2007 ERCOT Planning

Long-Term Hourly Peak Demand and Energy Forecast

May 8, 2007

Executive Summary

The 2007 long-term peak demand and energy forecast for the ERCOT region is presented in this report, including the methodology, assumptions and data upon which this forecast is based. The 2007 forecast is based on the latest historical hourly demands for the region, adjusted for economic and weather variables (primarily temperatures, heating and cooling degree-days). The forecast does not account for interruptible demand and ancillary service programs since those programs are accounted-for in the ERCOT Capacity, Demand and Reserves report as reductions to demand for the purpose of reserve calculations.

The 2007 summer peak demand forecast of 63,794 MW represents an increase of 2.3% from the 2006 actual peak demand of 62,339 MW which was also ERCOT's all-time peak demand. The ERCOT system forecast for 2007 is 3.5% higher than last year's forecast mainly due to a more optimistic economic outlook for the state of Texas, including ERCOT's territory, and adjustments to the model's weather sensitivity.

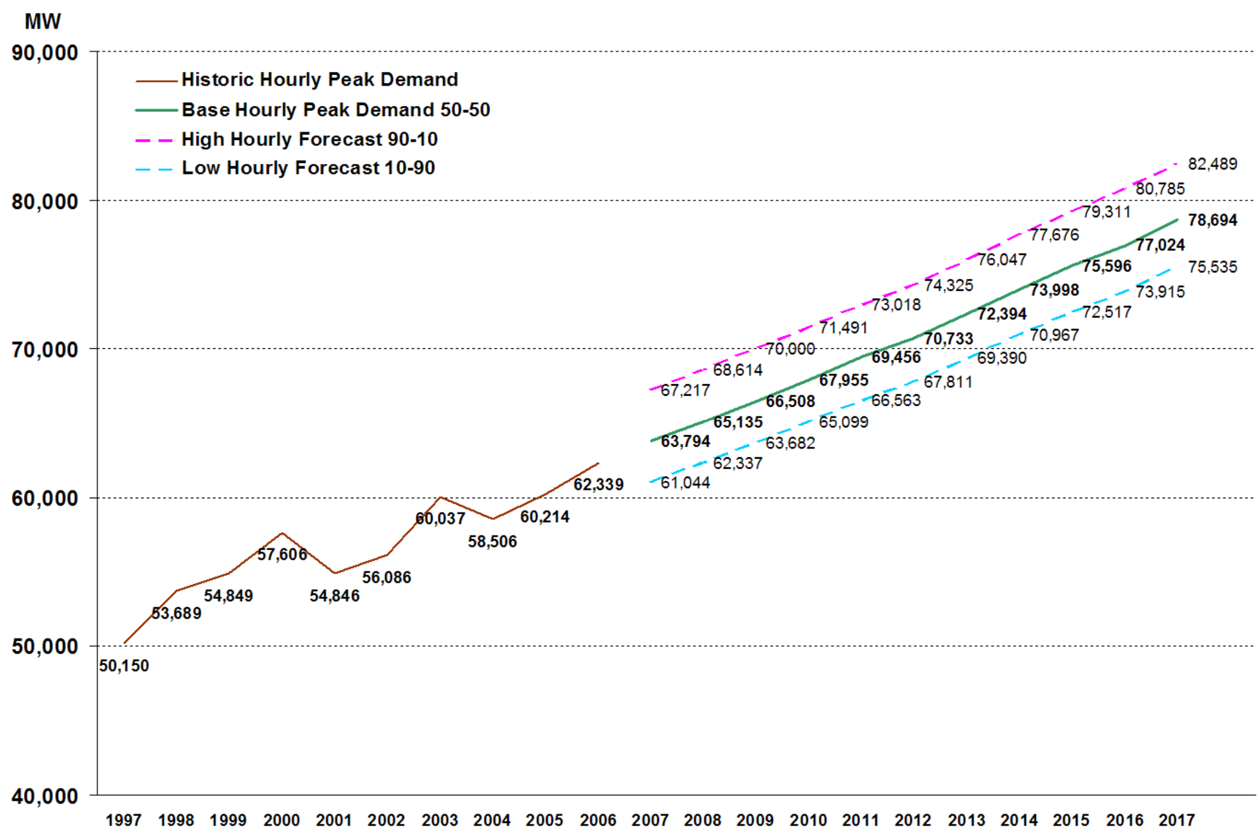


Figure 1 – Historic and Base Forecast Hourly Peak Demand

Figure 1 shows a plot of the historic peak demand from 1997 to 2006 and forecasts from 2007 until 2017. The forecast growth rate for the next ten years (2007-2017) is 2.12% and goes down to 1.92% when computed from 2007 to 2025. The historic compound growth rate for the last ten years has been approximately 2.45%, with strong growth for Texas experienced since 2003 and extraordinary demographic factors in 2006 such as the surge of in-migration from Hurricane Katrina causing an increase in the population in the Coastal weather zone.

The 2007 long-term hourly peak demand forecast, on the average, is 0.70% higher than the forecast produced last year for 2007 to 2015. The key factor driving the higher peak demands and energy consumption is the overall health of the economy, as measured by economic indicators such as the real per capita personal income, population, and various employment measures including non-farm employment and total employment. The model was also recalibrated to include the effects of having an additional year of historic data, which caused a portion of the forecast increase as well.

Also shown in Figure 1 are the forecast scenarios using statistical analysis and extreme weather profiles. The red dash line on the top is a plot of the system peak demand forecasts using temperatures above 90% of the historical temperatures (90^{th} percentile) experienced during the last twelve years. This extreme forecast is referred to in the figure as the extreme hourly forecast 90-10. The low hourly forecast 10-90 refers to the forecasts obtained by using temperatures above 10% of all temperatures during the last twelve years. The forecast for 2007 is 63,794 MW and the preliminary 90% band is 67,209 MW or 5.35% higher than the forecast using normal weather.

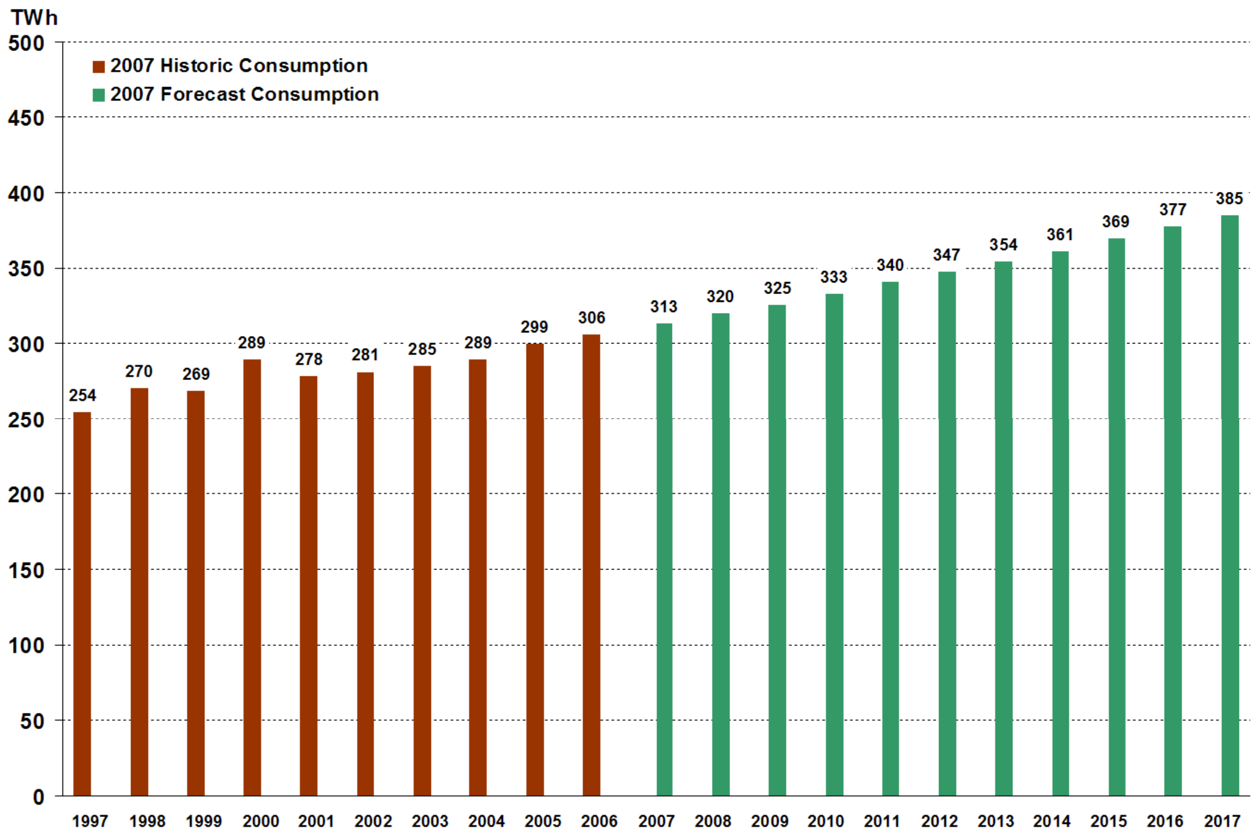


Figure 2 – Historic and Forecast Energy (TWh) Consumption

The energy forecast from 2007 to 2015 is, on the average, 0.06% higher than last year's forecast. The energy consumption is projected to grow at a 2.08% over the 2007 to 2017 period. A one time adjustment due to economic revisions and other factors, such as Hurricane Katrina, contributed to the energy growth rate of 2.40% from the actual energy in 2006 to the forecast for 2007. One of the key factors driving the long-term higher energy consumption is an improvement in the outlook of the overall health of the economy as captured by economic indicators such as the real per capita personal income, population, and various employment measures including non-farm employment and total employment. If income is growing at a faster rate than population, the average person expects to enjoy an overall higher standard of living. A higher standard of living generally translates into an improvement in comfort, which in many cases directly translates into increases in electricity consumption. The energy forecast scenarios show a rather slight degree of variability between the 90-10 high weather forecasts and the median (50-50) base case. The same holds true for the 10-90 forecast scenario. The projected energy consumption shows a similar growth as the 1997 to 2006 period (2.09%).

Introduction

This report gives a high level overview of the forecasts obtained from the 2007 Long-Term Forecast Model. The methodology is briefly described, highlighting the major aspects involved in producing the forecast, including the data input used in the process. Second, a historical perspective of the load growth in the ERCOT's territory is provided and final results of the forecast peak demands and energy from 2007 to 2025 are presented in a graphical form and summarized in a table. Third, a discussion of the major drivers of peak demands and energy consumption is included, along with the uncertainties associated with the forecast, and the differences with last year's forecast. The final hourly load shape forecast is presented in a graphical form giving a perspective or comparison of the actual and forecast trends out into the next eight years (2007-2015). Finally, the more detailed econometric forecasting methodology used by ERCOT is described in Appendix 3.

General Background: Forecast Development Description

The 2007 Long-Term Demand and Energy forecast was produced with a set of econometric models that use weather, economic and demographic data and calendar variables to capture and project the long-term trends in the historical data for the past five years.

A representative hourly load shape by weather zone is forecasted using an average weather profile of temperatures and Cooling Degree Hours (CDH) and Heating Degree Hours (HDH) obtained from historical data to project the load shape into the future. Other factors such as seasonal daily, weekly, monthly and yearly load variations and Holiday events, in addition to various interactions, such as of weather and weekends and weekdays are also considered. This hourly Load Shape only describes the hourly load fluctuations within the year and in itself does not reflect the long-term trend.

