Estimating Hourly Loads for 50 States, by Customer Segment

Josh Bode, Sr. Consultant, Freeman, Sullivan & Co.
Paul Mangasarian, Analyst, Freeman, Sullivan & Co.

ABSTRACT

Hourly load data is critical for understanding the potential of managing system peaks through load reduction. Although AMI systems are being deployed across the country, they will not be in place for several years for most utilities. In fact, hourly load data is often required prior to implementation of AMI systems in order to plan, design, assess and launch demand response (DR) and time varying pricing applications. The amount of load reduction that can be achieved is a function of customer load during system peaking conditions, percent load reductions, participation rates, and technology. A common shortcoming of many DR planning exercises is the lack of reasonably accurate estimates of hourly loads. Not all utilities have load research samples, or if they do, the data is not publicly available for regional or national planning purposes. In addition, not all utilities have or provide hourly load data that distinguishes buildings with and without central air conditioning. This paper summarizes analysis conducted as part of the recent National Assessment of Demand Response Potential[1] that combined publicly available load research data from 21 states with data on annual use, weather conditions, and air conditioning saturations in order to estimate hourly loads, by customer segment, for fifty states and Washington, D.C. The hourly load shapes were subsequently combined with system load data to estimate the customer segment load coincident with system peaking conditions. The database of hourly load shapes developed represents an extremely valuable resource that can be used by utility planners and regulatory agencies interested in estimating DR potential and/or coincident load by customer segment.

Introduction

Hourly load profiles have many applications for system and program planning. They are used in long term forecasting, in program design and targeting, to assess the potential for demand response and energy efficiency and for cost-effectiveness analysis (as an input that drives expected impacts). Average weekday profiles are insufficient for many planning applications designed to quantify or provide resources under system peaking conditions. Load patterns in average days can be substantially different than load patterns on system peaking days, particularly when air conditioner use is prevalent.

Hourly load data are not available for all utilities and states or for all customer segments within states. Not all utilities have load research samples and, if they do, the hourly load profiles are often not published, limiting their use in regional, system-wide (e.g. ISO level) or statewide planning. More importantly, the available hourly load profile data many times does not distinguish between residential customers with and without central air conditioning.

The work presented in this paper was conducted as part of the recent FERC National

---

[1] "A National Assessment of Demand Response Potential," by *The Brattle Group, Inc.*, Freeman, Sullivan & Co., and Global Energy Partners, LLC, *Federal Energy Regulatory Commission*, June 2009.

Assessment of Demand Response Potential. Although databases with hourly profiles are commercially available, they could not be used for the project. A key requirement of the FERC project was that it be transparent and rely on publicly available data. Most commercially available profile databases rely on a combination of building simulation modeling and calibration to available hourly load profiles. Building simulation modeling inherently makes assumptions about occupancy and operations that may or may not be calibrated to actual behavior.

As part of the project, a systematic search for publicly available hourly load shapes was conducted. In total, hourly load shapes were available for 21 states, with a wide variation in central air conditioner saturation, temperatures, humidity and geography. The hourly load data was for specific utilities or, in some instances, specific planning regions within the state (e.g. Texas). For some states, such as California and Texas, the more granular data represented substantial variation across the state in climate and/or humidity. The hourly load data was for specific rates. Although default rates typically vary by customer class and size, there is not a standard practice across utilities or states for determining whether customers are small, medium, or large. The commercial and industrial (C&I) hourly loads were categorized into three size strata – small (<20 kW), medium (20-200 kW), and large (>200 kW) – based on FERC Form 1 data on average hourly load available for each rate.

Figure 1 below provides an overview of the states for which hourly load profile data was obtained.

