A Probability–Based Approach for Selecting Hours in a TMY Data File Corresponding to a Utility's Peak

This paper was prepared by Frontier Associates LLC and proposes a formal probability-based method for estimating the summer and winter peak demand reduction from an energy efficiency measure. We envision that this approach would be particularly useful when typical meteorological year (TMY) data and model simulations are used to estimate peak impacts.

Background

It is not always clear how weather-sensitive energy efficiency measures will perform at the exact hour of the utility's annual summer or winter system peak. Many of our calculations and models provide 8760 hourly impact estimates for the change in load associated with the efficiency measure based on typical meteorological year (TMY) data. TMY data contain actual months of weather data from different past years. Consequently, the TMY year does not coincide with an actual year and cannot be matched against actual demand or load data for a utility system or market. The challenge is to determine which of the 8760 values output from a building energy use simulation model to select to represent the demand reduction at the time of the utility's system peak.

Methodology

Frontier's proposed approach to matching a seasonal peak on the ERCOT system with a TMY data file involves the following steps:

- Establish the number of peak hours to be predicted each season (i.e., summer and winter).
- Use a logistic regression model and hourly or interval-level (i.e., 15-minute) data since January 2007 to estimate the relationship between setting a peak hour and a set of explanatory variables, including a temperature variable and dummy variables representing the time of day.
- Use the estimated relationships to assign marginal probabilities to changes in the explanatory variables.
- With the estimated relationships, calculate the probability of setting a peak hour based on TMY weather data.
- Find and average the savings from a building energy use simulation model (that used the same TMY data file) which corresponds to the same hours.

Although the seasonal peak is a measurement based on a single hour (or 15-minute interval, in the case of the ERCOT utilities), predicting a larger set of peak hours might be appropriate in the first step. In Texas, the demand reduction or capacity component of the incentive payments to project sponsors are based on the capital cost of a combustion turbine which normally has an expected annual runtime of 10 to 40 hours in Texas. Building energy use simulation models have stochastic algorithms. So if a single pair of model runs (i.e., a base case and a change case) is used to calculate hourly savings, there is a high

Frontier Associates LLC
1515 S Capital of Texas Hwy., Ste. 110
Austin, TX 78746

probability that savings will be biased for any single hour. So, either multiple model runs must be used to average the estimated hourly savings, or a broader definition of peak (i.e., peak hours) must be used in our approach. Finally, estimating the probability of setting a set of peak hours is much easier than estimating a single peak hour or interval per year with a logistic model. If Y=1 on only six instances (i.e., for six years in the estimation period), more advanced techniques would be required in the estimation (e.g., the use of prior distribution and Bayesian estimation techniques). For these reasons, a set of 20 peak hours is used here.

Note that the second step ignores many other very important factors that might affect the timing of the peak, including actions by industrial energy consumers and load-serving entities to avoid wholesale market price spikes and reduce 4-CP transmission charges.[1] The day of the week is also not considered. However, the inclusion of other variables would prevent the application of this approach when only a TMY weather file and a building energy use simulation model are used to calculate the peak demand reduction associated with an efficiency measure. TMY data are pieced-together from recorded weather during numerous previous years to create a typical year with typical fluctuations. Since the TMY data do not represent weather data from any single "real" year, there would be no way of matching "real" energy price data, the day of the week, or other variables to the fabricated weather data.

Marginal probability can be obtained by estimating a logistic regression or "logit" model. Most statistical software packages can convert the results from a logit model into probabilities.

In the final step, either a simple average or a probability-weighted average (with the weights based on the probability of the seasonal peak being set in a particular hour in the TMY data file) could be used to estimate peak demand reduction.

Example Determination of Peak Hours

An example is illustrated below to further explain the five steps above. In this example, summer data is applied. However, this approach works equally well for estimating winter peak demand reduction. In our comparison of the results obtained from alternative approaches in the following section, we also provide some comparisons of the winter peak demand reduction associated with various measures.

Total ERCOT system load is used in this example, but the use of load data which is specific to transmission and distribution utility (TDU) service areas might be used in practice. Interval-level data were converted to hourly values to facilitate the estimation and provide a better match of load to hourly temperature data. The top 20 hours of each summer season of each year were coded 1, and all other hours were coded as 0. Variables representing the hour with intervals ending 61, 65, and 69 were included to capture time-of-day factors affecting electricity use. All hours outside of the afternoon were assumed to have zero probability of being a peak hour and were eliminated from the dataset to facilitate estimation. Additionally, variables representing month (July and August) were also included.