The long-term trend is provided by the energy forecast. The monthly energy forecast models by weather zones use Cooling Degree Days (CDD) and Heating Degree Days (HDD), economic and demographic data, and indicator variables for special events to project the monthly energy for next nineteen years (2007-2025).

Data Sources

Economic and demographic data at the county level are obtained on a monthly basis from Moody's Economy.com. These data are used as input to the monthly energy models.

Twelve years of weather data are available from WeatherBank for the 20 ERCOT weather stations. These weather stations are used to develop weighted hourly weather profiles for each of the eight weather zones. These data are used in the Load Shape models. Monthly CDD and HDD are used in the monthly energy models.

The economic and demographic, and weather data are provided by the vendors above, and as such, are proprietary data and under contracts which require that these data not be released to the public.

Settlement load data are available on an hourly basis since July 31, 2001. Prior to 2001, ERCOT has Transmission and Distribution Service Providers (TDSP) hourly data going back to 1995. Weather zone load data have been collected only from July 31, 2001.

ERCOT's Historic and Forecasted Peak Demands and Average Load Growth

The Figure 3 (below) compares the ERCOT's average hourly load with the annual system peak hour load. The growth of the average hourly load is considered almost as a fixed amount that can be estimated with a reasonable degree of accuracy. The peak demand growth, however, is a much more volatile variable and more difficult to predict. The many factors affecting peak hour demand and the high degree of uncertainty in the long run make it a challenging variable, in term of assessing its behavior in the future.

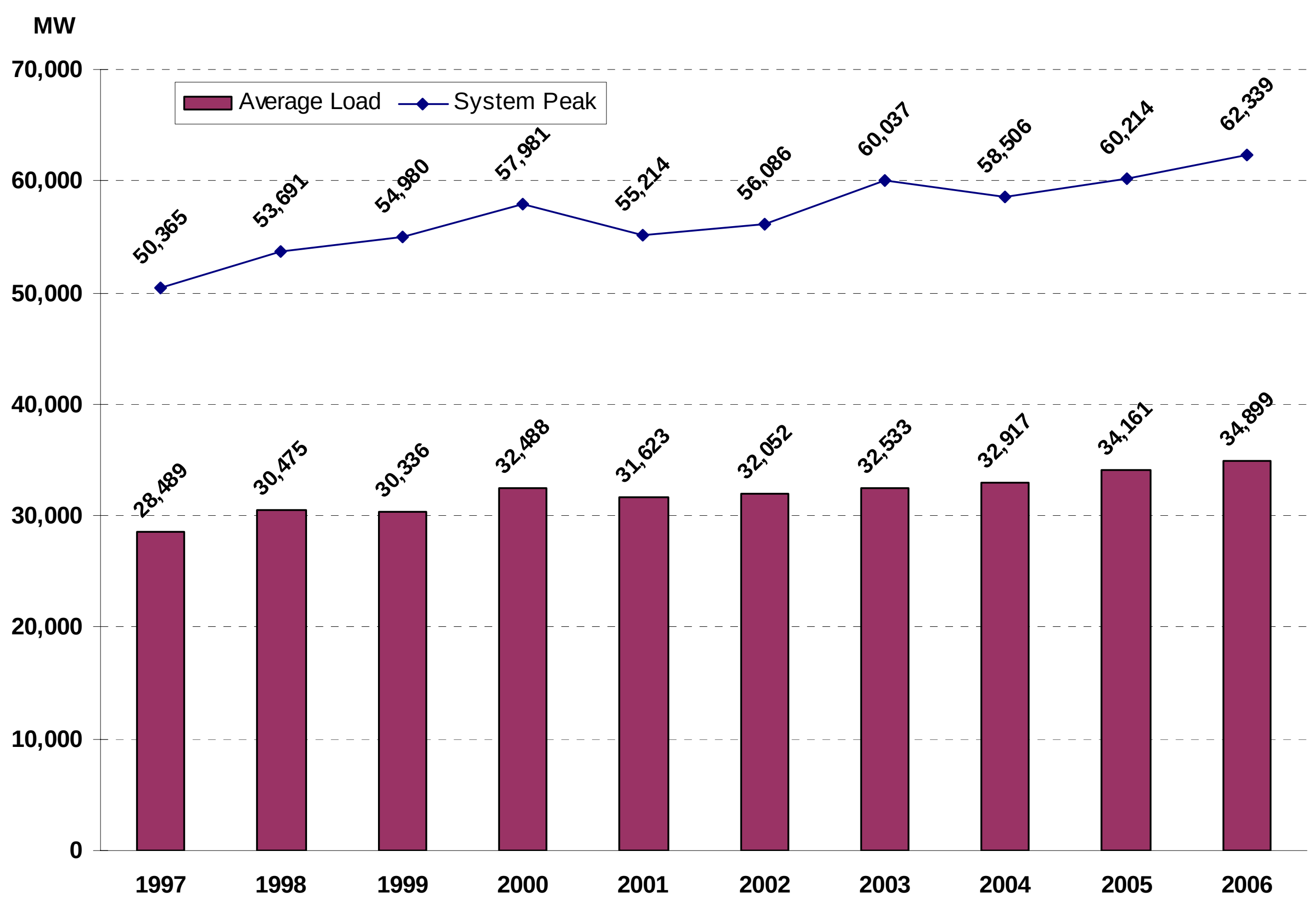


Figure 3 – ERCOT Historic Average Load versus System Peak Growth

Over the last ten years, ERCOT's average hourly load has grown 22.50%. On the other hand, ERCOT's system peak grew 23.78% or 1.30% more than the average. The average annual growth rate was 2.47% over this period. Over the last five years, a similar pattern can be detected. The average growth was 8.88% versus 11.15% for the system peak. The average growth rate over the five year period from 2002 to 2006 was 2.23%.The 2007 Long-Term peak demand and average load forecast is graphed below in Figure 4. Over the ten year period (2007-2017) the average load id projected to grow 22.90% or at a 2.28% growth rate. The system peak load's growth over the same period is 23.35% with a 2.33% growth rate.

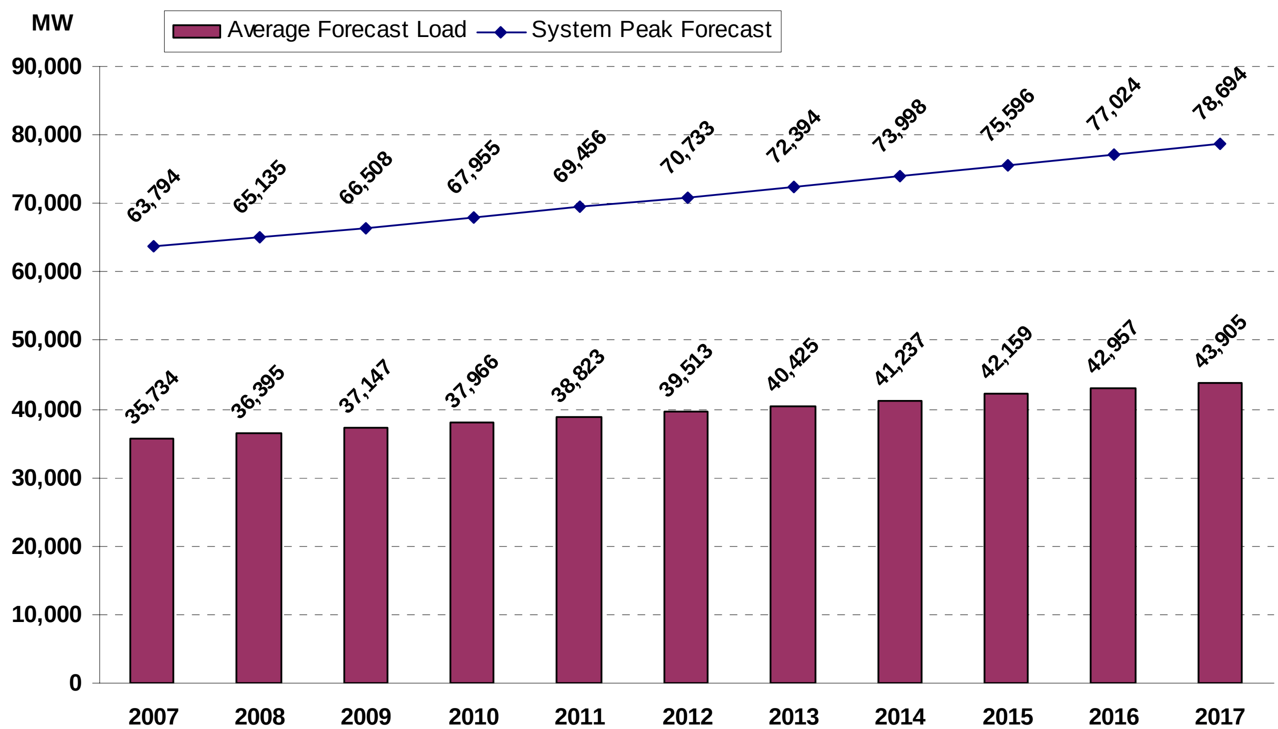


Figure 4 – ERCOT Forecast Average Load Growth versus System Forecast Growth

ERCOT's Peak Demand and Energy Forecasts

The annual historic and forecast peak demands, and the energy consumption, are plot in figure 5 below. The historic peak demand compound growth rate from 1997 to 2006 was 2.08% and the energy growth rate over the same period was 2.45%. By comparison, over the last five years, from 2002 to 2006, the peak and energy grew at 2.68% and 2.15% correspondingly. The 2007 peak demand and energy forecast produced growth rates of 2.36% for the peaks from 2007 to 2017 and 2.31% for the energy over the same period.