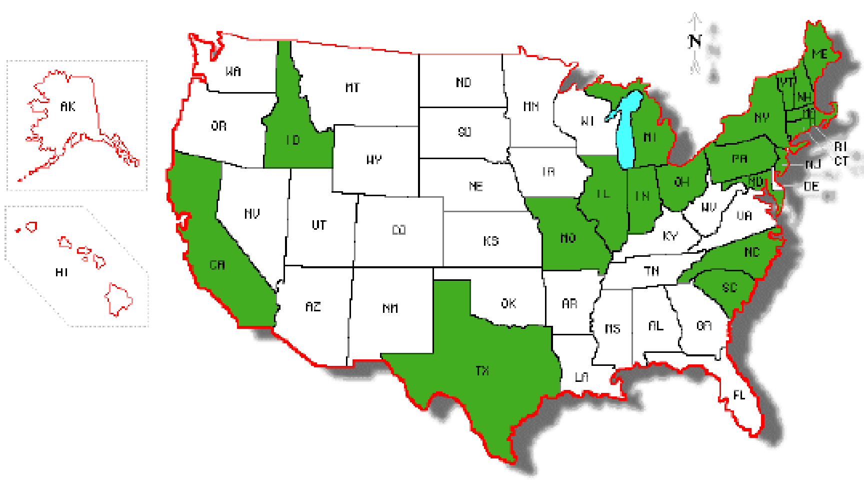

Figure 1: Hourly Load Profile Data Collected by State

Because hourly load data was available on a large cross section of utilities and states, it was possible to use regression analysis to estimate load shapes for each relevant customer segment and to use these models to develop load shapes for all other states and customer segments. In particular, the wide variation of air conditioning penetration, temperatures and humidity allowed for development of estimates for households with and without air conditioning across a wide spectrum of climate conditions.

Methodology

Although the hourly load shapes obtained represented a wide variation of challenges, the data needed to be matched with variables that explain the variation in load shapes. More importantly, the same type of data needed to be available at the state level for all states in order to produce estimates of loads for all fifty states plus Washington, D.C. There are several factors that drive customer loads, including household size, local codes and standards, hours of occupancy, lifestyle, weather and end-uses present at the location. Conceptually, the drivers of load can be divided into factors that affect the magnitude of the load and factors that affect the shape of the load. For example, square footage of a facility affects the magnitude of the load but may not affect the load shape if the operations and end-uses are similar.

A key decision in the analysis was to normalize load shapes by annual usage. By normalizing load shapes, the regression analysis could focus on explaining variation in load shapes rather than explaining variation in the scale of loads across different utilities and states. In addition, normalizing the load substantially reduced the variation in load associated with fixed regional or utility effects. Figure 2 shows the normalized hourly summer loads for an example utility. Although the large C&I load is more than 100,000 times the size of residential load, by normalizing the data, the load shapes can be directly compared on the same graph.