[1] An application of this probabilistic method in a situation where utility load data can be matched with actual weather and actual electricity prices can be found in Jay Zarnikau and Dan Thal, The response of large industrial energy consumers to four coincident peak (4CP) transmission charges in the Texas (ERCOT) market, *Utilities Policy*, Sept. 2013, Vol. 26, pp. 1-6.

Because summer peak loads are largely determined by air conditioning usage in Texas, a variable was constructed to represent the ratio between the actual temperature in a central location within the ERCOT market (Austin) for a given interval and the highest temperature reading during the given year (*RelativeMaxTemp*).

The resulting model was thus:

Logit(*CP*) = f(*RelativeMaxTemp*, *Interval61, Interval65, Interval69, July, August*) **(1)**

This relationship was estimated using R software as a general linear model with a binomial distribution.

Once the marginal probabilities are estimated, the probability of each hour of the TMY file setting a peak hour can be calculated. This can be simulated with R.

The 20 hours in the TMY file assigned the highest probability of being within the set of peak hours are identified. We then find those same hours in the output from a building energy use simulation model that used the same TMY data file. The average of the energy efficiency measure's hourly savings over those 20 hours provides an estimate of the savings associated with the efficiency measure coincident with the summer peak.

For the 20 peak hours analysis, the logistic regression result is as follows:

Table 1. Logistic Regression Statistical Results for the 20 peak hour analysis

	Estimate	P-value
Intercept	-39.3331	<0.0001 ***
RelativeMaxTemp	36.1022	<0.0001 ***
Interval61	1.7570	0.000131 ***
Interval65	1.9924	<0.0001 ***
Interval69	1.5439	0.001012 **
July	0.9284	0.016848 *
August	1.7722	<0.0001 ***

As we can see from the table, all of these results are significant, so the estimated coefficients can be interpreted as follows:

For the "RelativeMaxTemp" variable, the estimated coefficients mean that for every unit increase in "relative maxtemp" (the ratio between the actual temperature in a central location within the ERCOT market (Austin) for a given interval and the highest temperature reading during the given year), the log odds of being a peak hour versus not being a peak hour increase by 36.1022.

For the "Interval61", "Interval65" and "Interval69" variable, "1.7570" means the log odds of being a peak hour in 3:00-4:00pm versus being a peak hour in 2:00-3:00pm increase by 1.757, and the log odds of being a peak hour in 4:00-5:00pm versus being a peak hour in 2:00-3:00pm increase by 1.9924. The log odds of being a peak hour in 5:00-6:00pm versus being a peak hour in 2:00-3:00pm increase by 1.5439.

For the "July" and "August" variables, "0.9284" means the log odds of being a peak hour in July versus being a peak hour in June increase by 0.9284, and the log odds of being a peak hour in August versus being a peak hour in June increase by 1.7722.

To be more specific, consider an example of an hour (3:00pm-4:00pm) in August, the hourly temperature being 100F, the annual highest temperature being 102F, so the estimated log of the odds ratio of being a peak hour versus being a non-peak hour in this specific hour is: 39.3331+36.1022*100/102+1.757+1.7722=-0.4096. Thus the probability of getting a peak hour during that time is exp(-0.4096)/(1+exp(-0.4096)=0.4.

Based on these results, all summer afternoon hours are ordered according to the probability and top 20 hours are selected:

Table 2. 20 peak hours with the highest probability

Date	Time	temperature	maxtemp	RelativeMaxTemp	logodds	probability
7/28/1995	17:00	102.02	102.02	1	-0.3101	0.42309
8/5/2004	16:00	100.04	102.02	0.980592041	-0.40961	0.400743
8/5/2004	17:00	98.06	102.02	0.961184082	-0.86764	0.295746
7/28/1995	16:00	100.94	102.02	0.98941384	-0.92768	0.283395
8/20/2004	16:00	98.06	102.02	0.961184082	-1.10304	0.249171
7/28/1995	18:00	100.94	102.02	0.98941384	-1.14078	0.242177
8/20/2004	17:00	96.98	102.02	0.950597922	-1.24982	0.222731
7/27/1995	17:00	98.96	102.02	0.970005881	-1.39295	0.198937
8/3/2004	16:00	96.98	102.02	0.950597922	-1.48522	0.18464
8/4/2004	16:00	96.98	102.02	0.950597922	-1.48522	0.18464
8/11/2004	16:00	96.98	102.02	0.950597922	-1.48522	0.18464
8/19/2004	16:00	96.98	102.02	0.950597922	-1.48522	0.18464
8/26/2004	16:00	96.98	102.02	0.950597922	-1.48522	0.18464
8/3/2004	17:00	96.08	102.02	0.941776122	-1.56831	0.172457
8/4/2004	17:00	96.08	102.02	0.941776122	-1.56831	0.172457
8/19/2004	17:00	96.08	102.02	0.941776122	-1.56831	0.172457
8/26/2004	17:00	96.08	102.02	0.941776122	-1.56831	0.172457
7/27/1995	16:00	98.96	102.02	0.970005881	-1.62835	0.164056
7/24/1995	17:00	98.06	102.02	0.961184082	-1.71144	0.152977
7/26/1995	17:00	98.06	102.02	0.961184082	-1.71144	0.152977