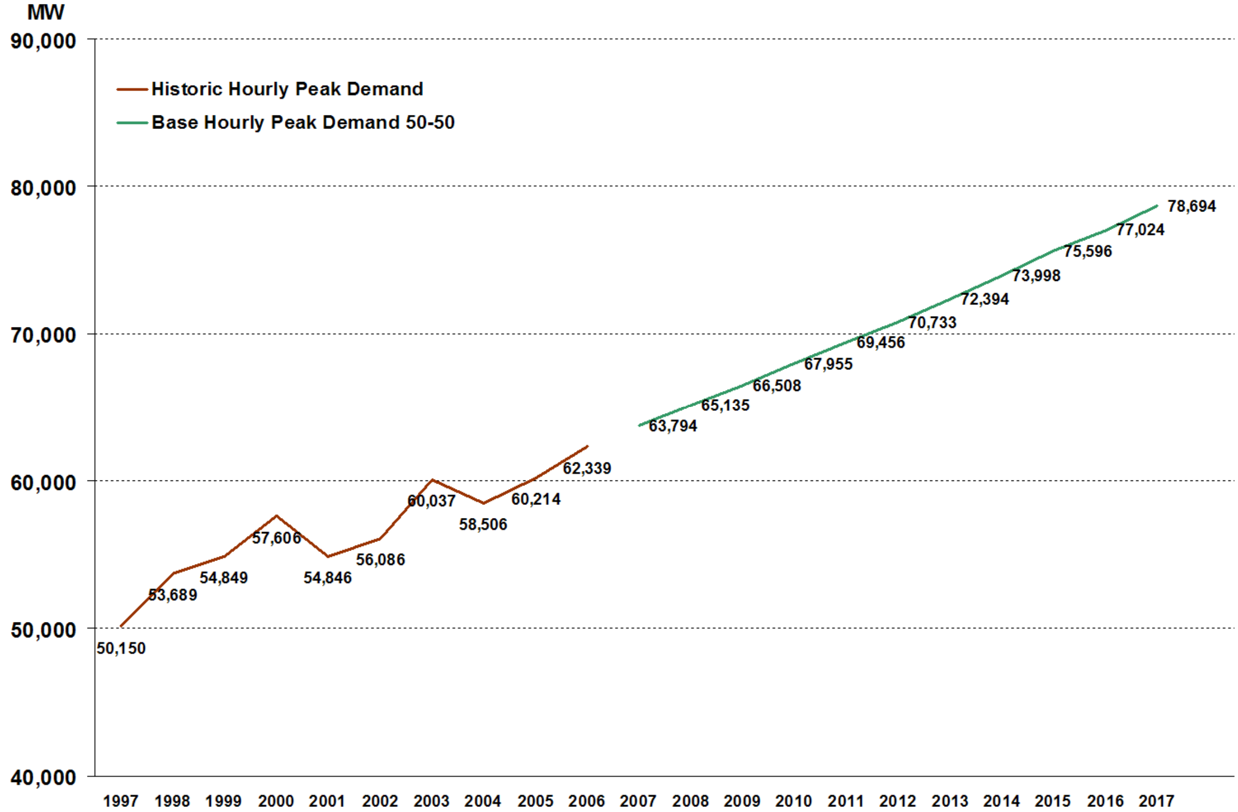


Figure 5 – Historic and Forecast Hourly Peak Demands

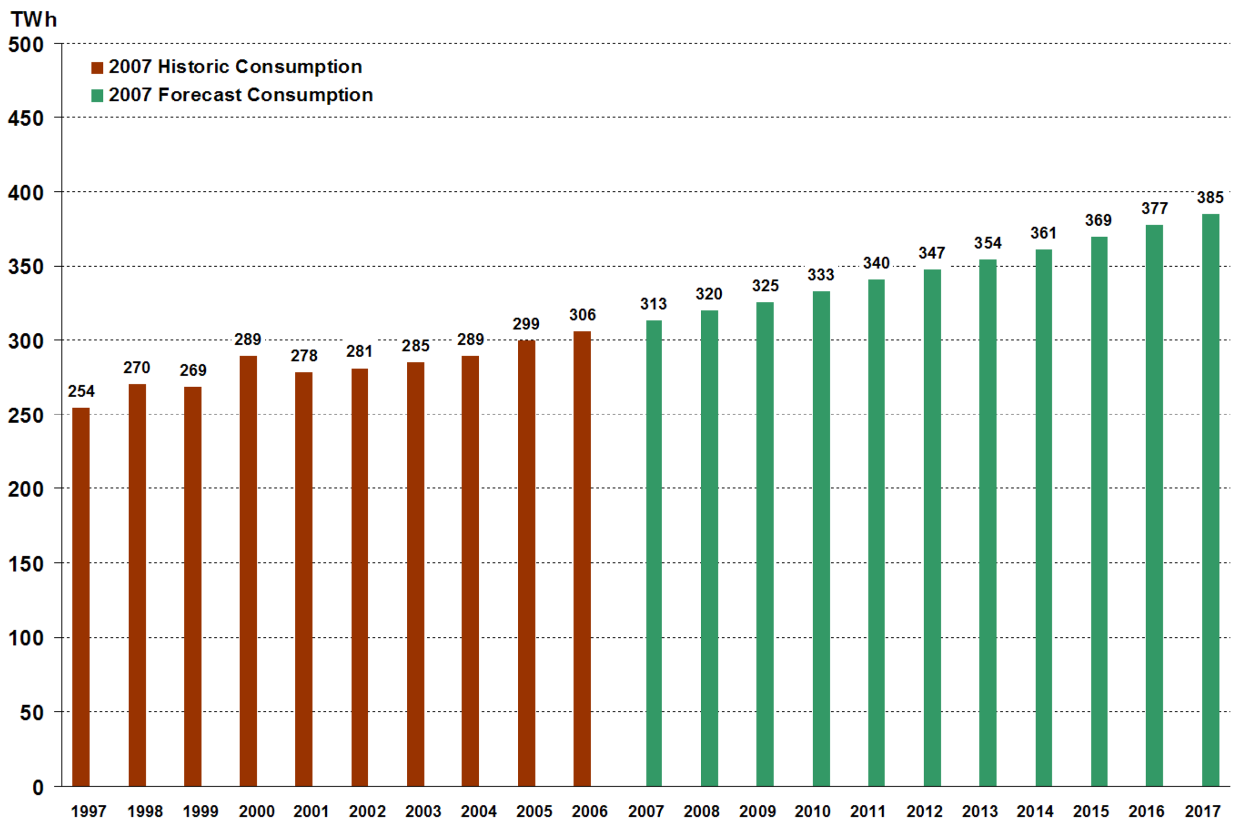


Figure 6 – Historic and Forecast Energy Consumption

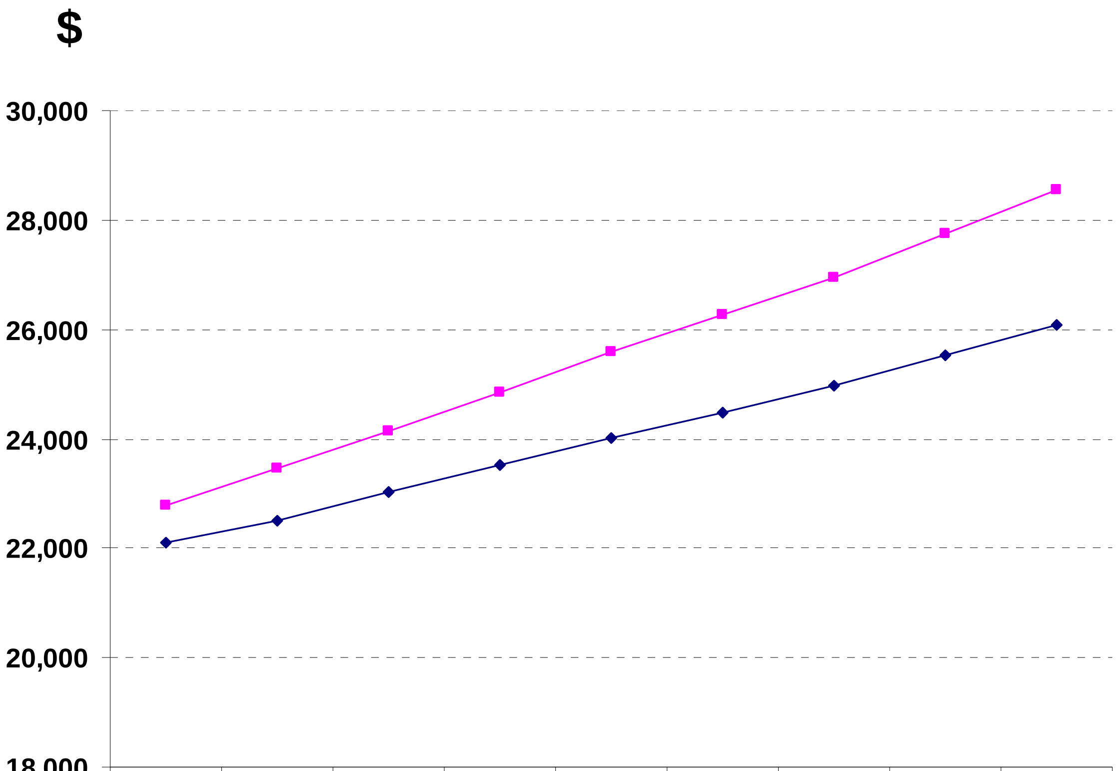
Economic Outlook and Factors Driving Peak Demand and Energy

Growth in electricity demand and consumption is closely correlated with three main factors: 1) Weather, 2) Economics, and 3) Demographics. Economic and demographic changes can affect the characteristics of electrical demand in the medium to the long-run. Weather, on the other hand, drives most of the variation in electric demand in the short-run. Thus, since weather also affects the variation in the electric demand in the long-run, long-term forecasting using historic average weather profiles to indicate the future variation in weather.

The regional economic outlook for Texas is projected to outperform the U.S. as a whole. Three of its major metros, Houston, Dallas, and Austin, which are among the top 50 in the U.S., are leading the South. Employment growth in Texas, shows a stronger performance for the Dallas-Forth Worth area, and the Austin-San Antonio areas. The Houston area is expanding but is expected to lose some momentum due to the energy industry.

Some of the indicators that were used in the forecast are economic and demographic drivers such as real per capita personal income, population, employment in the financial services, non-farm employment, and total employed. These are presented in the figures below.

Moody's economic outlook is more optimistic than last year's. The total employed figures, which include farm indicators, are an exception in ERCOT's territory as the some of the zones like the west and far west include large territories of rural land where they have not experienced the growth as much.

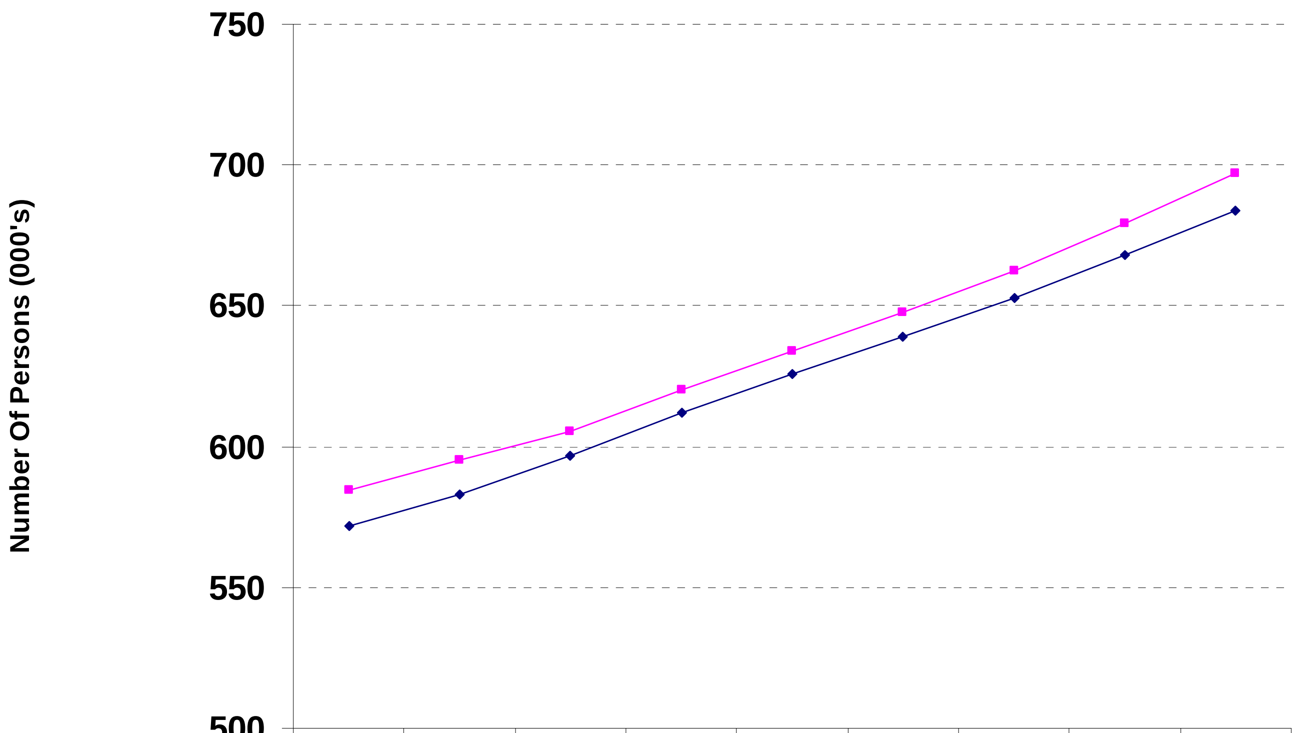


	2007	2008	2009	2010	2011	2012	2013	2014	2015
2006 Forecast	22,092	22,518	23,035	23,533	24,015	24,480	24,978	25,515	26,078
2007 Forecast	22,776	23,445	24,132	24,847	25,600	26,282	26,961	27,736	28,547

