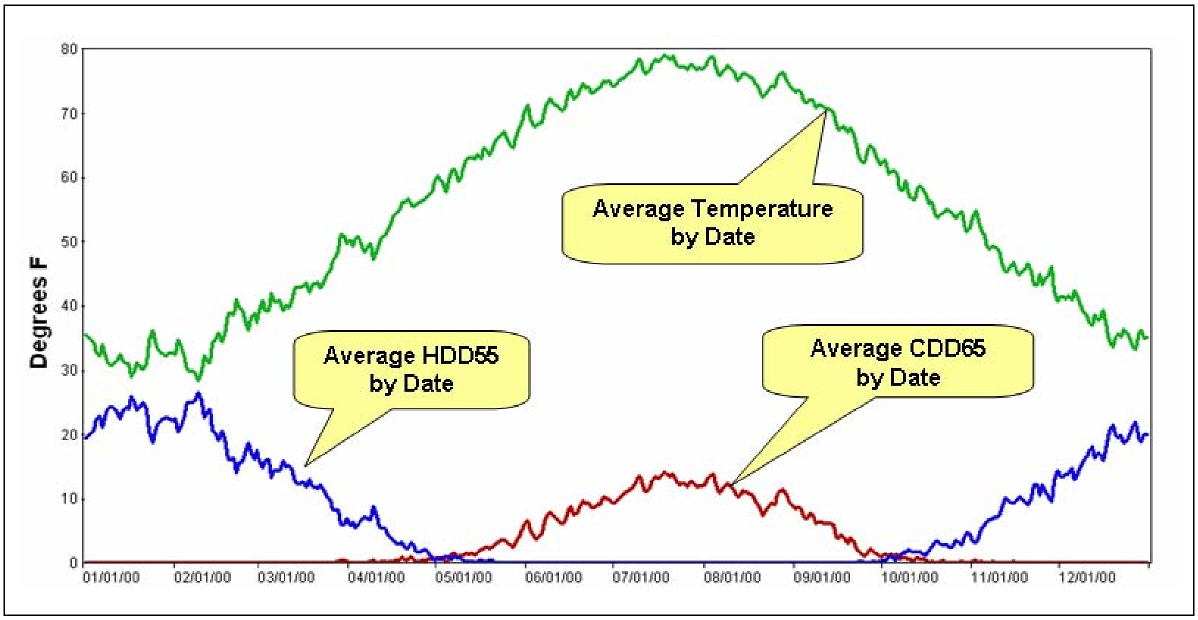

Figure 4: *Example of Average by Data over 30 Years*

Although these values are averaged over 30 years, there remain some irregularities, based on the specific 30 year time span that is chosen. To remove these irregularities, centered moving averages can be used. Figure 5 shows CDD and HDD values using a 30 day centered moving average for a variety of trigger temperatures.

The smoothed CDD and HDD averages by date are appropriate for use in energy normalization calculations, such as:

- Computing normal energy for billing cycles
- Computing normal energy for calendar months
- Computing normal energy for unbilled days
- Energy variance by calendar month

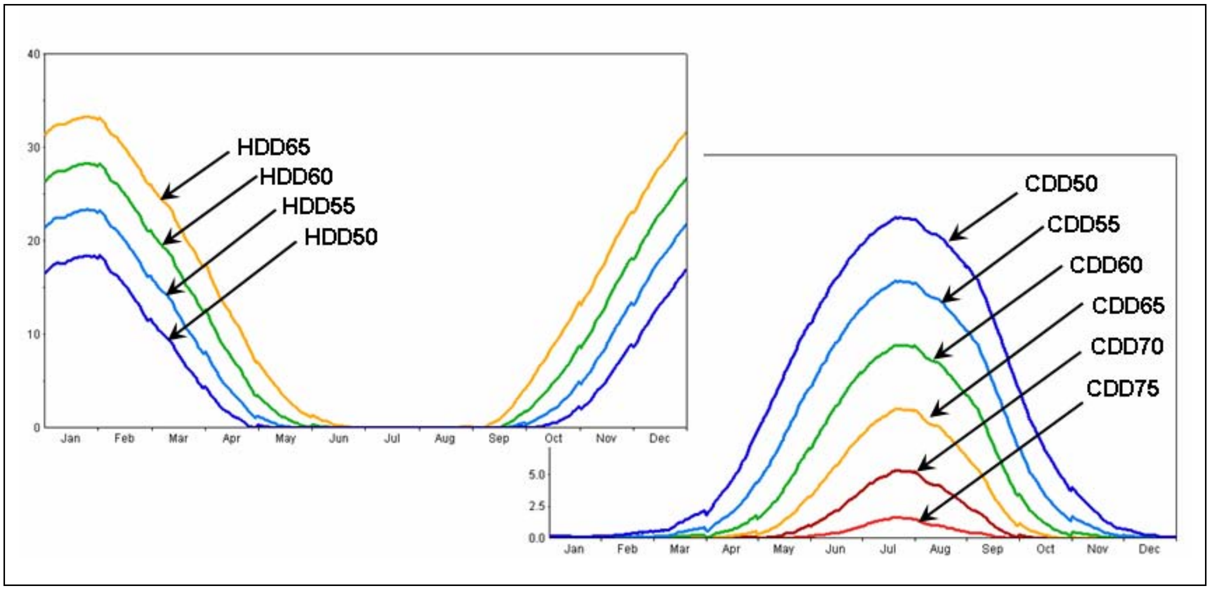

Figure 5: *Example of CDD and HDD – Smoothed 30 Year Average by Date*

As will be shown below, HDD and CDD values can be averaged by date, and these values will be appropriate for energy forecasting and energy normalization. NOAA provides daily normals computed in this way. While these values are appropriate for energy normalization, they are not intended to be used for peak normalization. The principal is:

Weather data (HDD and CDD) can be ranked by date across years. The results can then be smoothed using centered moving averages. The results are appropriate for energy normalization and energy forecasting.

Nonlinear Relationship between Energy and Weather

It is well understood that the relationship between temperature and energy use is nonlinear. As shown in Figure 6, electricity usage typically has a balance point (about 60 degrees in the figure). To the right of the balance point, as temperatures rise, energy use begins to rise. The early degrees above 60 are relatively low powered, and the sensitivity increases to the high powered degrees above 75.

To the left of the balance point, as temperature drops, energy use also begins to rise. Again, the early degrees are relatively low powered, and the later (colder) degrees below 50 are relatively high powered.

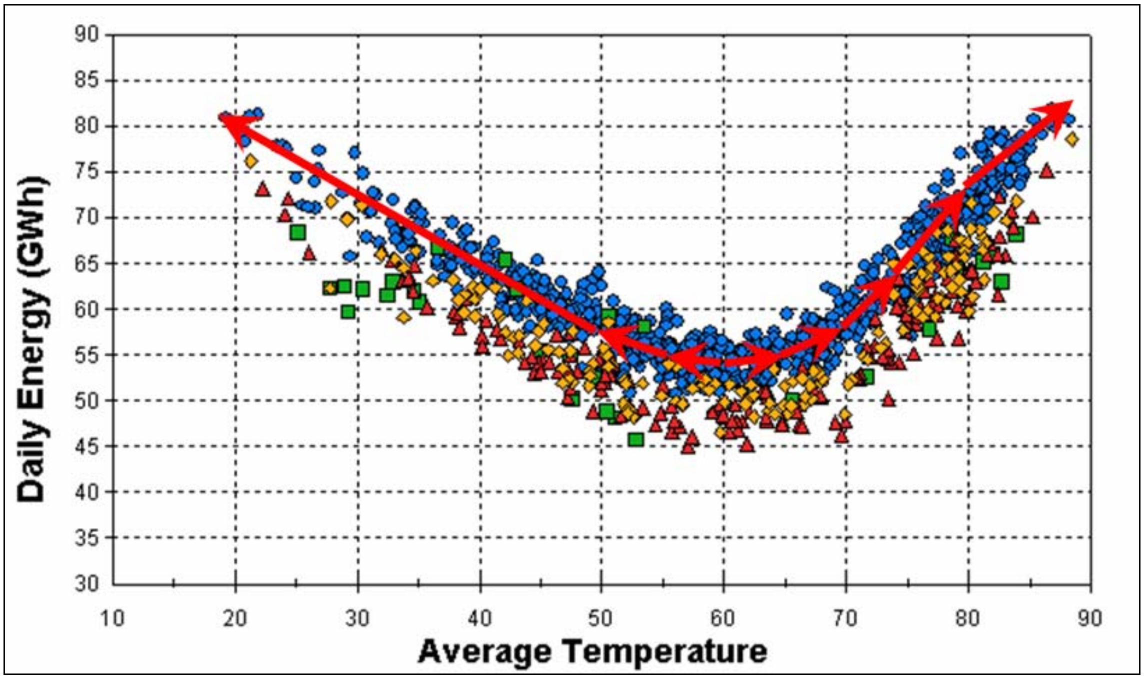

Figure 6: *Example of Relationship between Energy and Temperature*

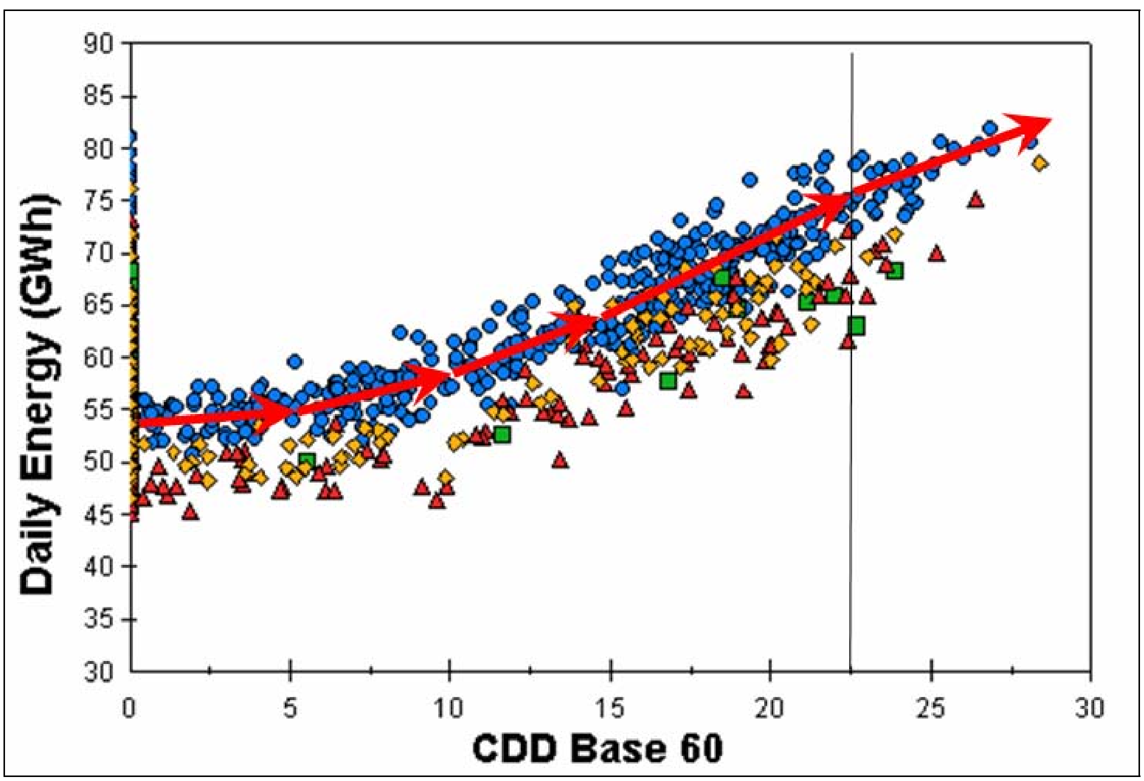

Figure 7: *Example of Relationship between Energy and CDD60*

Focusing on the warm side, Figure 7 shows the relationship between daily energy (on the Y axis) and cooling degree days (CDD) with a base of 60. This is the way a regression sees the data with cooling degree days as the explanatory variables.

The presence of this strong nonlinear relationship has implications for the construction of weather variables because of the potential of aggregation bias. This type of bias occurs whenever there is a convex or concave function. In the case of energy usage and temperature, the relationship is convex. For convex functions, the average of predicted values for two X inputs always lies above the predicted value at the average of the X inputs. There are two operations that this impacts.

- Averaging across stations to define weather zone conditions.
- Averaging across years to define normal weather.

An example of averaging across stations is provided in Figure 8. In this example, there are two regions and one weather station for each region. Assume that half of the load is in each region. In the first region, the average temperature for the day is 60 degrees and the corresponding load is point A. In the second region, the average temperature for the day is 80 degrees and the corresponding load is point B. If we weight the stations together, the average temperature is 70 degrees, and the predicted average load for the two regions at this average temperature is the point C. In fact, this grossly underestimates the true average load across regions, which is the point D. This is a classic example of aggregation bias. If we average temperatures across the stations, and then compute CDD values from this average, we will consistently understate the total load for the two regions.

