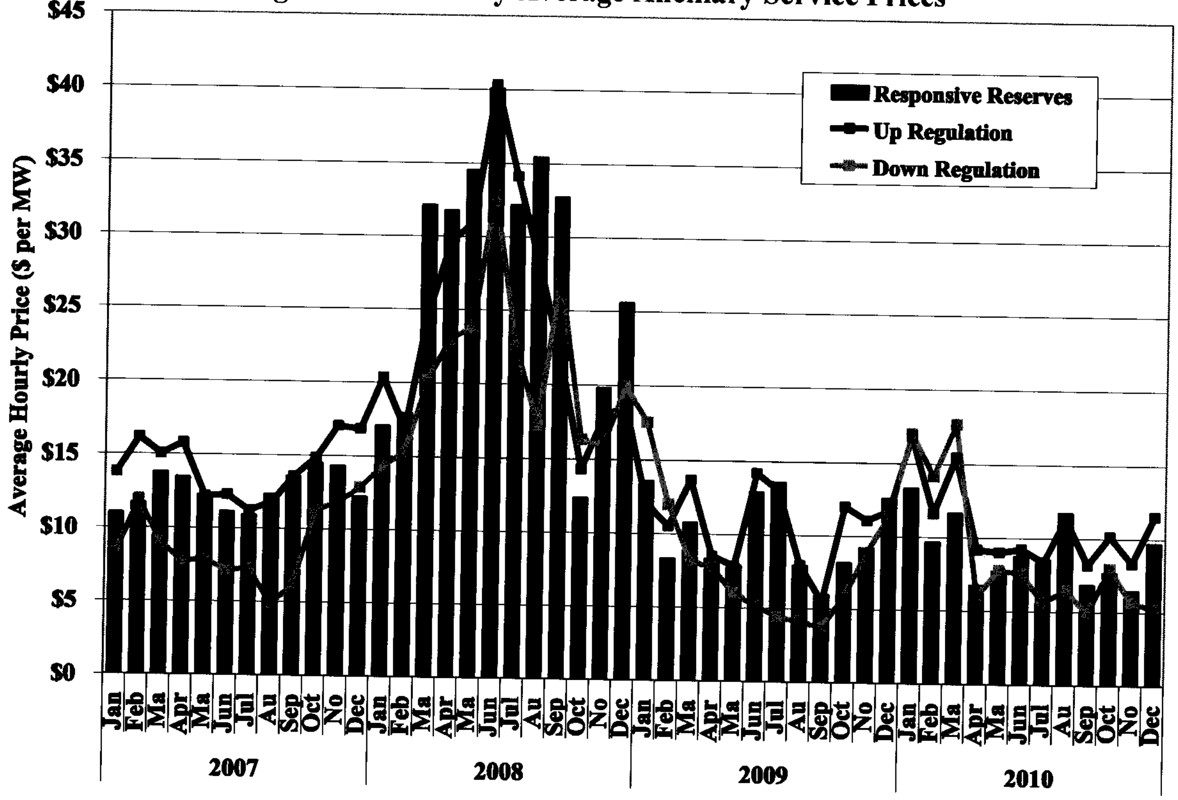
B. Ancillary Services Market

The primary ancillary services are up regulation, down regulation, responsive reserves, and non-spinning reserves. Market participants may self-schedule ancillary services or purchase their required ancillary services through the ERCOT markets. This section reviews the results of the ancillary services markets in 2010.

In general, the purpose of responsive and non-spinning reserves is to protect the system against unforeseen contingencies (*e.g.*, unplanned generator outages, load forecast error, wind forecast error), rather than for meeting normal load fluctuations. ERCOT procures at least 2,300 MW of responsive reserves to ensure adequate protection against the loss of the two largest generating units. Non-spinning reserves are procured as a means for ERCOT to increase the supply of energy in the balancing energy market through supplemental generator commitments. In the zonal market, balancing energy deployments that occur in the 15-minute timeframe and regulation deployments that occur in the 4-second timeframe are the primary means for meeting load fluctuations across and within each 15-minute interval. One significant change under the nodal market is that deployments of energy occur more frequently, typically every five minutes. The more frequent deployment of energy means less regulation capacity is required to meet load fluctuations.