Figure 7- Real Personal Per-Capita Income

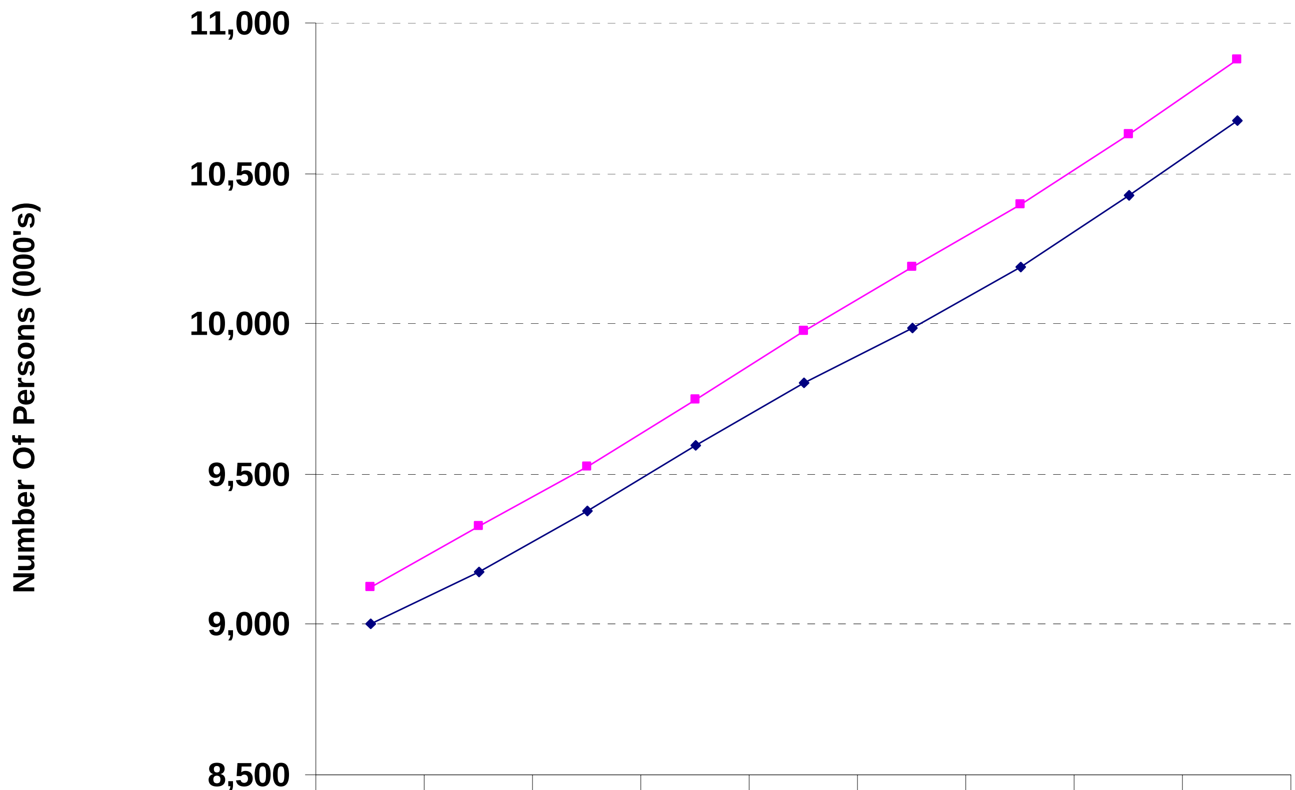
	2007	2008	2009	2010	2011	2012	2013	2014	2015
2006 Forecast	20,986	21,327	21,713	22,137	22,566	22,981	23,398	23,837	24,282
2007 Forecast	21,015	21,372	21,759	22,175	22,602	23,011	23,425	23,862	24,309

Figure 8 – Population in the ERCOT Territory



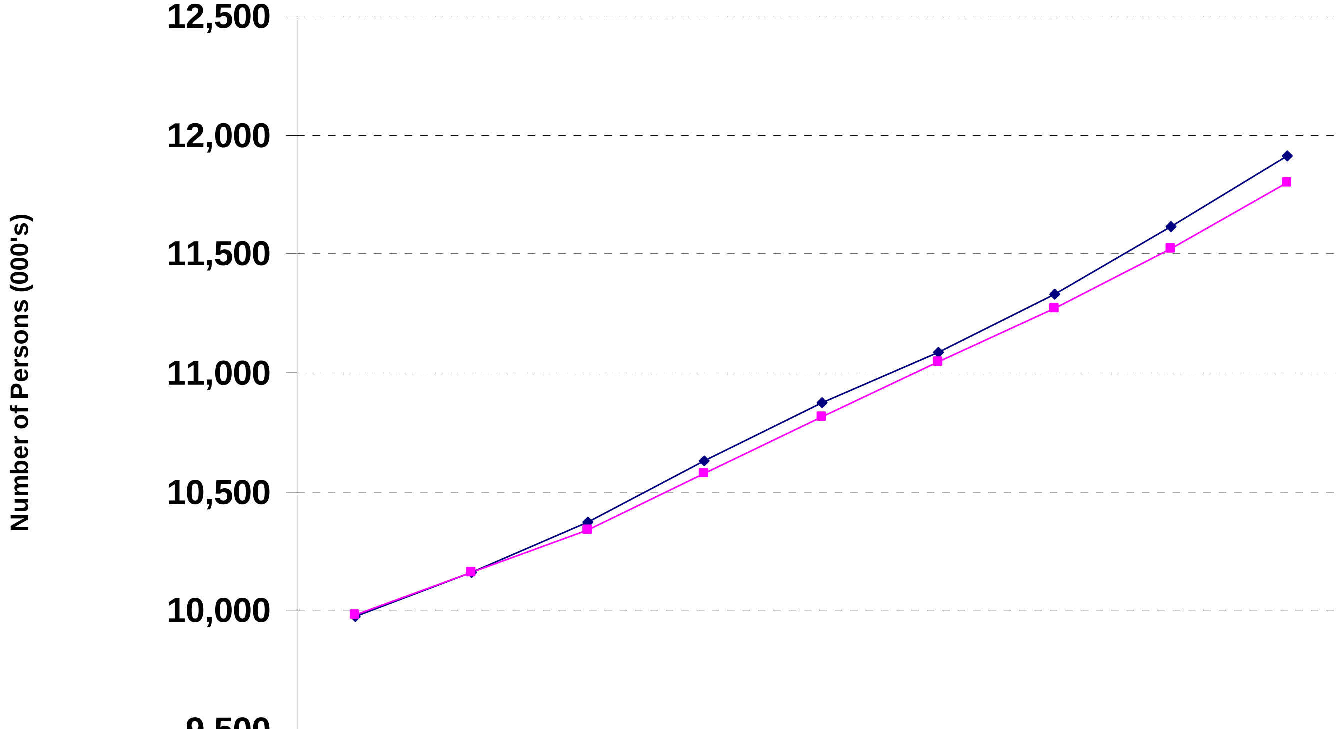
	2007	2008	2009	2010	2011	2012	2013	2014	2015
2006 Forecast	572	583	597	612	626	639	653	668	684
2007 Forecast	584	595	605	620	634	648	662	679	697

Figure 9 – Employment in the Financial Services



	2007	2008	2009	2010	2011	2012	2013	2014	2015
2006 Forecast	9,003	9,173	9,377	9,597	9,801	9,983	10,191	10,429	10,675
2007 Forecast	9,123	9,328	9,526	9,749	9,975	10,187	10,399	10,628	10,877

Figure 10 – Total Non-Farm Employment



	2007	2008	2009	2010	2011	2012	2013	2014	2015
2006 Forecast	9,975	10,158	10,370	10,633	10,875	11,089	11,333	11,617	11,911
2007 Forecast	9,985	10,160	10,338	10,577	10,818	11,044	11,271	11,522	11,797

Figure 11 – Total Persons Employed

ERCOT's Peak Demand and Energy Uncertainty

One measure of the uncertainty associated with extreme weather impacts on the peak demands can be obtained by using a more extreme weather profile to obtain the forecasts. ERCOT developed weather profiles that rank at the 90th percentiles of all the temperatures in its hourly temperature database and did the same to develop with the 10th percentile of all temperatures. Strictly speaking these are not confidence bands in the statistical sense, but common use has been to use this term to refer to the results. A more appropriate term would be to use scenarios associated with the 90th percentile temperature distribution or 90th percentile scenario forecasts. ERCOT has also, in the past, run Monte Carlo simulation to assess the extreme temperatures on the peak demands.

For the 2007 LTFM the 90% Confidence Bands were developed and are depicted in the figures below. The high forecast for 2007 is 5.35% higher than the 2007 forecast with an average weather profile.