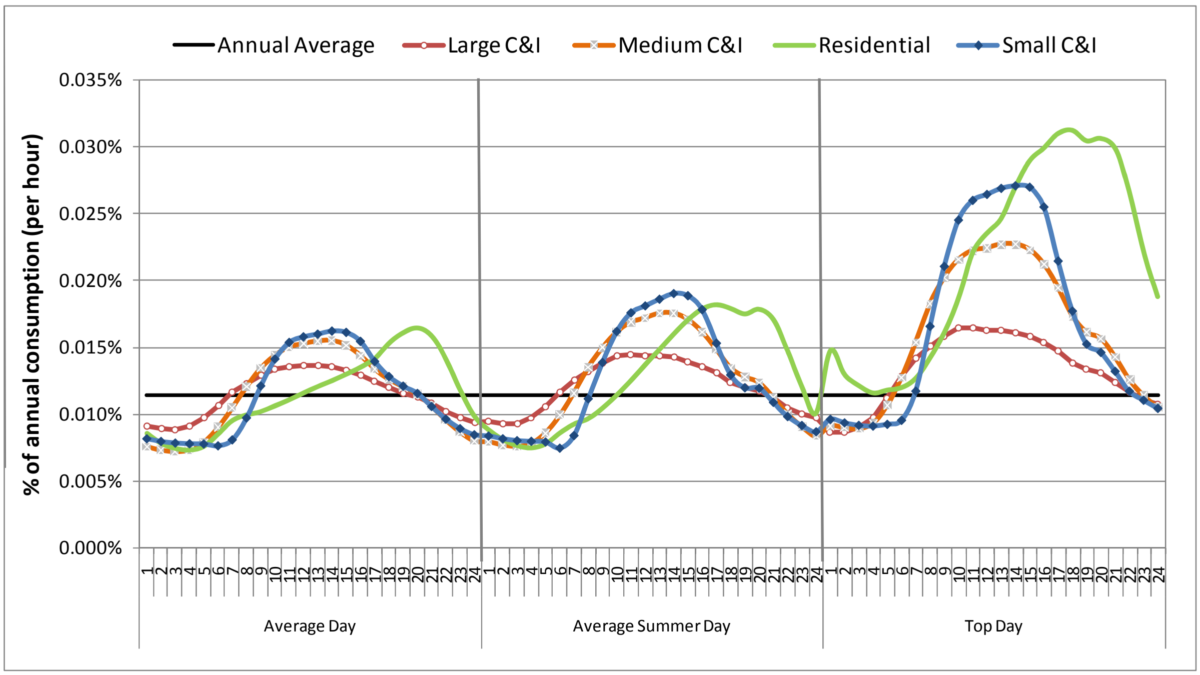

Figure 2: Normalized Hourly Summer Load Shapes for Example Utility

Figure 3 presents the process used to estimate load shapes for all fifty states using available utility level hourly load data. The normalized utility load data was augmented by weather data from the National Climatic Data Center and central air conditioning data to develop the regression models.[2] A separate regression model was developed for each customer type and used to predict the state level normalized hourly loads. Subsequently, the normalized hourly loads were combined with annual consumption data to reflect unique regional factors that affected the scale of the hourly load, but not the shape. In addition, the predicted state level hourly load data was combined with the appropriate system load data from the same time period in order to estimate load coincident with system peaking conditions.

[2] As part of the project, Global Energy Partners collected estimates of central AC saturation for all fifty states. The central AC saturation data can be found in Appendix D of the National Assessment of Demand Response Potential

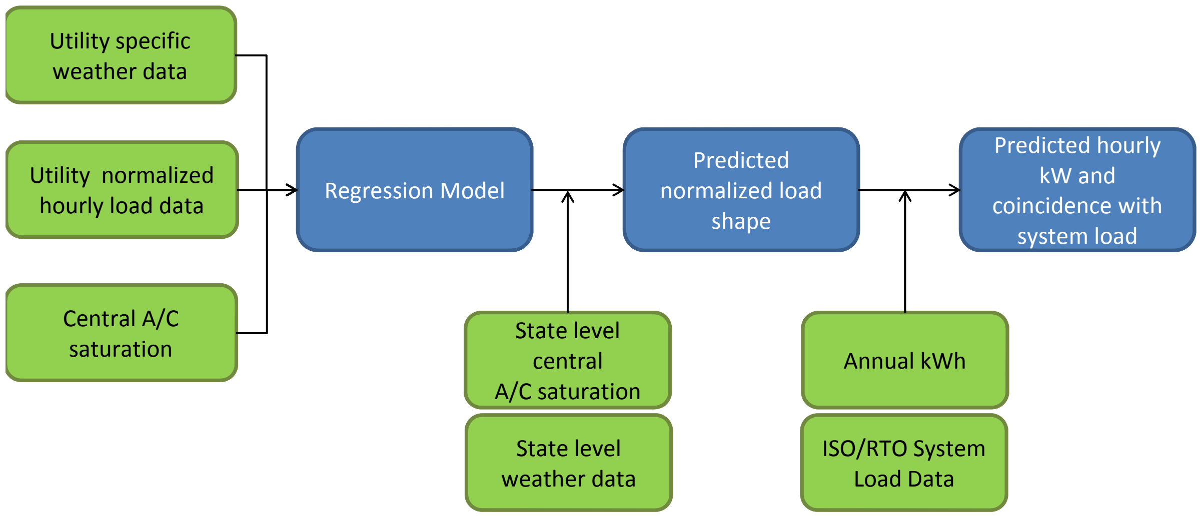

Figure 3: Estimation and Prediction Process

The statistical models were estimated using panel regressions. Each utility load profile represented an individual panel (broken down by utility, region, state and customer class). Each panel contained data in hourly form, for at least one consecutive year's worth of data (8,760 hourly observations), with some panels containing several years of data. The regression models were designed to accurately predict normalized hourly load for electricity customers nationwide given the time of day, day of week, and month, with a focus primarily on the accuracy of the predictions in the months and hours of the day when a DR event is likely to be called. Hourly loads were estimated for the four customer classes: small, medium and large C&I and residential. Separate models were estimated for residential customers in the New England states and non-New England states. This segmentation was intended to reflect inherent differences in the housing stock. Homes in the New England states are typically older and smaller with a much lower central AC penetration due to a lack of centralized vents, but much higher concentration of room air conditioners, a variable for which there is no reliable data source. With the effect of the temperature-based variables in the model scaled directly by central AC penetration, segmenting the residential class ensures appropriate coefficients for these variables. Without the segmentation, the model produced biased estimates at the low end of the saturation of central air conditioning due to the bias in the New England states.