Once these calculations were performed for various regions of Texas using utility hourly or interval load data and local weather data, the hours of interest in the TMY weather file would be selected and applied whenever a building energy use simulation model was used to calculate the impacts of an efficiency measure. The statistical analysis would need to be repeated only when new TMY files were available from the U.S. government or when it was suspected that the relationship between a utility's peak and weather or the timing of the peak was changing.

Example Application of the Selected Peak Hours to Energy Efficiency Savings Profiles

Application of this approach to a simulation of the savings associated with ceiling insulation and air infiltration in an electrically-heated home in Austin is presented in this section. We also examine the savings from two lighting-related energy efficiency measures.

We developed a whole home simulation model using EnergyGauge, a simulation software tool that uses a DOE-2 simulation engine. We selected prototype home characteristics using available data on the construction, occupancy, and equipment characteristics of Texas homes. The prototype home characteristics for this example are listed in Table 3. The rows labeled "Ceiling Insulation" and "Air Infiltration" state the base and change conditions.

Table 3. Home Characteristics Inputs used in Simulation Model

Input	Value	Source
Conditioned Area	1,915 square feet	Weighted average total conditioned square feet of Texas single family detached (SFD) homes.
Site Plan	1 story square, 43'9"x43'9"	78% of Texas SFD homes are 1 story per 2009 RECS; a square home is agnostic to orientation.
Bedrooms	3	Majority of SFD homes (53%) have 3 beds.
Bathrooms	2	A plurality of SFD homes (41%) have 2 baths.
Foundation	Slab-on-grade, no insulation	Majority (76%) of SFD homes have a slab.
Ceiling Insulation	For Air Infiltration measure R-22. For Ceiling Insulation measure: Base R-2.5, Change R-30.	The average ceiling/wall insulation level for homes existing before 1998 is R- 20.51/10.94, per the SPS 1998 baseline study. Otherwise assuming all homes built from 1998 on had an average of R- 30/13 (IECC 2009 code requirements). Per RECS, 78% of Texas SFD homes are pre-1998, and 22% are from 1998 on. Taking the weighted average U values of insulation the result is an overall average of U- 0.0882/0.0454, or R- 11.3/22.0.
Wall Insulation	R-11.3	See above.

Input	Value	Source
Window Area	210 square feet	Per RECS, the average Texas home has 14 windows, assuming an average size of 3'x5' that makes for 210 SF of windows.
Air Infiltration	For Ceiling Insulation measure: 12.2 ACH50 For Air Infiltration measure: Base 12.2 ACH 50, Change 7.43 ACH50	Based on LBNL's ResDB (http://resdb.lbl.gov/): US average of 0.61 Normalized Leakage (NL) rate for SFDs; per ResDB 0.5NL = 10 ACH50, so 0.61 NL = 12.2 ACH50.
Window U-value	.78	Combined the prevalence of single, double, and triple paned glass in Texas SFDs from RECS (58/41/1%) with the average U and SHGC for each pane level from LBNL's RESFEN database (http://windows.lbl.gov/software/resfen/resfen.html; excluding windows with high solar gain coatings).
Window SHGC	.56	See above.
Thermostat Settings	Heating: 71.3 during the day when someone is home, 67.7 during the day when no one is home, 69.8 at night Cooling: 74.1 during the day when someone is home, 76.6 during the day when no one is home, 73.9 at night	Weighted average reported thermostat set points from RECS. Times associated with these set points are assumed to be the same as those specified by EnergyStar.
Duct Losses	18% total loss	From LBNL's ResDB (http://resdb.lbl.gov/). National average total duct leakage is 18% of air flow.
Air Conditioning	11.3 SEER	Result of combining the average age of central cooling equipment from RECS with annual shipment-weighted SEER values from DOE.
Electric Heater	COP of 1	Fundamental property of electric resistance.