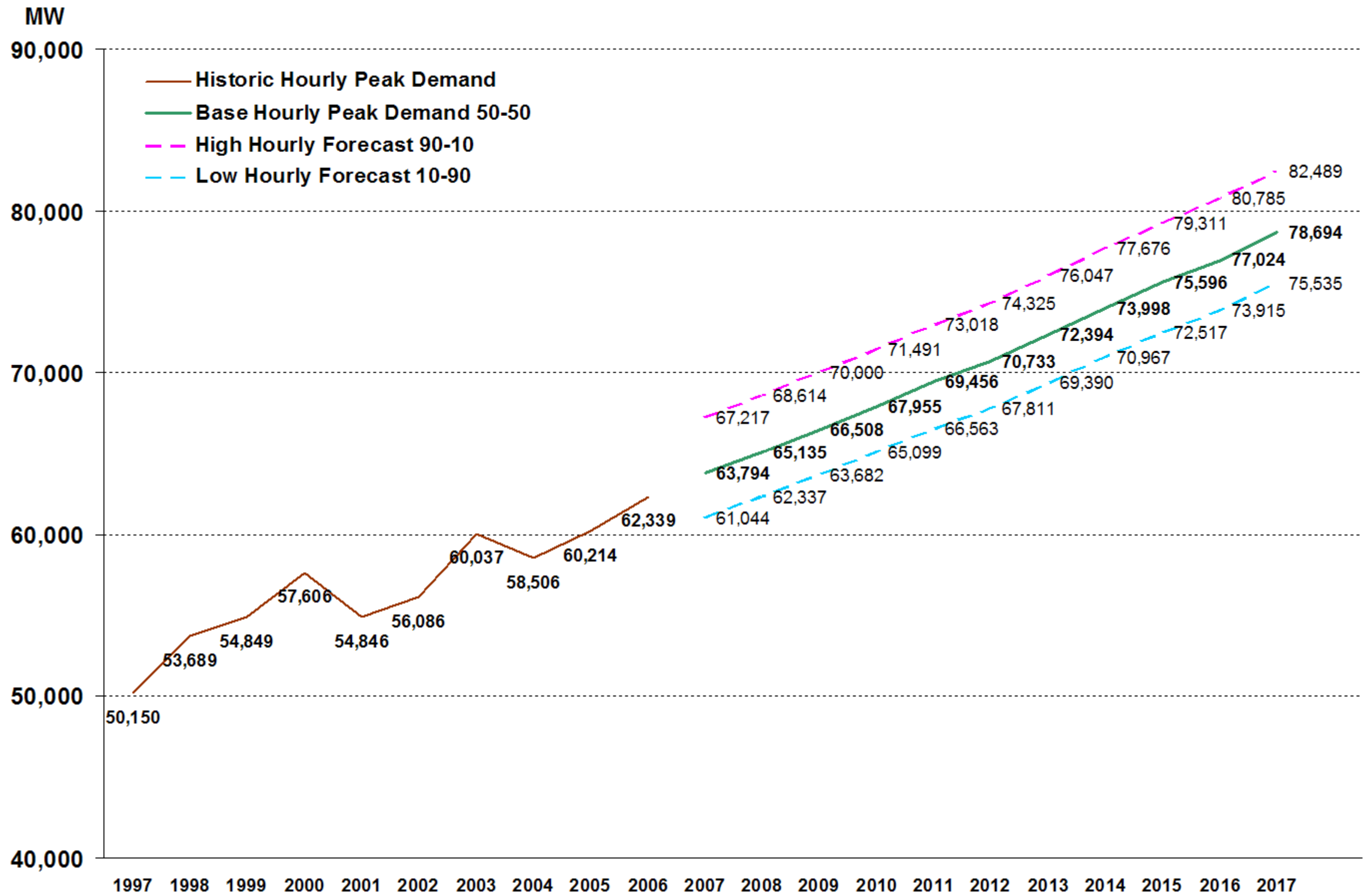


Figure 12 – Historic and Forecast Hourly Peak Demand

Differences with Last Year's Forecast

In the long-term, this year's forecast is very similar to last year's forecast for the same period. In general, the forecast is slightly higher due to a stronger economic outlook for Texas. There was a short run adjustment made for 2007 due to the improved economic forecasts and the forecasting models were recalibrated based on having an additional year of actual data. The figure below shows the difference between the two forecasts from 2007 to 2015.

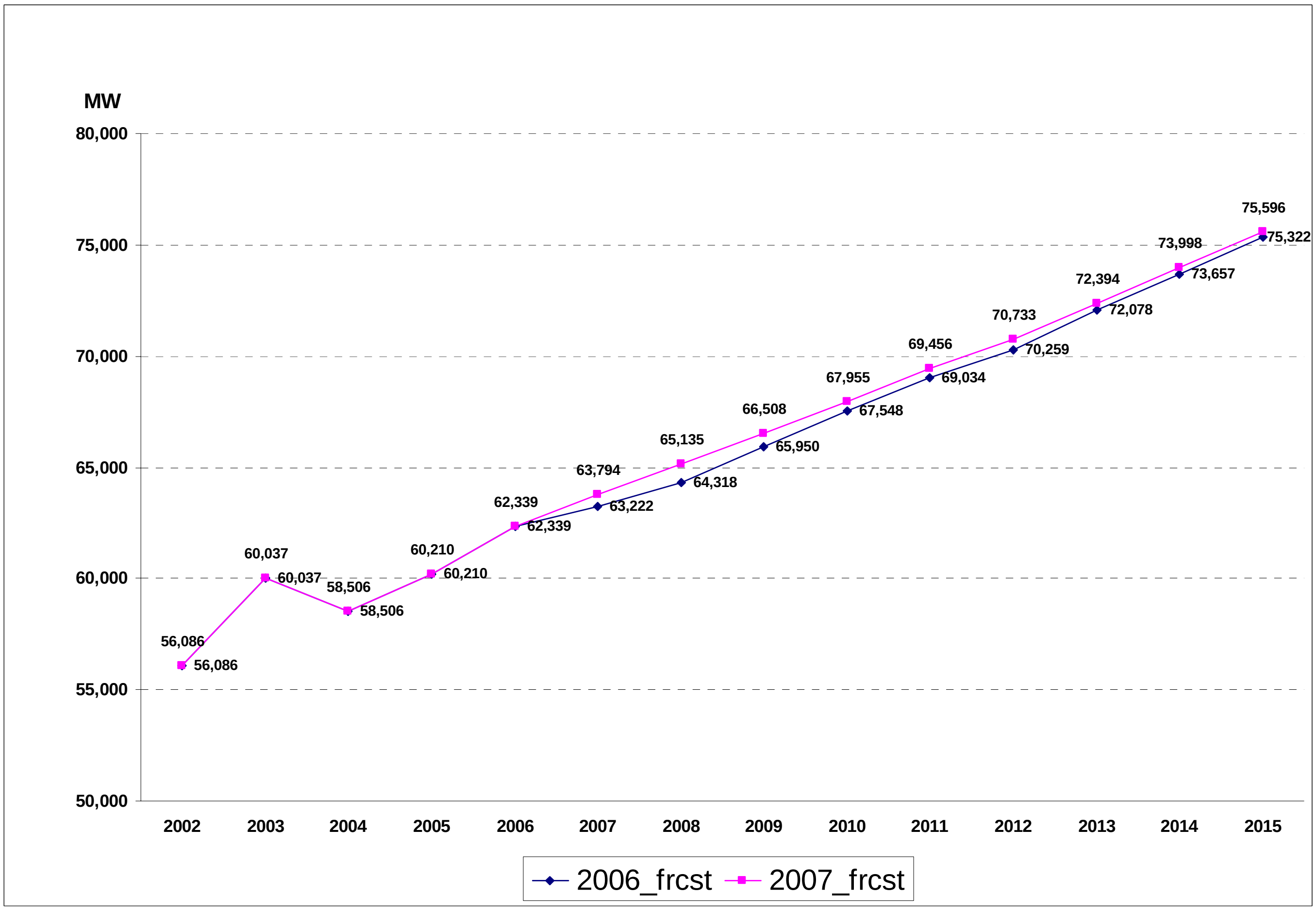


Figure 13- Comparison of 2006 and 2007 Forecast

ERCOT's Load Shape Forecast

The process used to develop ERCOT's peak demand forecast produces an hourly Load Shape for each weather zone. The hourly load peak demand forecast also contributes the system peak demands that are used in the resource adequacy assessment, NERC summer and Long-Term assessments, and other reports. The 2007 Long-Term System Hourly Load forecast over the next five years (2007-2012) and the forecast (fitted) results are shown in the figure below.

Figures 14 and 15 depict the forecast load shapes for 2007 to 2012. Each of these load shapes is derived using an average weather profile. Because of this, the load shapes are basically the same for each forecast year. The upward trend comes from the economic forecasts that drive the energy consumption forecasts. Figure 16 shows one 24 hour day for August 2007.