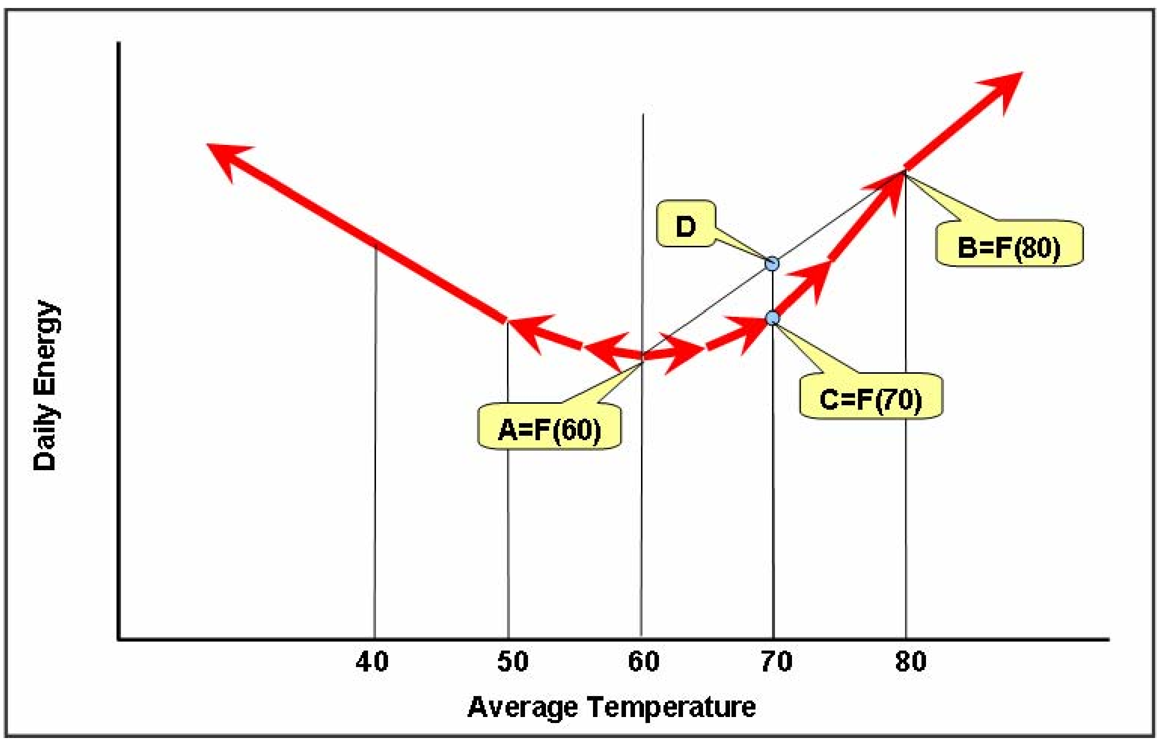

Figure 8: *Example of Aggregation Bias – Averaging across Stations*

The second example is similar, but involves aggregation across years. This example would apply in the swing months (spring and fall), when weather can be hot or cold. Suppose the data are for May 15. In the example, there are two years. In the first, May 15 is a warm day, with average temperature of 70 (point A). In the second year, May 15 is a cool day, with average temperature of 50 (point B). If we average the temperatures across the two years and then compute HDD and CDD variables, we get an average temperature of 60 and a load with no heating and no cooling (point C). The appropriate average load for the two years is point D, which has some cooling load from point A and some heating load from point B.

Our interest in weather normalization is not in the average temperature. It is in the average load. Ideally, we would like to simulate past weather patterns through a load model and average the resulting simulated loads. The center of the distribution of simulated loads is the normal load. This can be closely approximated by computing HDD variables for each year, computing CDD variables for each year, averaging these variables and plugging them into energy models.

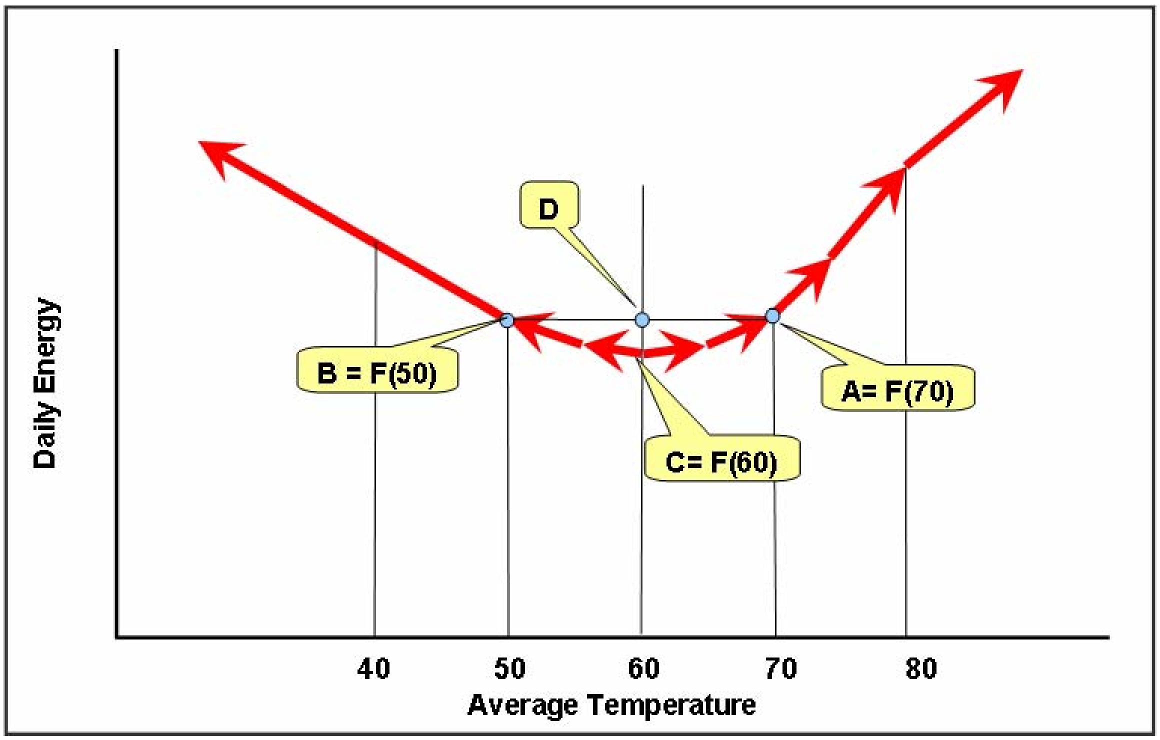

Figure 9: *Example of Aggregation Bias – Averaging Across Years*

The calculation of normal weather corresponding to this example is shown at the right. If we first average the May 15 values across the two years, we get an average temperature of 60 degrees. At this average, there are no HDD or CDD measured from a base of 60 or a base of 65. If we first compute the HDD and CDD values for each year and then average, we get HDD and CDD values at both 60 and 65 degrees.