For each customer segment, functional form was closely considered and then several specifications were tested using a fixed-effects panel regression model, with correction for auto-correlation. This approach controls for fixed omitted variables and ensures more robust regression models. Unlike random effect models, it does not assume that fixed effects are uncorrelated with the explanatory variables. In order to ensure the models could be applied outside of the estimation dataset, the amount of variation due to utility specific fixed effects was minimized by using normalized load shapes.

The selection of the final regression model was based on its accuracy under normal and extreme weather conditions and on its theoretical consistency. The same specification was used for all customer segments with the main difference being that central AC penetration varies in the residential segment, while it is held constant for the C&I segments due to lack of state specific commercial and industrial customer central AC saturation data. C&I load is far less dependent on central AC load, and variation in central AC penetration is significantly lower in these segments. As a result, the lack of central AC data for these segments did not compromise the accuracy of the results.

The final models depict normalized energy use for customers across states and classes as a function of variables that capture typical load shapes associated with operational schedules and, for the residential model, variables designed to capture central air conditioning load based on central air conditioning penetration and cooling-degree-hours. The dependent variable in each regression consisted of normalized hourly energy use. The explanatory variables for the residential model were:

- Hourly binary variables to define the typical load profile for a day;
- Monthly binary variables to capture seasonal variation;
- Day-of-week binary variables to capture variation in energy use throughout the week;
- A weekend and holiday binary variable interacted with hourly binary variables to capture the different hourly load profile typically found on weekends or holidays;
- A Monday or Friday binary variable interacted with hourly binary variables to capture the different hourly load profiles found on Mondays and Fridays;
- Cooling-Degree-Hours * Central Air Conditioning Penetration interacted with hour binary variables to capture the impact of air conditioning load across the hours.

Mathematically, the regressions can be expressed by:

$$normalizedkW_{xt} = a_x + \sum_{i=5}^{9} b_i \cdot Month_i + \sum_{k=1}^{7} c_k \cdot Dayofweek_k + \sum_{j=1}^{24} d_j \cdot Hour_j + \sum_{j=1}^{24} e_j \cdot Hour_j \cdot Weekendholiday + \sum_{j=1}^{24} f_j \cdot Hour_j \cdot MondayOrFriday + \sum_{j=1}^{24} g_j \cdot Hour_j \cdot CoolingDegreeHours \cdot CACpenetration + U_{xt}.$$

In this equation,

$normalizedkW_{xt}$ represents the normalized hourly usage for state or utility x at time t;
a - g are estimated parameters;
$Month_i$ is a dummy variable for month i;
$Dayofweek_k$ is a dummy variable for day of week k;
$Hour_j$ is a dummy variable for hour j;
weekendholiday is a dummy variable specifying the day as either a weekend or holiday;
MondayOrFriday is a dummy variable specifying the day as either a Monday or Friday;
CoolingDegreeHours is the cooling degree hours measured as the maximum of 0 or temperature - 65
U_{xt} is the error term;

The accuracy of the models' predictions across all the states hinges on the amount of variation in the load profiles used as inputs. A detailed comparison of predicted to actual loads demonstrated that the models predicted well for all customer classes across various metrics. Overall, the small, medium, and large C&I regression had R-squared values of 0.81, 0.67, and 0.40, respectively. The R-squared is a measure of the amount of variation around the mean explained by the regression. The following figures show the comparison of predicted to actual values for the commercial and industrial classes.