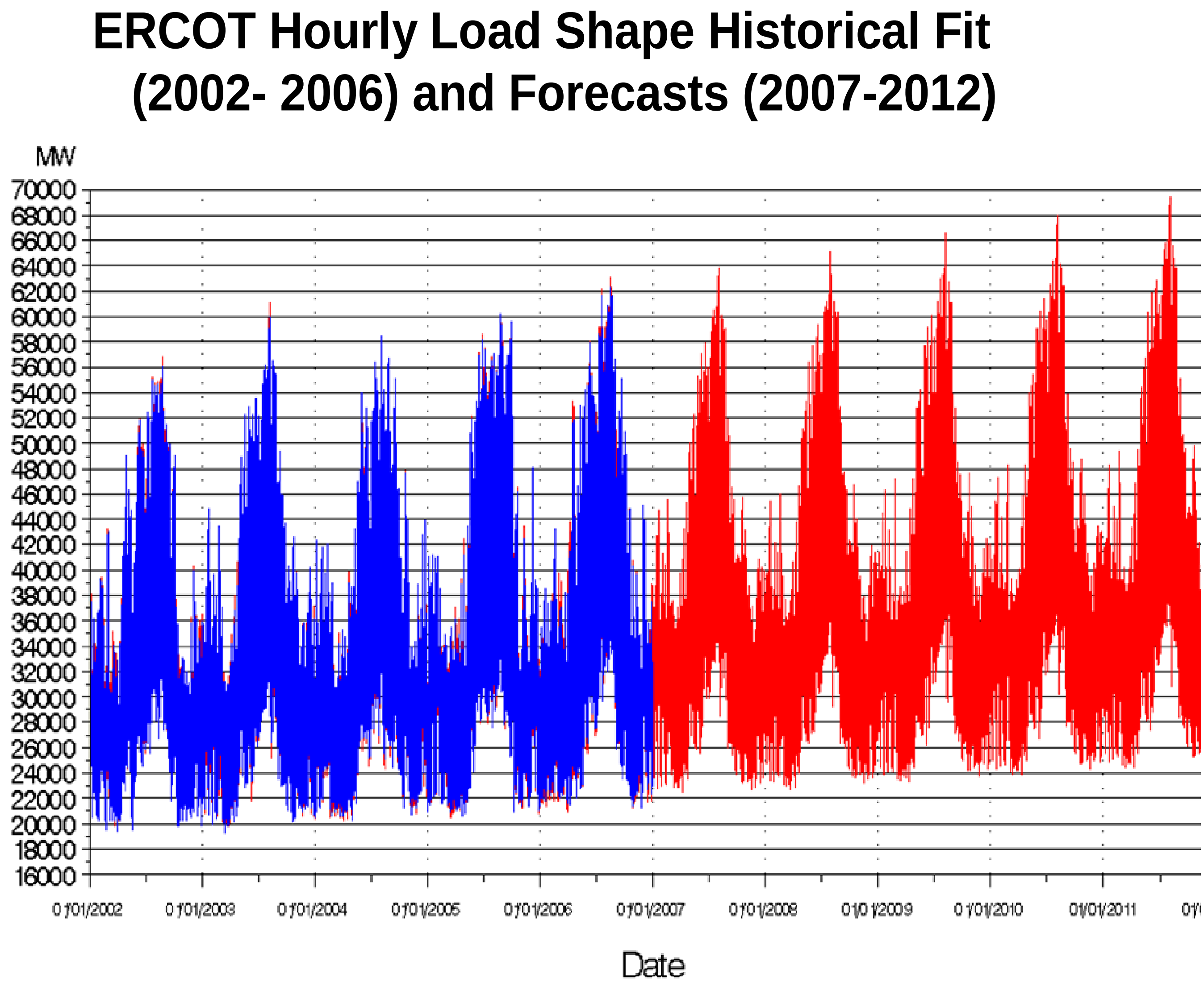


Figure 14 – Hourly Load Forecast including Historic Fit

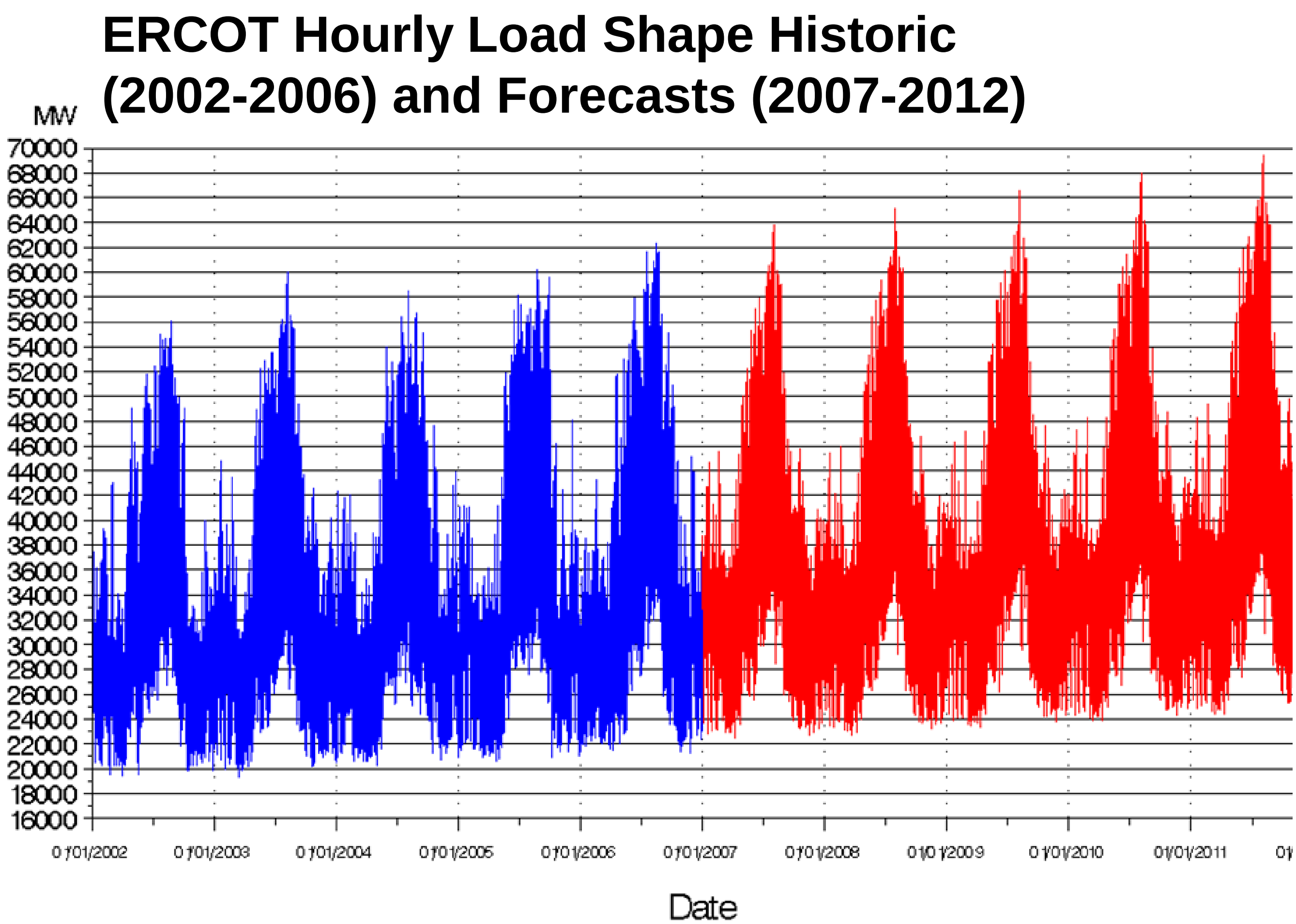


Figure 15 – Hourly Load Forecast and Actuals

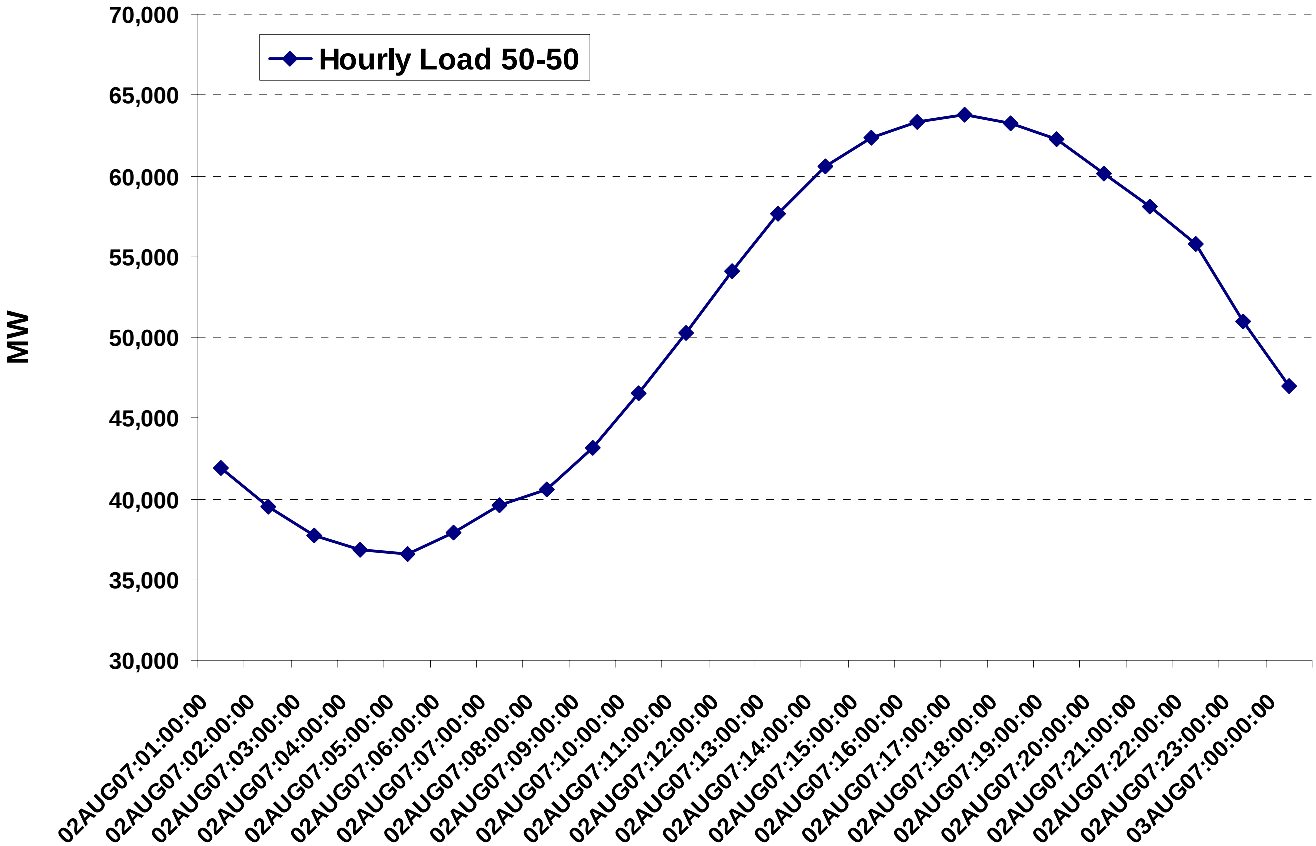


Figure 16 – Hourly Peak Loads for August 2, 2007

ERCOT's Peak Demand and Energy Forecast by Weather Zone

There are eight defined weather zones at ERCOT. The weather zones are: 1) North, 2) North Central, 3) East, 4) Far West, 5) West, 6) South Central, 7) Coastal, and 8) South. The largest MSAs are located in the North Central, South Central and Coastal zones. The Dallas/FW is in the North-Central, Austin and the San Antonio areas are contained within the South-Central and Houston is in the Coastal zone. These three areas have an optimistic outlook and are expanding rapidly. The Houston area is growing at a fast pace. However, its future outlook shows a potential slowdown due to the energy industry. Thus, the forecasts for these major zones show a stable and strong growth. The forecasts for the smaller zones show an average or below average trend in growth.

The annual forecasts data by weather zone are included in Table 2 of appendix 2.

APPENDIX 1: PEAK DEMANDS AND ENERGY CONSUMPTION DATA

A summary of the 2007 Long-Term Forecast Model (LTFM) results is condensed below. This table includes the forecast energy, the MWh historic values, the coincident and zonal peaks, the diversity factors, the coincident factors, and the load factors and the diversity factors in % terms. For reference, historic data for 2002-2006 is included. The MW peak is a coincident peak and the zonal peak refers to the aggregate of individual non-coincident peaks. The Forecast Energy MWh column, from 2002-2006, contains the fitted backcast values for that period and forecasts from 2007 to 2025. The MWh_Hist contains the historic energy consumption for 2002-2006. The following quantities in the table below can be defined as follows (numbers are rounded):

Load Factor: (energy/(peak*number of hours)
Diversity Factor: (Non-Coincident Peak – Coincident Peak)
Diversity Percent: (Diversity Factor/Coincident Peak)
Coincident Factor: (1-Diversity Percent)

Year	Forecast Energy MWh	MWh Hist	MW Peak	Zonal Peak	Diversity Factor	Coincident Factor	Diversity %	Load Factor
2002	281,930,582	280,772,959	56,086	57,233	1,146	98.00%	2.04%	57.15%
2003	284,207,211	284,983,916	60,037	60,376	339	99.44%	0.56%	54.19%
2004	287,569,872	289,140,984	58,506	59,316	810	98.63%	1.38%	56.42%
2005	300,553,020	299,253,971	60,214	61,364	1,150	98.13%	1.91%	56.73%
2006	305,552,884	305,687,145	62,339	63,312	974	98.46%	1.56%	55.98%
2007	313,027,658		63,794	64,831	1,037	98.40%	1.63%	56.01%
2008	319,688,988		65,135	66,111	976	98.52%	1.50%	56.03%
2009	325,408,664		66,508	67,577	1,069	98.42%	1.61%	55.85%
2010	332,578,515		67,955	69,038	1,084	98.43%	1.59%	55.87%
2011	340,089,254		69,456	70,555	1,099	98.44%	1.58%	55.90%
2012	347,087,436		70,733	71,843	1,109	98.46%	1.57%	56.02%
2013	354,122,426		72,394	73,436	1,041	98.58%	1.44%	55.84%
2014	361,232,831		73,998	75,084	1,086	98.55%	1.47%	55.73%
2015	369,322,241		75,596	76,752	1,156	98.49%	1.53%	55.77%
2016	377,330,064		77,024	78,194	1,170	98.50%	1.52%	55.92%
2017	384,606,172		78,694	79,883	1,190	98.51%	1.51%	55.79%
2018	391,597,067		80,161	81,354	1,194	98.53%	1.49%	55.77%
2019	398,301,224		81,622	82,750	1,128	98.64%	1.38%	55.71%
2020	404,587,586		82,871	84,043	1,172	98.61%	1.41%	55.73%
2021	411,162,342		84,363	85,608	1,245	98.55%	1.48%	55.64%
2022	417,594,564		85,681	86,941	1,260	98.55%	1.47%	55.64%
2023	423,892,847		87,015	88,290	1,275	98.56%	1.47%	55.61%
2024	430,373,659		88,180	89,453	1,274	98.58%	1.44%	55.72%
2025	436,287,512		89,883	91,128	1,245	98.63%	1.38%	55.41%

Table 1 – Forecast Results of the 2007 Long-Term Forecast Model

APPENDIX 2: WEATHER ZONE LOAD DATA

Table 2 - Historic and Forecast Yearly Coincident Peak Demands by Weather Zones (MW)

Year	NORTH	NORTH CENTRAL	EAST	FAR WEST	WEST	SOUTH CENTRAL	COAST	SOUTH	SYSTEM LOAD
2002	1,904	20,527	2,175	1,830	1,595	9,492	14,578	3,985	56,086
2003	2,070	22,303	2,319	1,805	1,675	10,016	15,823	4,025	60,037
2004	2,047	20,749	2,265	1,658	1,562	9,619	16,611	3,996	58,506
2005	2,080	21,975	2,351	1,661	1,542	10,162	16,282	4,159	60,214
2006	2,361	22,687	2,432	1,598	1,612	10,718	16,739	4,191	62,339
2007	2,086	23,782	2,251	1,412	1,638	11,329	17,174	4,123	63,794
2008	2,117	24,059	2,363	1,415	1,683	11,708	17,631	4,158	65,135
2009	2,145	24,472	2,323	1,429	1,725	12,075	18,112	4,227	66,508
2010	2,183	24,914	2,353	1,435	1,770	12,475	18,554	4,271	67,955
2011	2,229	25,365	2,382	1,441	1,820	12,901	19,002	4,317	69,456
2012	2,263	25,743	2,402	1,442	1,863	13,292	19,377	4,351	70,733
2013	2,325	26,267	2,517	1,448	1,914	13,725	19,794	4,405	72,394
2014	2,377	26,788	2,462	1,509	1,964	14,111	20,312	4,474	73,998
2015	2,447	27,360	2,484	1,461	2,022	14,570	20,727	4,525	75,596

APPENDIX 3: METHODOLOGY

A Modified Approach to Long-Term Load And Energy Forecasting: Its Uses In An ISO's Environment For Resource Adequacy And Transmission Planning

Introduction

The main focus of this paper is the benefits of a modified approach to long-term demand and energy forecasting model in an ISO's setting. The forecasts that were produced by a regression model are input into several planning processes that are important in the long-term planning of an electrical grid. The development of this forecasting methodology was designed to address the needs for forecasts in several processes. The load forecasting methodology that was adopted is discussed and its results are outlined. The objective of this methodology is to determine a long-term view of the peak demands that ERCOT (total load served in the ERCOT region including exports across DC ties and excluding private use network loads) can expect to face, in order to secure sufficient resources in the next five to ten years. The discussion covers the success experienced in using this methodology and details of the process involved in producing the forecasts. More specifically, this paper details:

- A methodology developed specifically for ERCOT to meet its specific needs.
- How the methodology chosen has been used to successfully meet ERCOT's planning objectives.