	Year 1	Year 2	Average
Daily Avg	70.0	50.0	60.0
CDD 60	10.0	0.0	5.0
CDD 65	5.0	0.0	2.5
HDD 60	0.0	10.0	5.0
HDD 65	0.0	5.0	2.5

The point is that if we average temperatures across years and then compute CDD and HDD values, we will get a downward bias in normalized energy use. Using this average there is no heating or cooling despite the fact that there is either heating or cooling in both of the historical years. This will give a normalized energy value that is too low.

The correct process is to compute HDD and CDD values first and then average these values across years. There is a general principal here:

In the presence of nonlinear relationships, apply all nonlinear operations (such as computing HDD and CDD values) first and then average across years. When using models with multiple HDD and CDD variables, this will closely approximate the result that would occur by simulating energy use with each weather pattern and averaging the energy results.

This is not just a hypothetical possibility. Figure 10 shows the results of aggregation bias when temperatures are averaged over 30 years, and HDD and CDD values are then computed from the averages. As shown, in this example, CDD and HDD values are significantly understated in the swing months (April, May, October, and November).

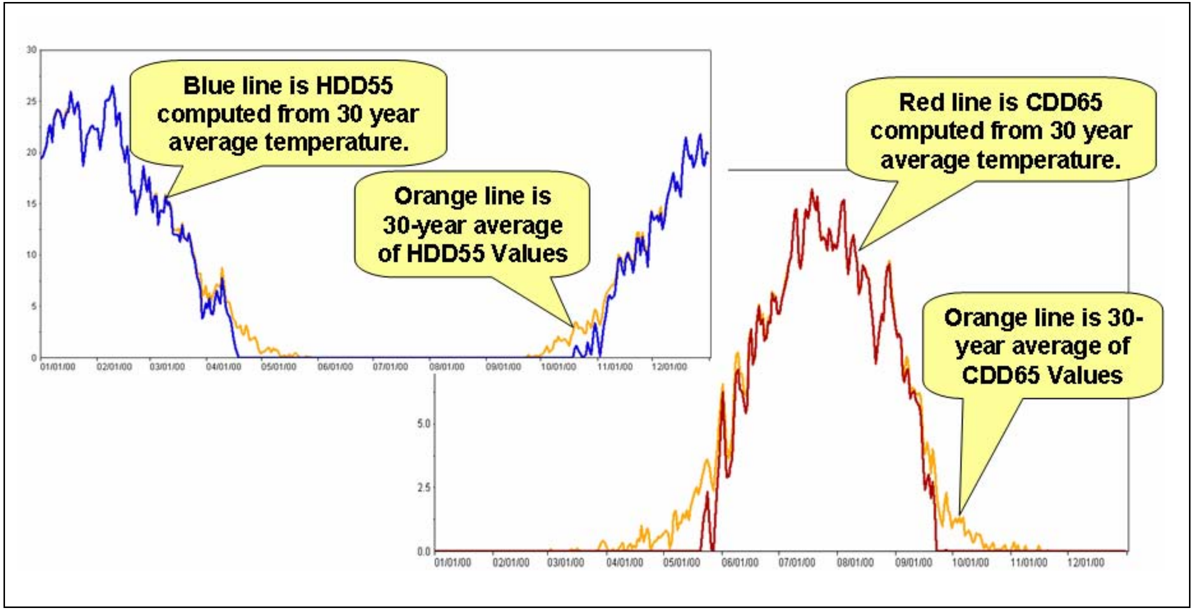

Figure 10: *Example of Aggregation Bias with HDD65 and CDD65*

To reiterate, if we average temperatures across years first and then compute HDD and CDD, there will be a downward bias in normalized energy use. Similarly, if we estimate with actual daily weather and forecast with normal weather, we will under-forecast the true expected value.

Calculating Average Temperature

For many years, the industry standard has been to compute average temperature for a day as the average of the maximum and minimum values for that day. This process was natural in past decades when many weather stations only reported max and min values. On days with typical weather patterns, the average of the max and min values will closely approximate the average of the hourly values. There is an obvious problem however, on days when typical weather patterns are disrupted by weather fronts, afternoon thunderstorms, and similar disturbances.

An example is provided in Figure 11. In the figure, hourly weather values are shown for the week of July 8. On the 8th, the weather pattern is typical, with the warmest temperatures in the late afternoon. On the 9th, temperatures drop in the afternoon and on the 10th and 11th, temperatures drop dramatically.