Our first analysis in this section provides a summary of the ancillary services prices over the past four years. Figure 15 shows the monthly average ancillary services prices between 2007 and 2010. Average prices for each ancillary service are weighted by the quantities required in each hour. This figure shows that ancillary service capacity prices in 2010 were similar to those in 2009. Price movements in the ancillary services markets can be primarily attributed to the variations in energy prices that occurred over the same timeframe.

Figure 15: Monthly Average Ancillary Service Prices



Under the zonal market, ancillary services markets are conducted prior to the balancing energy market. This practice requires that participants include their expected value of foregone sales in the balancing energy market in their offers for responsive reserves and regulation. Providers of both responsive reserves and up regulation can incur such opportunity costs if they reduce the output from economic units to make the capability available to provide these services. Likewise, providers of down regulation can incur opportunity costs in real-time if they receive instructions to reduce their output below the most profitable operating level. Further, because generators must be online to provide regulation and responsive reserves, there is an economic risk during low-price periods of operating uneconomically at minimum output levels (or having to operate above minimum output levels if providing down regulation).

Figure 15 shows that average down regulation prices have generally been lower than prices for up regulation service over the last four years, indicating that the opportunity costs were greater for providers of up regulation. Exceptions to this trend were observed in September and October 2008, and during the past two winter seasons. Figure 15 also shows that, on average, the price of up regulation is slightly higher than the price of responsive reserves. This outcome is consistent with expectations. Although a supplier incurs opportunity costs to provide either service, there are additional costs associated with providing up regulation. These additional costs include the costs of frequently changing resource output levels, and the risk of having to produce output when regulating at balancing energy prices that are less than the unit's variable production costs.

One way to evaluate the rationality of prices in the ancillary services markets is to compare the prices for different services to determine whether they exhibit a pattern that is reasonable relative to each other. Table 1 compares the average prices for responsive reserves and non-spinning reserves over the past four years in those hours when ERCOT procured non-spinning reserves.[8]

Table 1 shows that responsive reserves prices are higher on average than non-spinning reserves prices during hours when non-spinning reserves were procured. It is reasonable that responsive reserves prices would generally be higher since responsive reserves are a higher quality product that must be delivered in 10 minutes from on-line resources while non-spinning reserves must be delivered in 30 minutes.

Table 1: Average Hourly Responsive Reserves and Non-Spinning Reserves Prices During Hours When Non-Spinning Reserves Were Procured

	2007	2008	2009	2010
Non-Spin Reserve Price	$6.07	$7.97	$3.08	$4.25
Responsive Reserve Price	$16.74	$36.39	$9.68	$9.09

In contrast to the previous data that show the individual ancillary service capacity prices, Figure 16 shows the monthly total ancillary service costs per MWh of ERCOT load and the average balancing energy price for 2007 through 2010.

[8] ERCOT began procuring non-spinning reserves in every hour beginning in November 2008, primarily to address the increasing uncertainty in net load associated with increasing levels of intermittent generation resources. Non-spinning reserves were purchased in approximately 14 percent of hours in 2007, and increased to 51 percent of the hours in 2008.