The simulations assumed:

- Air infiltration: 2.8 ton AC, 3.5 ton heating capacity
- Ceiling Insulation: 4.3 ton AC, 4.8 ton heating capacity

Table 4 compares estimates of the demand reduction of various scenarios associated with Frontier's proposed approach described above along with other definitions mentioned in this memo. For the 20 peak hours in the TMY file, Frontier gets 6 peak hours in July and 14 peak hours in August, ranging from 3:00pm – 6:00pm. Scenarios include overall demand reduction from two efficiency measures: ceiling insulation type and air infiltration. The estimated average summer demand reduction in Austin for a prototype home for these 20 hours is 2.09 kW for the ceiling insulation efficiency measure and 0.34 kW for the air infiltration efficiency measure. Plus, Frontier has also estimated the demand reduction of indoor and outdoor lighting in Austin. For indoor lighting kW savings, Frontier assumed an average of 30% of the original usage would be saved if some energy-saving indoor lighting equipment is installed. Thus an average of 0.062 kW savings could be calculated based on the EnergyGauge home simulation model during 20 summer peak hours. For outdoor lighting in Austin, Frontier considered a variety of outdoor lighting equipment and assumed that 5 kW savings when the outdoor light is on is a reasonable deduction. Since none of the summer 20 peak hours is at night, Austin outdoor lighting demand reduction is 0 kW. However, the demand reduction in winter would be 5 kW since all of the peak hours happen after sunset and before sunrise.

Table 4. Summer Peak Demand Reduction for Various Efficiency Measures from Different Approaches

	Ceiling Insulation Austin (kW)	Air Infiltration Austin (kW)	Indoor Lighting Austin (kW)	Outdoor Lighting Austin (kW)
Frontier's Proposed Approach	2.089	0.341	0.062	0
Top 2 Hours of All Peak Months (170 Hours)	1.531	0.257	0.087	0
Top Hour Top Month (8 Hours)	1.920	0.347	0.052	0
Heat Wave	2.036	0.344	0.056	0
510 Hours	1.511	0.241	0.069	0

Generally, the definitions involving the highest number of hours yield the smallest estimated peak demand reduction. Our proposed method provides estimates similar to the "heat wave" method for the summer.

Frontier Associates LLC
1515 S Capital of Texas Hwy., Ste. 110
Austin, TX 78746

Winter savings calculations by Frontier's proposed approach can be implemented in the similar steps above, except "*RelativeMaxTemp*" needs to be replaced by "*RelativeMinTemp*" to represent the ratio between the actual temperature in a central location within the ERCOT market (Austin) for a given interval and the lowest temperature reading during the given year. The winter kW savings estimated by various approaches are as follows:

Table 5. Winter Peak Demand Reduction from Different Approaches

	Ceiling Insulation Austin (kW)	Air Infiltration Austin (kW)	Indoor Lighting Austin (kW)	Outdoor Lighting Austin (kW)
Frontier's Proposed Approach	2.253	0.810	0.134	5
Top 2 Hours of All Peak Months (124 Hours)	0.601	0.197	0.183	5
Top Hour Top Month (6 Hours)	2.756	0.746	0.064	5
Heat Wave (Cold Snap)	1.327	0.265	0.123	5
510 Hours	0.933	0.239	0.105	3.583

Please note that the "Top Hour Month" calculation is based on the average impact of the measure over 6 hours. A more-recent proposal from the EM&V team suggests that 22 hours be used in this calculation.

Please also note that the "Cold Snap" calculation process in winter is somewhat different from the "Heat Wave" calculation process applied in summer. Based on the EM&V team definition, "the calculation would need to be performed for both the AM and PM periods. For each one of these periods, a 2-hour rather than a 3-hour peak period might be appropriate, and the month range would also need to be selected (Dec – Feb is the current definition). For the temperature selection criteria, the full 4-hour periods could be used instead of the DEER 12-to-6pm period." For TMY data, the strongest "cold snap" happens on December 19th, so the 3 chosen days would be December 19th, December 20th and December 21st. And the coldest hours within the winter peak period (6-10am and 6-10pm) would be 6-8am and 8-10pm. Consequently, an average of 2*2*3=12 hours kW savings is calculated.