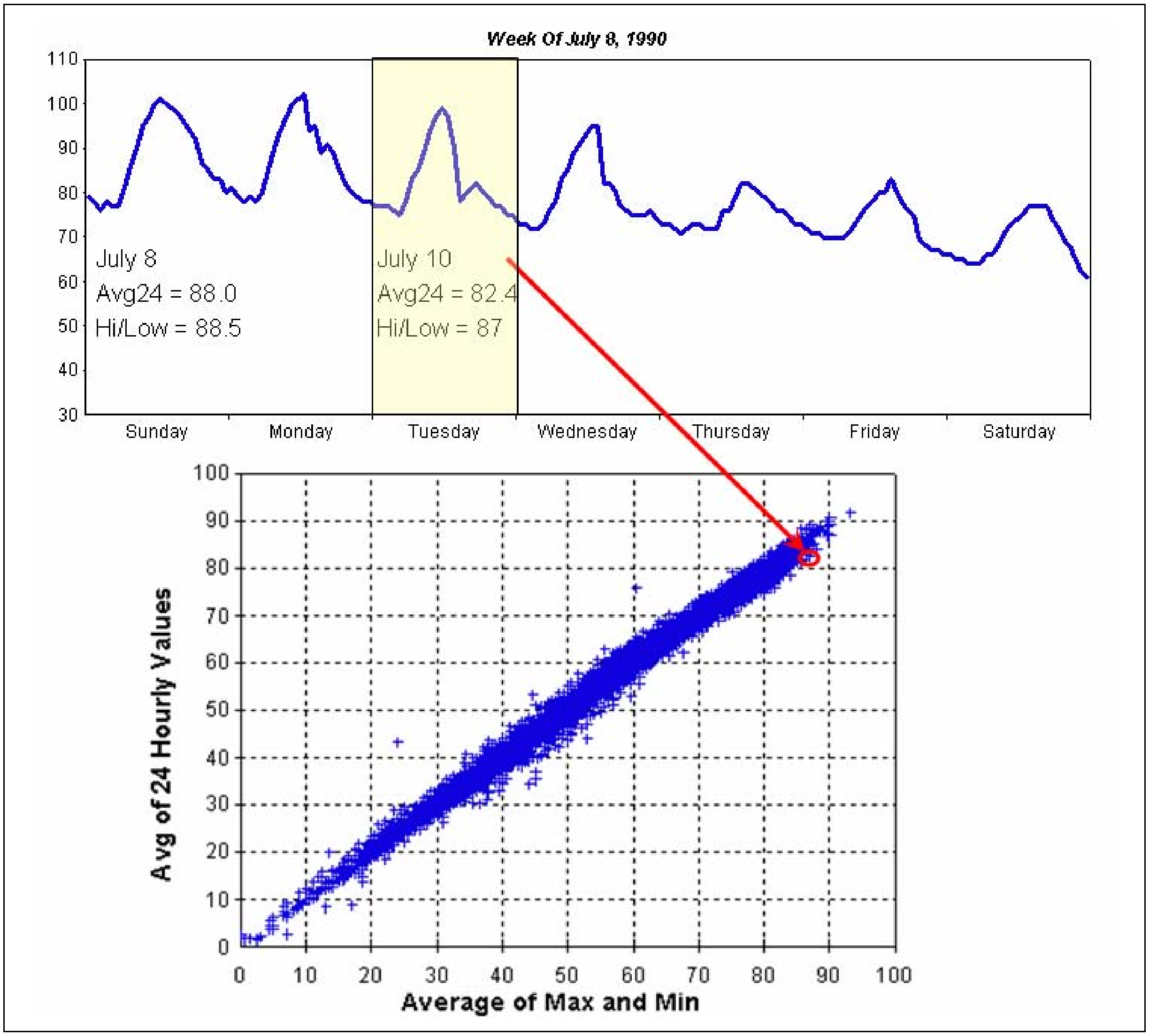

Figure 11: *Example of High+Low/2 versus Average of Hourly Values*

As shown, the average temperature on the 8th computed as average of the high and low values is close to the average of the 24 hourly values (about 88 degrees in both cases). This is not the case on the 10th, where the average over 24 hours (82.4 degrees) is almost 5 degrees below the average of the high and low values (87 degrees).

Ultimately, the data will tell us which approach is the most powerful predictor of energy use. In every test performed to date, the average of 24 hourly values performs better than the average of the high and low values. As a result, we conclude:

Given the capability of modern weather measurement equipment to gather hourly data, it is recommended that HDD and CDD values for each day be computed from the average of the 24 hourly values.

Peak Producing Weather

As discussed above, one way to simulate monthly and annual peaks is to create a chaotic weather pattern with typical extremes in each month. This weather pattern can then be used to simulate daily and hourly energy over a year. For estimating peaks, however, it is usually better to deal only with the peak data. This is done using the following procedure.

- Identify the system peak days in each month or season.
- Store the weather data for the peak day and prior days.
- Average peak day and prior day weather across years.

The result of these steps is a database of peak producing weather. These values can be averaged to get normal peak producing weather, or the extreme values can be analyzed to define design conditions (e.g., 1 in 20).

For summer and winter months, the order of operations is not critical, since all the summer peak days will be very warm and all the winter peak days will be very cold. Under these circumstances, the danger of aggregation bias is slight because all the data is in a relatively linear part of the weather relationship. In the swing months, it is probably best to identify each month as either a warm peak month or cold peak month, and average only warm days or cold days respectively. Again, it is best to compute CDD and HDD values first, and then average these values across years.

Peak producing weather is not necessarily the weather on the hottest day in summer or the coldest day in winter. Prior day weather matters and the timing of weekend days and holidays matter. This point is illustrated in Figure 12, which shows the daily weather data (high, low, and average temperatures) for each day in July and November of a selected year. The circled values are the actual system peak days in these months.