Figure 16: Ancillary Service Costs per MWh of Load

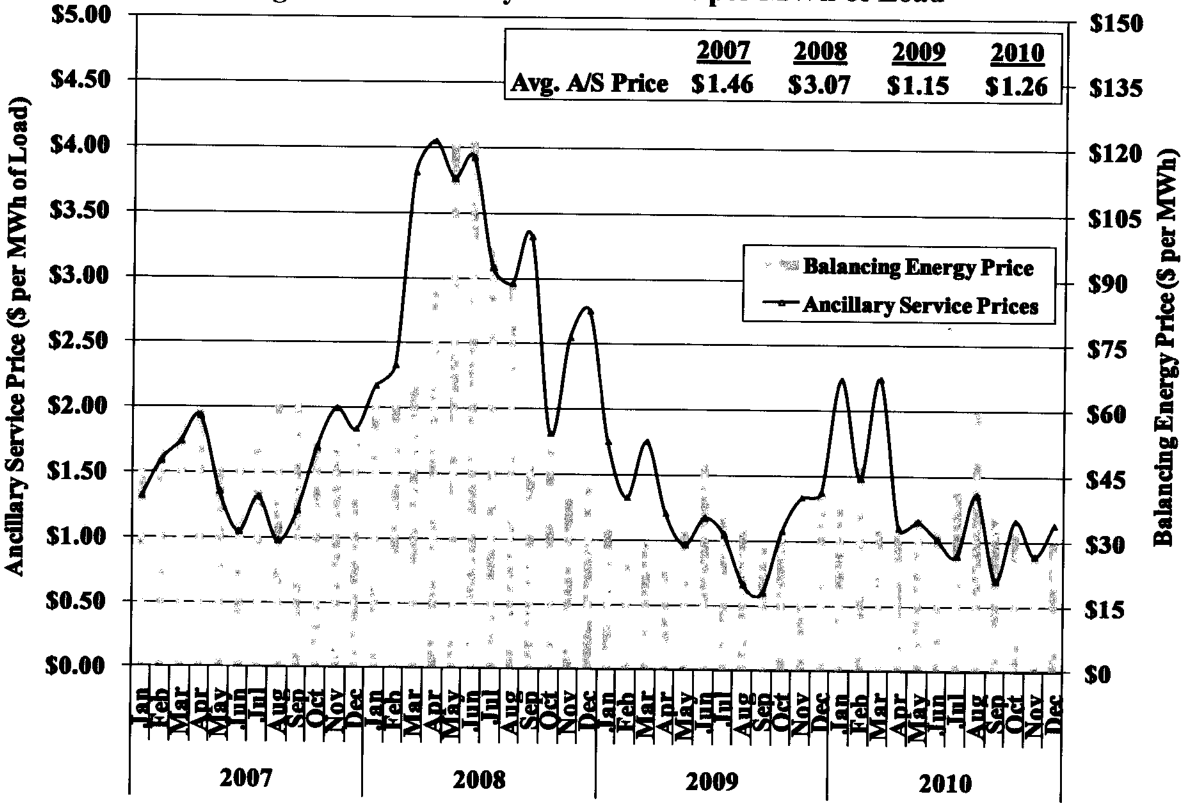
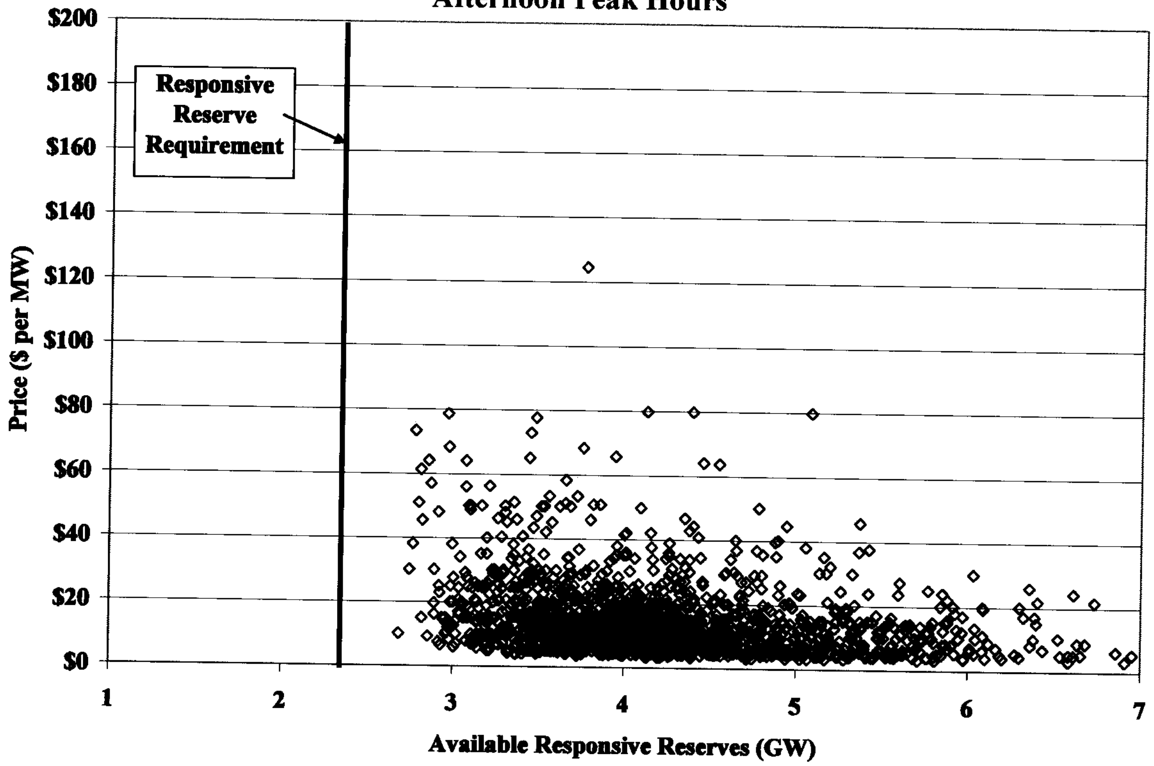