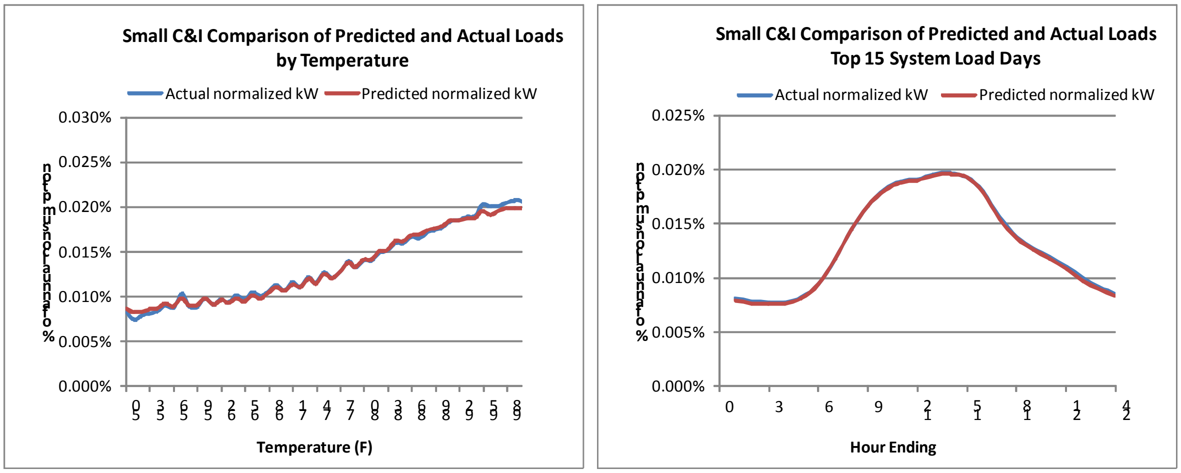

Figure 4: Comparison of Actual and Predicted Values for Small C&I Customers

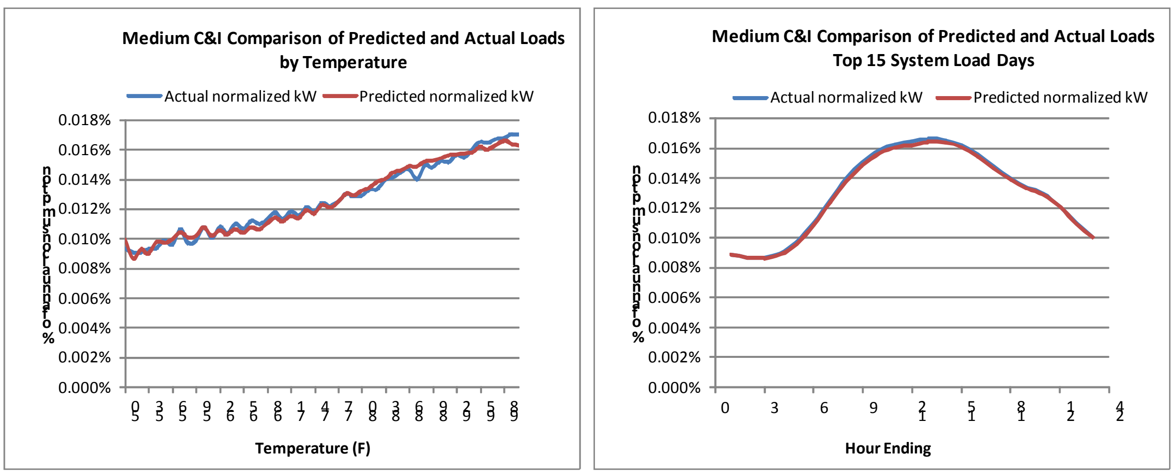

Figure 5: Comparison of Actual and Predicted Values for Medium C&I Customers

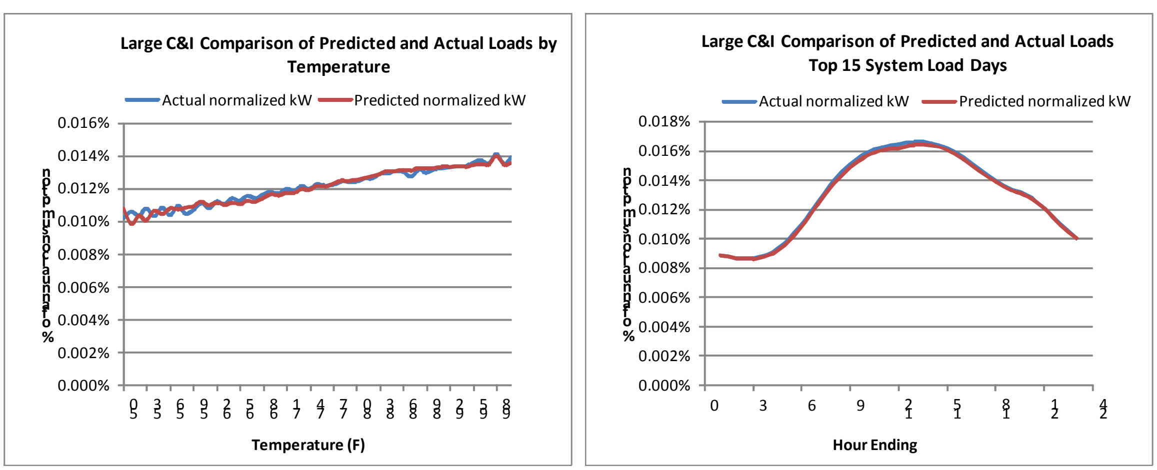

Figure 6: Comparison of Actual and Predicted Values for Large C&I Customers

Despite limitations in the input data, model accuracy is excellent even at the high end of the temperature spectrum and across all hours of the day during peak (top 15) system load days for all C&I segments.

The figures below compare predicted and actual values for the residential model. As with the C&I models, the residential models predict relatively well across the temperature spectrum, though they exhibit a small downward bias for more extreme temperatures. The source of the bias is