- In July, the warmest weather occurs on 4th, with hot weather on the preceding days as well. The actual system peak occurs on July 23rd, which was a Tuesday in this year.
- In November, the coldest weather occurs on November 28. The actual system peak occurs on November 18th, a Monday following a relatively cold weekend.

As an alternative, it is possible to average extreme weather across years. This will typically result in more extreme weather values than those computed over actual peak days. This follows from the fact that the most extreme weather conditions will occur on a weekend or holiday about 30% of the time and will not produce peak loads because of attenuated business loads on these days. The resulting principal is:

To define normal peak producing weather, identify actual system peak days and the weather on these days and on preceding days. Use this weather to define normal peak producing weather by month or by season.

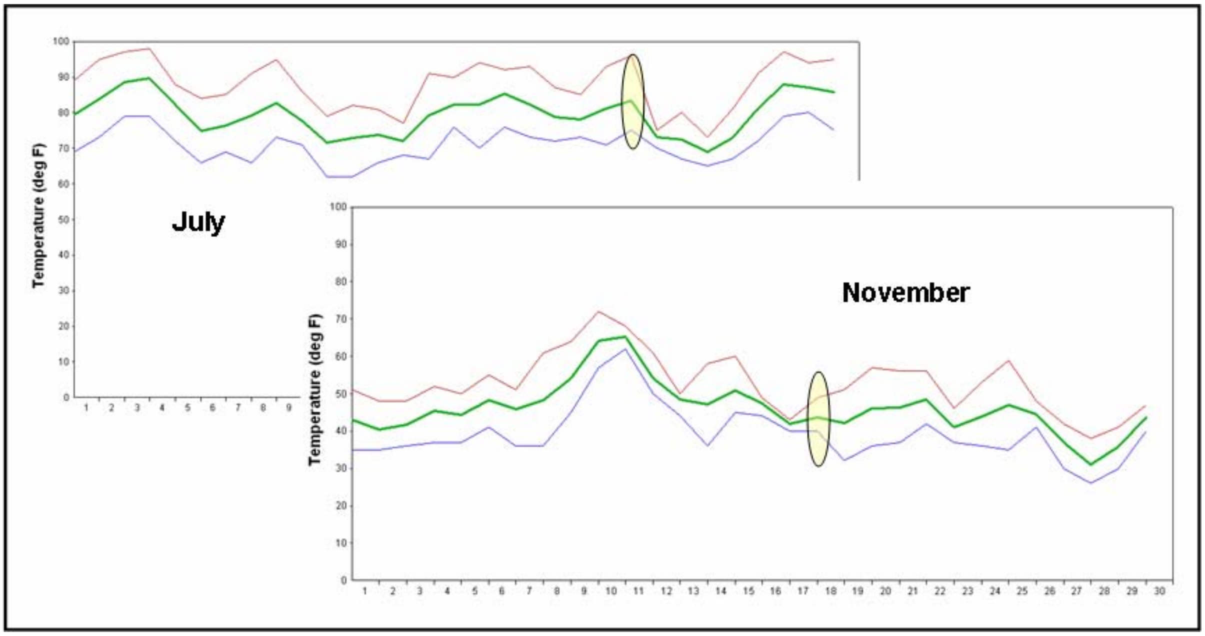

Figure 12: *Examples of Peak Day Weather*

Conclusion

The above principals suggest some best practices. First, for estimating monthly energy models with retail data, the dependent variable will usually be sales on a billing cycle basis, and the right hand weather variables will typically be the HDD and CDD values that occurred over the days in the cycles.

For normalizing historical monthly sales and for forecasting monthly sales with normal weather, the following steps are recommended.

- Estimate monthly energy models using multiple HDD and CDD variables to approximate the nonlinear response of daily energy use, and therefore monthly energy use, to weather conditions.
- If possible, use insights from daily load research data to understand or structure the appropriate weighting between multiple HDD variables and multiple CDD variables.
- Use hourly weather data to construct daily average temperatures as an average of the 24 hourly values. Use these average values to compute daily HDD and CDD values at each base temperature.
- Once daily HDD and CDD values are computed for each station, average these values across weather stations using appropriate weights. If appropriate, use different station weights for winter months and summer months.
- For monthly sales, sum daily HDD and CDD values across days in each billing cycle and weight the sums across cycles. Different cycle weights may be appropriate for different classes.
- For simulating calendar month sales, sum daily HDD and CDD values across days in the calendar month. No weights are involved in this calculation.
- For simulating monthly unbilled sales, sum daily HDD and CDD values across unbilled days for each cycle and weight these sums across cycles.
- To define normal HDD and CDD values, average HDD and CDD values across years for each date. It may be desirable to smooth these results using centered moving averages. The result is a single set of daily normals for each HDD and each CDD variable.

- To define normal HDD and CDD values for cycle months and calendar months, combine the daily data normal values using the same process as for the actual values.

The same daily normal values that are used to construct monthly normal values can also be used in daily models of system loads to normalize daily system energy. The result will be as depicted in the Figure 12, which shows actual daily energy in the top diagram and normalized daily energy in the bottom diagram.