Why it is needed

The development of a long-term trend outlook uses a regression model that forecasts peak demands that are most likely to occur under normal weather conditions to determine the approximate timing for scheduling the building of transmission lines to balance the supply and demand for electric power in the ERCOT electrical grid. The load forecast is an input to the reserve margin calculation. As such, the load forecast is a key component necessary for meeting this objective, which is used to ensure a balanced system.

A resource adequacy assessment begins with the calculation of a reserve margin as,

Reserve margin= (Resources – Firm Load Forecast)/Firm Load Forecast

This calculation is the foundation of the process for determining the adequacy of the system. The review of resource adequacy is an annual process that ensures that enough resources will be available to meet demand in the medium-to long-term time frame.

The forecast is also used in the medium-range planning of resources by the outage coordinators to schedule plant outages for the next year.

Another aspect of system adequacy, where the load forecast plays an important role, entails performing a load sensitivity assessment. This assessment is related to the risk associated with the volatility of the load due to weather. The 90% approximate forecast limits due to the volatility associated with forecasting the load, using temperatures at the 90th percentile of the distribution, are calculated for the next ten to fifteen years to assess the risks of extreme weather volatility on the peak demands. These load volatility estimates are an input into the loss-of-load-probability studies (LOLP), which are used to determine the target reserve margin.

Reviews of the reserve margin to ensure its adequacy are performed every few years through a LOLP study. In this study, expected load, load forecast error, the load volatility due to weather, generation fleet, maintenance schedules, and unit forced outage rates are input into a unit commitment and dispatch model in order to simulate the interrelationships between these variables over a number of replications. This simulation yields an expected un-served energy value. Then, the target reserve margin is obtained by finding the minimum point of the intersection where the LOLP is the ERCOT/NERC standard of one event every ten years.

Load volatility estimates derived from the load forecast are also used by NERC in the summer and winter reliability assessments. These load forecasts feed into the reporting requirements of FERC 714.

The long-term hourly load forecast by weather zones also serves an important function in performing economic analyses. It is an input to the UPLAN software which determines whether or not to undertake transmission projects.

As described above, the load forecast is a major input to several planning processes. The long-term forecast can affect the adequacy of the system grid. Some of the consequences of load forecast errors and their impact on system adequacy can be:

- Building excessive additional generation capacity and/or transmission facilities
- Inadequate levels of resources and generation leading to blackouts and price spikes
- Sending incorrect signals to the market regarding the value of capacity and energy

Finally, the energy consumption forecast provides the means to determine the annual $/MWh ERCOT fee for the annual budget review, conducted by the Texas PUC.

Availability of methods

There are a wide variety of methods that can be used to forecast system peak and energy consumption. Such methods range from simple trending methods to more complex ones such as end-use forecasting or hybrids end use and econometric techniques, sophisticated Box-Jenkins Transfer function (Dynamic Regression) models and now neural network models that can be adapted to produce long-term forecasts

For ERCOT, data requirements were a major determinant of which method was feasible and appropriate to implement. There were specific requirements to be met in terms of the end product. The following describes the specific nature of these data needs.

Forecast Level of Detail

An hourly forecasted load shape by weather zones for the next five to ten years was needed as an input into UPLAN for economic analysis of transmission projects. The hourly loads from the load shape, combined with the results of a monthly energy forecast, were considered a feasible way to produce a system peak forecast for each year in the five-to-ten-year horizon. The system peaks and energy consumption forecasts were thought to be a high priority for this important process as these forecasts could as well be used as inputs into the resource adequacy process.

Load and Weather Data level of Detail

ERCOT Staff decided to produce long-term forecasts for eight major areas in Texas where weather data was available and coincided with the available data appropriate for load analysis. Thus, from ERCOT's standpoint, weather zones were the logical choice. In addition, these zones also coincided with the major areas of interest for the analysis of transmission projects. In summary, the total load by weather zone was chosen as meeting the objective of the forecast needs. These forecasts then could be aggregated to a system level.

Economic, Demographic and Price Data Level of Detail

Besides hourly load, ERCOT also secures weather data, economic and demographic data from outside providers. In regard to prices, which are considered an important driver for inclusion in a demand equation, it is not clear as to whether the wholesale prices that ERCOT collects are really the most relevant for a forecasting application, in terms of being the prices ultimately faced by the consumer. Since the wholesale prices are collected on an hourly basis, and retail prices are better reflected by an average over a longer time period, such as a month, wholesale hourly prices do not capture the correlation with the MWh consumption correctly. Several attempts to include market clearing prices of energy (MCPE s) in the forecasting models were made but were unsuccessful. The models obtained showed price to be insignificant or to indicate a nonsensical relationship regarding the direction of the effect of price (wrong sign on the coefficient) and thus should not be included in a long-term demand equation. To make matters more challenging in this respect, an objective and credible forecast of these prices would represent a major accomplishment in itself. Inclusion of a price variable in the forecasting models could potentially provide a means to calculate an unbiased and credible forecast of the price effect on the long-term load response.

Method Selection

There is no single best forecasting method. The choice of a forecasting method in this case was based on the specific circumstances of the situation being faced. Given the requirements at the time, in terms of available data, the capabilities needed of any chosen method, and the intended use of the resulting forecasts, a regression with autocorrelated errors was deemed as the most appropriate choice available to meet ERCOT's objectives. This methodology is unique in that it directly and successfully forecasts an hourly load shape using a regression model estimated by seasons. This methodology could potentially be applied to other entities facing similar requirements.

Forecast Process --- General Description

The forecast process starts with the development of regression equations from historical data for demand peaks and energy. These use the following input drivers:

Trend Variables
- Population
- Income
- Economic

Calendar Variables
- Seasonal Variation
- Daily Variation
- Weekly Variation
- Holidays

Weather profiles from actual data that use an average representation of weather not prediction of weather
- Temperature
- Humidity
- Cooling Degree Days (CDD)
- Heating Degree Days (HDD)

The results are forecasts for energy and peak.

The data used to prepare the forecast came from the following sources:

1. Economic Data
   - Economic data obtained from Economy.com
   - Data includes economic and demographic data (such as income, employment, housing permits, GDP, population and migration patterns) for Texas at the state, county, metropolitan statistical areas (MSAs). Some of these data is also available at the national level

2. Weather Data
   - Ten years of weather data obtained from Weather Bank for 20 weather stations
   - The data is first weighted by individual weather stations using ERCOT's standard factor, and then for the total system using weights proportional to the load in each weather zone

3. Load Data
   - Settlement load data available on an hourly basis since July 31, 2001
   - Prior to 2001, we have Transmission and Distribution Service Providers (TDSP) hourly data

The weather data is used in the development of weather normalized profiles by weather zone and is accomplished by calculating the normalized temperature profile by weather zone. The weather profiles use the rank-average method which involves the following steps:

1) Rank the hourly temperatures for each year for each weather zone from highest to lowest
2) Determine the median temperature from all years for every hour
3) Calculate the sum of the absolute values of the difference of the median and the hourly temperatures for all hourly temperatures in each year
4) Determine the year with the minimum summed value and select this year as the typical year profile
5) Use this year's profile to re-sort the median temperatures

A major issue in the preparation of the long-term forecast relates to the variable selection process. The process in this case generally entails performing the following analyses with the following considerations:

- Multiple regression analysis was used to develop the forecasting equations
- Initial selection of variables came from a stepwise procedure to determine those that were the most statistically significant
- A subset of those variables was chosen on the basis of empirical results and judgment
- Variables selected for inclusion had to meet the following: 1) justifiable on a logical basis , 2) historically measurable and 3) must have an available forecast
- Ordinary least squares techniques, with some of the models including autoregressive error terms, were used to calculate the appropriate coefficients on each variable and to choose the best equations

Load shape and Energy forecasts were developed from monthly energy and hourly load shape equations for each season of the following form:

- The general formulation of the energy equations include:

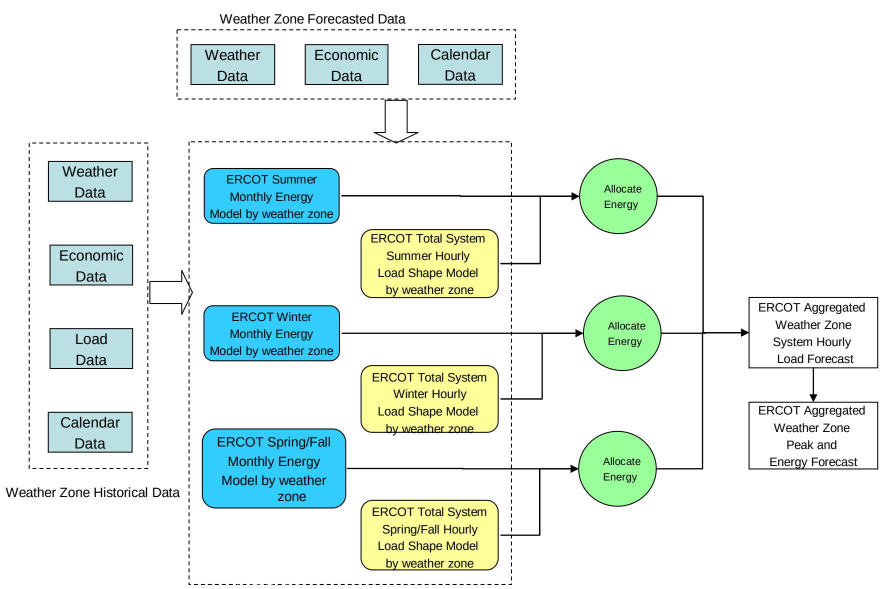
  Energy Month i = f {CDD, HDD, Income, Population, Employment, Monthly Indicators}

- The general formulation of the load shape equations include selected variables from some of the following:

  Load hour i =f {Max Temps, Lagged Temps, Heat Index, Non-Linear Temp Components (square and cube), Temp Gains (diff between daily high and low temps), Temp Build-up, Dew Point, Month*Temp Interactions, CDD, HDD, Hour of Day Indicators, Weekday/Weekend, Holidays}

<u>Putting it all together</u>

Weather Zone Forecasting Process



The Weather Zone forecasting process flow is as follows:

1. Obtain weather and economic variables by weather zone (historic and forecast)
2. Develop regression equations by weather zone describing the historic actual:
--Monthly Energy
* Using a different equation for each season
-- Hourly Load Shape
* Using a different equation for each season
3. Incorporate forecasted values of economic and normalized temperatures for 2006-2020 by weather zone into monthly energy equation to produce forecasted monthly energy
4. Incorporate normalized temperatures for 2006-2020 by weather zone into monthly load shape equation to produce forecasted load shape
5. Produce hourly demand forecast by weather zone by fitting forecasted monthly energy under projected hourly load shape

Hourly Forecast

The calculation of an hourly forecast is a result of the process described above and yields the following results:

- The forecasted hourly shape from the load shape equations is scaled to produce the final hourly forecast
  - Each hour's load is scaled so that the amount of energy under the load shape for a month is equal to the amount of energy projected for that month by the energy forecast from the energy equations
  - The percent of a month's energy that is contained in each hour from the load shape equation is maintained
- The peak forecast is the highest hourly load from this final hourly forecast

Mathematical/statistical rigor

(A) Derivation:

The mathematical/statistical intricacies of the models are presented below. The peak demand forecasts are obtained by combining the results of two models: an hourly model that forecasts the load shape and a monthly energy forecast which includes economic and demographic variables to determine the long-term trend. The hourly load shape model is of the following form:

$$Y_t = \alpha_o + \sum_{i=1}^{23} \beta_i HR_{i,t} + \sum_{i=1}^{n} \gamma_i W_{i,t-s} + \sum_{i=1}^{n} \Omega_i DT_{i,t} + \sum_{i=1}^{n} \Theta_i WI_{i,t} + \sum_{i=1}^{n} \delta_i SV_{i,t} + \sum_{i=1}^{n} \nu_i \mathrm{E}_{i,t} + \frac{\varepsilon_{i,t}}{\left(1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p\right)}$$

Where:

- Y_t is the hourly load (MW)
- $HR_{i,t}$ are hourly indicator variables
- $W_{i,t-s}$ are weather variables and their lags
- $DT_{i,t}$ are day type variables
- $WI_{i,t}$ are weather interaction variables
- $SV_{i,t}$ are sunlight variables
- $E_{i,t}$ are special events variables
- $\varepsilon_{i,t}$ is a random error term
- $\Phi's$ are autocorrelation terms specified with a lag (backshift) operator, $L^s = \mathrm{X}_{t-s}$

This model specified in mathematical form can be generalized as follows:

$$Y_t = \beta_o + \sum_{i=1} \beta_\kappa \mathrm{X}_{\kappa,t} + \frac{\varepsilon_t}{\Phi(L)}$$

Where:

$\beta_o, \beta_{1,} \ldots, \beta_\kappa$ = coefficients to be estimated

$X_{K,t} = {}_K$ regressor variables, K=l, …, m

ε_t = a random error term

Φ (L) = an autoregressive structure of order ρ where ρ = 24 or an AR(ρ) process

" $= \left(1 - \Phi_1 L - \Phi_2 L^2 - \Phi_3 L^3 - \ldots - \Phi_p L^p\right)$

Φ_j = autoregressive coefficients

L^j = Lag operator, $L^j = X_{t-j}$

Thus, the model to be estimated can be derived as follows:

(1) $$\Phi(L)Y_t = \Phi(L)\beta_o + \Phi(L)\sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t} + \varepsilon_t$$

Where the constant term $\alpha_o = \Phi(L)\beta_o$.

Expanding the expression on the right hand side,

$$\Phi(L)\sum_{\kappa=1}^{m}\beta_{\kappa,t} = \left(1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p\right) - \left(\beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_m X_{mt}\right)$$ and gathering common terms together we obtain

$$\Phi(L)\sum_{\kappa=1}^{m}\beta_{\kappa,t} = \sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t} - \left[\ \Phi_1 + \Phi_2 + \ldots + \Phi_p\ \right] \bullet \sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t-p}$$

Or more succinctly,

(2) $$\Phi(L)\sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t} - \sum_{j=1}^{p}\Phi_j \sum_{\kappa=1}^{m}\beta_{\kappa,t-j}.$$

The expression on the left hand side of the equation is

$$\Phi(L)Y_t = \left(1 - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p\right)Y_t$$
$$\Phi(L)Y_t = Y_t - \Phi_1 Y_{t-1} - \Phi_2 Y_{t-2} - \ldots - \Phi_p Y_{t-p}$$

Or more compactly stated,

(3) $$\Phi(L)Y_t = Y_t - \sum_{j=1}^{p}\Phi_j Y_{t-j}$$

Substituting (2) and (3) into (1) we get,

$$Y_t - \sum_{j=1}^{p}\Phi_j Y_{t-j} = \gamma_o + \sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t} - \sum_{j=1}^{p}\Phi_j \sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t-j} + \varepsilon_t$$

or

$$Y_t = \gamma_o + \sum_{j=1}^{p}\Phi_j Y_{t-j} + \sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t} - \sum_{j=1}^{p}\sum_{\kappa=1}^{m}\beta_\kappa X_{\kappa,t-j} + \varepsilon_t$$

Where

$$\varepsilon_t = \mu_t - \sum_{j=1}^{p}\Phi_j L^j \mu_{t-j}.$$

(B) Estimation:

In vector notation [1],

$$y_t = x_t'\beta + \mu_t$$

Where $\left(x_t = x_{1t,} x_{2t,} \ldots, x_{Kt} \right)'$

$$\mu_t = \varepsilon_t + \varphi_1 \mu_{t-1} + \varphi_2 \mu_{t-2} + \ldots + \varphi_p \mu_{t-p}$$ [2]

And $\varepsilon_t = N\left(o, \sigma^2\right)$, normally and independently distributed with mean o and variance of σ^2

y_t	=	dependent values
x_t'	=	a column vector of regressor variables
β	=	a column vector of structural parameters

The autoregressive parameter vector, $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_p)'$ and its variance covariance matrix:

$$\mu = \left(\mu_{1,} \mu_{2,\ldots,} \mu_p\right)' = \Sigma$$
$$\mathrm{E}\left(\mu \mu'\right) = \Sigma \sigma^2 U$$

Since the stepwise-like procedure BACKSTEP is specified for testing the statistical significance of the φ's, the TOEPLITZ matrix is used, with the (i,j)th

element $\gamma_{|i-j|}$ is equal to $R\hat{\varphi} = r$

Where r = $\left(r_1, r_2, \ldots, r_p\right)'$ and r_i is the lag i sample autocorrelation. The matrix [R, r] is treated as sum-of-squares cross products matrix coming from a simple regression using N-K observations, where K = number of estimated parameters.

This method of estimation is known as the Yule-Walker (YW) method. It alternates the estimation of β using generalized least squares (GLS) with the estimation of the φ's using the YW equations applied to the sample autocorrelation function (SA).

The steps are:

1) Form OLS estimates of β.
2) Estimate φ from the SAC function of the OLS residuals using the YW equations.
3) Estimate U from the estimate of φ and Σ from U and the OLS estimate of σ^2.

[1] This material comes from the SAS Autoreg Procedure in the ETS manual.

[2] SAS parametization computes the signs of the autoregressive parameters reversed from what is presented in most of the literature. The parametization shown here is in agreement with most of the literature.

The second model forecasts the long-term trends in energy consumption (MWh) utilizing economic, demographic, weather, and season variables. The form of the model is as follows:

$$Y_t = \beta_o + \sum_{i=1}^{n} \gamma_i CDD_{n,t} + \sum_{i=1}^{n} \Theta_i HDD_{n,t} + \sum_{i=1}^{s} \delta_i \mathrm{E}_{it} + \sum_{i=1}^{11} \alpha_i S_{it} + U_{it}$$

Where: Y_t = Monthly energy consumption (Mwh)
$CDD_{n,t}$ = Cooling Degree Days (n terms using different basis)
$HDD_{n,t}$ = Heating Degree Days (n terms using different basis)
E_{it} = Economic and Demographic variables
S_{it} = Seasonal indicator variables
μ_t = random error term

This model represented in general form is as follows:

$$Y_t = \beta_o + \sum_{i=1}^{p} \beta_\kappa \mathrm{X}_{\kappa,t} + \mu_t$$

Where,

$\beta_o, \beta_1, \ldots, \beta_p =$ coefficients to be estimated

μ_t = random error term

This energy equation is estimated using ordinary least squares (OLS).

(C) Allocation of Energy Under Load Shape:

Let $Y_{LSi,t} =$ hourly load shape forecast from the first model,

$Y_{\mathrm{E}t} =$ monthly energy forecast from the second model,

Then, the long-term load forecast is obtained as follows:

$$Y_{LSi,t} = Y_{LSj} \cdot \frac{\Sigma Y_{Et}}{\Sigma Y_{LSi}}$$

Where:

Y_{LSj} is the load at hour j, j=1, …,8760

Thus, the annual system peak demand is obtained as,

Y peak = max $\{Y_{LSit;}\, i = 1, \ldots 8760; t = 1, \ldots 12\}$

<u>Conclusions-- Forecast Performance, Results, Findings and Properties</u>

Model validation using actual temperatures in the forecast period – To validate the model, it was estimated with data up to December 2005 and a forecast was produced for January 2006 to December 2006 using the actual temperatures. Shown in Figure 1, is a comparison of the actual system loads, the forecasted system loads, using normal temperatures, and the forecasted system loads using the actual temperatures. A forecast for the summer season only was also produced using the actual temperatures and is also shown in Figure 1. The results were very encouraging as the system peak that occurred on August 17, 2006 was forecasted for the year 2006 with a 0.78% error and 0.45% for just the summer. The forecast performance for the first twenty days of August 2006 is graphed in Figure 2, reflecting the two different allocations of energy under the load shape, first for the entire year 2006 and the other for just the summer season. The results for both approaches are shown in a combination of the major zones into a major group format are given in Chart 1 and 2.

The forecasting model can be used to perform weather impact assessments. Forecasting load volatility using the model with an extreme weather profile – The forecast using a 90th percentile weather profile is also shown in Figure 1. The actual system peak load 62334 MW occurred on August 17, 2006.

There are strengths and weaknesses associated with the process described in this paper. They are:

ERCOT's model strengths

- The methodology is statistical and mathematical in nature, but it still allows for judgment to be incorporated into the results by selecting variables that contribute to the generation of a forecast that passes, not only statistical tests, but common sense criteria.
- This approach was implemented in an automated fashion using macro routines in SAS. With so many models to maintain (8 zones * 3 seasons per zone = 24 models total), it is advantageous to have the ability to make changes and produce normal or extreme weather or any other type of forecasts very quickly.
- The chosen methodology remains consistent in the face of changes in the structural pattern of new incoming data. This is an indication of the robustness of the approach and the model.

ERCOT's model weaknesses

- The initial set-up for the infrastructure for using this approach is time consuming and complex.
- The model was developed from a top-down approach analyzing total ERCOT (system) load. Thus, it does not allow analysis at a more disaggregated level such as focusing at the class level, i.e., residential, business commercial, large industrial customers, etc.

An important aspect associated with any forecasting model is the robustness of the forecasts coming out of the model. Another related consideration is whether these forecasts can be considered reliable enough to lend the model some credibility. In this case, there are forecasts produced with a very similar model for 2005, using the same methodology but, with system load data instead of disaggregated data for weather zones. The model presented here aggregates across zones can be used to obtain the system peak. The results produced by the model for 2005 are very similar in terms of the magnitude of the percent forecast errors. The overall error was between 0 and + 0.5%. This pattern of successful forecasting gives this methodology some credibility and shows its robustness.