unclear, but may be due to measurement error for the input values or an omitted variable. In addition, the prediction accuracy was assessed for regions with high and low central AC saturations, as determined by the median state central AC saturation. While the predicted coincident load is more accurate for areas with high central AC saturation, the regression models still exhibit a downward bias during high system load days. While not ideal, this under prediction means that DR potential estimates are conservative because the load estimates are conservative. Furthermore, predictions are more accurate for States with high temperatures from which the majority of residential DR potential will come.

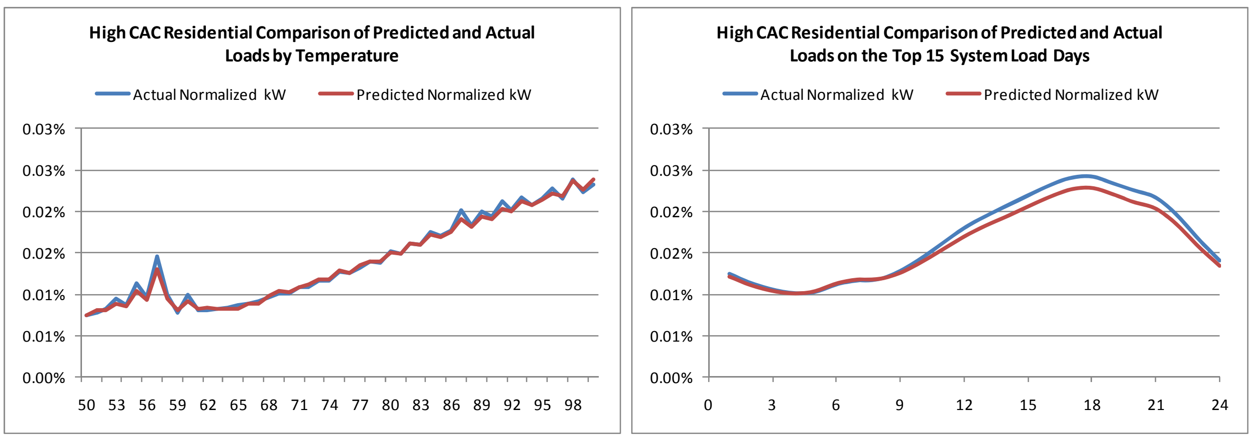

Figure 7: Comparison of Actual and Predicted Values for Residential Customers with High Central AC Saturation