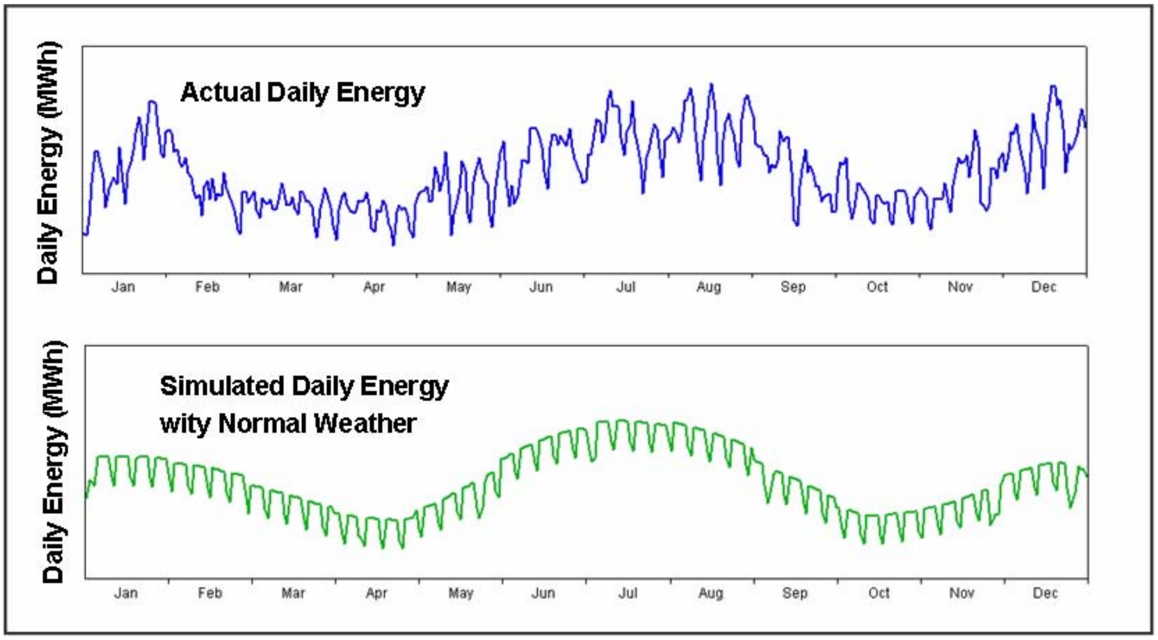

Figure 13: *Depiction of Actual and Normal Daily Energy*

While the above recommendations will work well for simulating normal energy and for energy forecasting, the normal weather variables are not appropriate for simulating normal peaks or for forecasting peaks with normal weather. Recommendations for peak modeling are:

- For estimating and simulating monthly and annual peaks, it is best to build models of actual peaks, explained by the actual weather that produced those peaks. Nonlinearities are less important here, since peak days typically occur on the extreme hot or cold days, where weather response is relatively linear.
- To identify normal peak producing weather, identify historical peak days. Organize the weather data for each peak day and for prior days. These values can then be averaged over years to define normal peak producing weather.
- Some consideration must be given to the swing months, since they may have warm day peaks in some years and cold day peaks in other years. Use the historical data to define each month as a warm peak month or a cold peak month, and average only over the appropriate subset of the data to define normal peak producing weather for these months.

If it is necessary to generate an 8,760 hourly profile for dispatch simulation or other similar purposes, define a third set of normal weather that is based on rank and average methods. Rank and average daily weather values will embody normal extreme values as well as normal averages. The recommended process for generating the 8,760 profile is as follows.

- The normal daily weather values will need to be assigned to a calendar for each forecast year. It may be desirable to rotate this weather so that it starts on the same day type each year (e.g., the first day is always assigned to the first Monday of each year.
- Once the normal (rank and average) daily weather pattern is assigned to days, hourly models can be used to generate 8,760 estimates for each class and the results can be added across classes and adjusted for losses (a bottom-up approach). Or the weather can be used to simulate hourly loads for the system as a whole (a top-down approach).
- The resulting system load estimates can then be calibrated to the calendar month energy and peak forecasts from earlier steps, resulting in a single consistent set of monthly energy values, monthly peak values, and hourly system loads.

Itron Inc.

Itron Inc. is a leading technology provider to the global energy and water industries. Itron Inc. consists of Itron in North America and Actaris outside of North America. Our company is the world's leading provider of metering, data collection and utility software solutions, with nearly 8,000 utilities worldwide relying on our technology to optimize the delivery and use of energy and water. Our products include electricity, gas and water meters, data collection and communication systems, including automated meter reading (AMR) and advanced metering infrastructure (AMI); meter data management and related software applications; as well as project management, installation, and consulting services.

To know more, start here: www.itron.com

Itron Inc.
Corporate Headquarters
2111 North Molter Road
Liberty Lake, Washington 99019
U.S.A.
Tel.: 1.800.635.5461
Fax: 1.509.891.3355

Itron Inc.
Energy Forecasting - West
11236 El Camino Real
San Diego, California 92130-2650
Phone: 1.858-724-2620
Toll Free: 1.800.755.9585
Fax: 1.858.724.2690
Email: forecasting@itron.com
www.itron.com/forecasting

Publication 100904WP-01
09/08