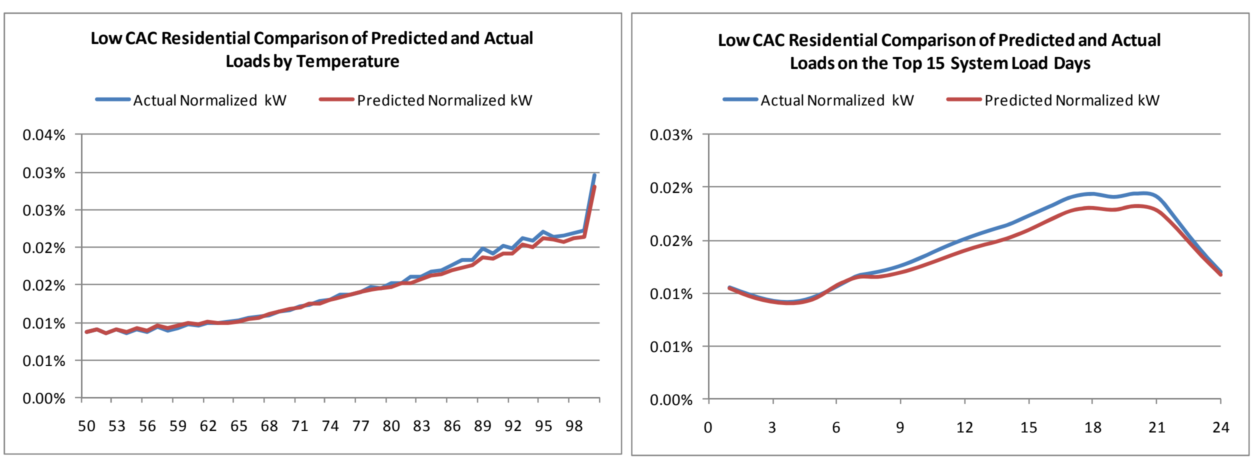

Figure 8: Comparison of Actual and Predicted Values for Residential Customers with Low Central AC Saturation

Results

The analysis produced hourly load data for all fifty states, by customer segment, matched with the system peak. Table 1 summarizes the average, hourly energy use between 2:00 and 6:00 pm on the top 15 system load days for each state and customer segment. The data was used as the basis for estimating load impacts for price-based DR options for each customer segment in the FERC National DR Potential Assessment study. While it is widely known that there is wide variation in the system coincident loads across states, Figure 9 reflects expected patterns and provides, for the first time (to our knowledge), estimates of system coincident loads for all fifty states.

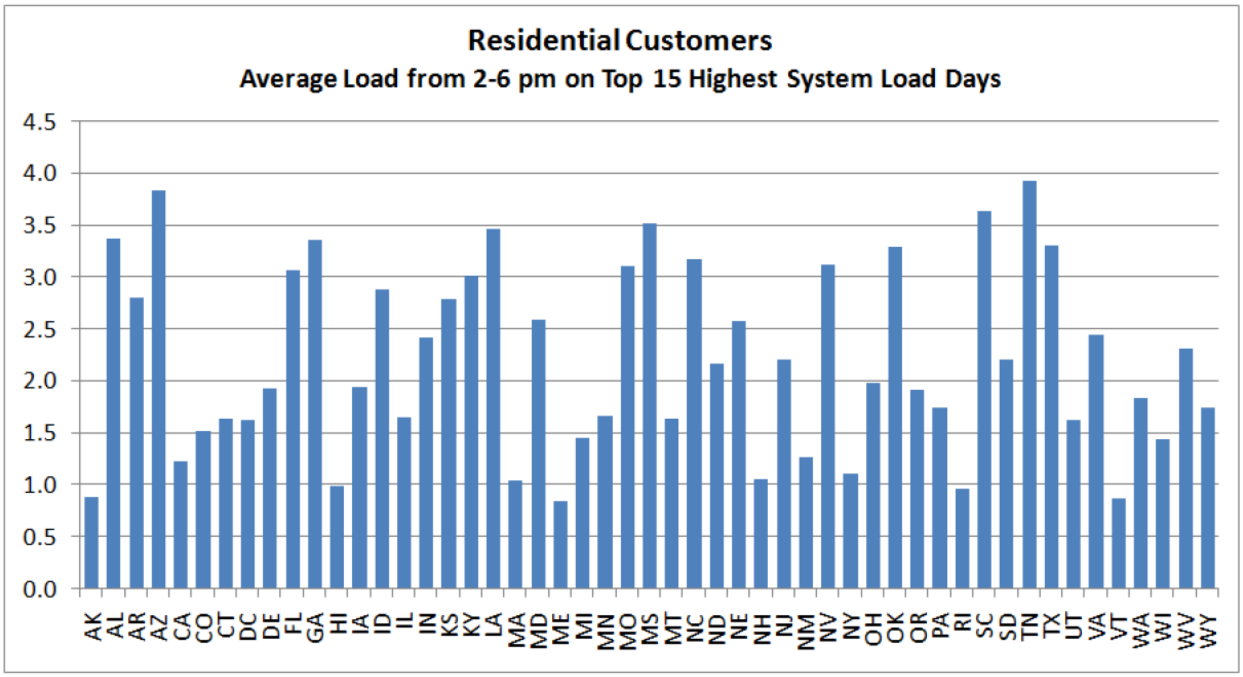

Figure 9: Comparison of Residential Customer Loads by State, Top 15 System Load Days, 2:00-6:00 pm

Table 1: Average Per-Customer Load by Rate Class, 2:00-6:00 pm, Top 15 System Load Days

	Average peak load per customer (kW)			
State	Residential	Small C&I	Medium C&I	Large C&I
Alaska	0.9	4.5	80	1,029
Alabama	3.4	15.1	192	748
Arkansas	2.8	9.1	93	801
Arizona	3.8	16.9	165	822
California	1.2	3.2	38	555
Colorado	1.5	1.9	40	901
Connecticut	1.6	3.9	63	206
District of Columbia	1.6	9.5	158	745
Delaware	1.9	15.2	125	951
Florida	3.1	2.9	40	696
Georgia	3.4	5.4	60	602
Hawaii	1.0	4.2	45	842
Iowa	1.9	4.1	47	709
Idaho	2.9	3.9	31	636
Illinois	1.7	7.3	28	450
Indiana	2.4	6.3	52	798
Kansas	2.8	6.4	44	318
Kentucky	3.0	10.5	176	959
Louisiana	3.5	14.6	39	771
Massachusetts	1.0	6.0	24	642
Maryland	2.6	13.1	32	606
Maine	0.8	2.0	30	571
Michigan	1.5	6.2	48	609
Minnesota	1.7	3.2	42	327
Missouri	3.1	5.0	110	748
Mississippi	3.5	8.8	78	1,215
Montana	1.6	12.3	157	1,101
North Carolina	3.2	5.6	168	1,373
North Dakota	2.2	9.7	129	614
Nebraska	2.6	4.5	128	291
New Hampshire	1.1	4.7	32	306
New Jersey	2.2	7.1	77	395
New Mexico	1.3	4.8	61	707
Nevada	3.1	12.1	112	931
New York	1.3	5.7	81	820
Ohio	2.0	8.5	65	604
Oklahoma	3.3	3.8	70	778
Oregon	1.9	4.5	75	680
Pennsylvania	1.7	8.2	43	644
Rhode Island	1.0	2.7	32	393
South Carolina	3.6	7.6	172	1,696
South Dakota	2.2	9.3	87	402
Tennessee	3.9	11.5	186	376
Texas	3.3	3.7	47	2,086
Utah	1.6	4.9	86	1,322
Virginia	2.5	4.6	88	708
Vermont	0.9	2.2	49	773
Washington	1.8	6.5	110	771
Wisconsin	1.4	4.1	61	782
West Virginia	2.3	6.3	78	1,431
Wyoming	1.7	14.9	66	1,551

Conclusions and Recommendations

The database of hourly load shapes developed represents an extremely valuable resource that can be used by utility planners and regulatory agencies interested in estimating DR potential and/or coincident load by customer segment. In addition, the underlying statistical model can be combined with temperature, central AC saturation data and annual consumption to produce hourly load estimates at the utility level, for utilities without class load research samples. While the primary goal of the analysis was to estimate DR potential for each of fifty states, the models and output produced can be employed in a variety of resource planning exercises.

The hourly load shapes for each state and customer segment were developed quickly and out of necessity as inputs for the larger FERC National DR Potential Assessment. While the results are relatively accurate, there is substantial room for improvement, particularly in distinguishing end-use loads such as central AC, and lighting.

Several future steps could be undertaken to improve the model and make them more useful, including:

- Increasing the amount of hourly load profiles in the estimation dataset
- Reducing error in the input values – some judgment was required in linking load with specific weather stations, and some central AC saturation estimates were a decade old
- Increasing the explanatory variables by incorporating data on:
    - Business mix – which can vary substantially by region, particularly for large C&I customers
    - Central AC saturation for non-residential customers
    - Electric heating saturation
    - Water heating saturation
    - Pool pumps
    - Retail prices, particularly if time varying – e.g., time of use rates
- Publicly posting the database of hourly load shapes and make the data available for downloads
- Provide tools that allow users to quickly customize hourly load data based on weather data, annual consumption, and appliance saturation data.

Although AMI systems that can support hourly data collection are being deployed across the country, they will not be in place for several years and they are unlikely to be universal over the next ten years. More importantly, load data is often required prior to the implementation of AMI systems in order to plan and launch information feedback and time-varying pricing applications. Put differently, there are several reasons to refine, update, and maintain a publicly available set of hourly load shapes.