Figure 16 shows that total ancillary service costs are generally correlated with balancing energy price movements, which, as previously discussed, are highly correlated with natural gas price movements. The average ancillary service cost per MWh of load increased to $1.26 per MWh in 2010 compared to $1.15 per MWh in 2009, an increase of 10 percent. Total ancillary service costs were equal to 3.3 and 3.2 percent of the load-weighted average energy price in 2009 and 2010, respectively.

Figure 17 plots the hourly real-time responsive reserves capability against the responsive reserves prices during the peak afternoon hours of 2 PM to 6 PM. The capability calculated for this analysis reflects the actual energy output of each generating unit and the actual dispatch point for LaaRs. Hence, units producing energy at their maximum capability will have no available responsive reserves capability. Consistent with ERCOT rules, the responsive reserve that can be provided by each generating unit is limited to 20 percent of the unit's maximum capability. The figure also shows the responsive reserves requirement of 2,300 MW in 2010 to show the amount of the surplus in each hour.

Figure 17: Hourly Responsive Reserves Capability vs. Market Clearing Price Afternoon Peak Hours



In a well functioning-market for responsive reserves, we would expect excess capacity to be negatively correlated with the clearing prices. Unfortunately, the data in this figure indicate only a weak negative correlation under the zonal market. As in prior years, the frequency with which price exceeds $20 per MW at times with significant excess responsive reserve capability available is surprising. In these hours the marginal costs of supplying responsive reserves should be very low. These results continue to reinforce the potential benefits that should result from jointly optimizing the operating reserves and energy markets.

One of the most obvious improvements brought about by the nodal market implementation was to the market for regulation reserves. In the first month of the nodal market, the total cost of regulating reserves decreased by $8.5 million when compared to the costs in December of 2009. The reduction is primarily attributable to the combination of less regulation capacity procured because of more frequent five minute energy deployments in the nodal market compared to 15 minutes in the zonal market.

II. DEMAND AND RESOURCE ADEQUACY

The first section of this report reviewed the market outcomes and provided analyses of a variety of factors that have influenced the market outcomes. This section reviews and analyzes the load patterns during 2010 and the existing generating capacity available to satisfy the load and operating reserve requirements.

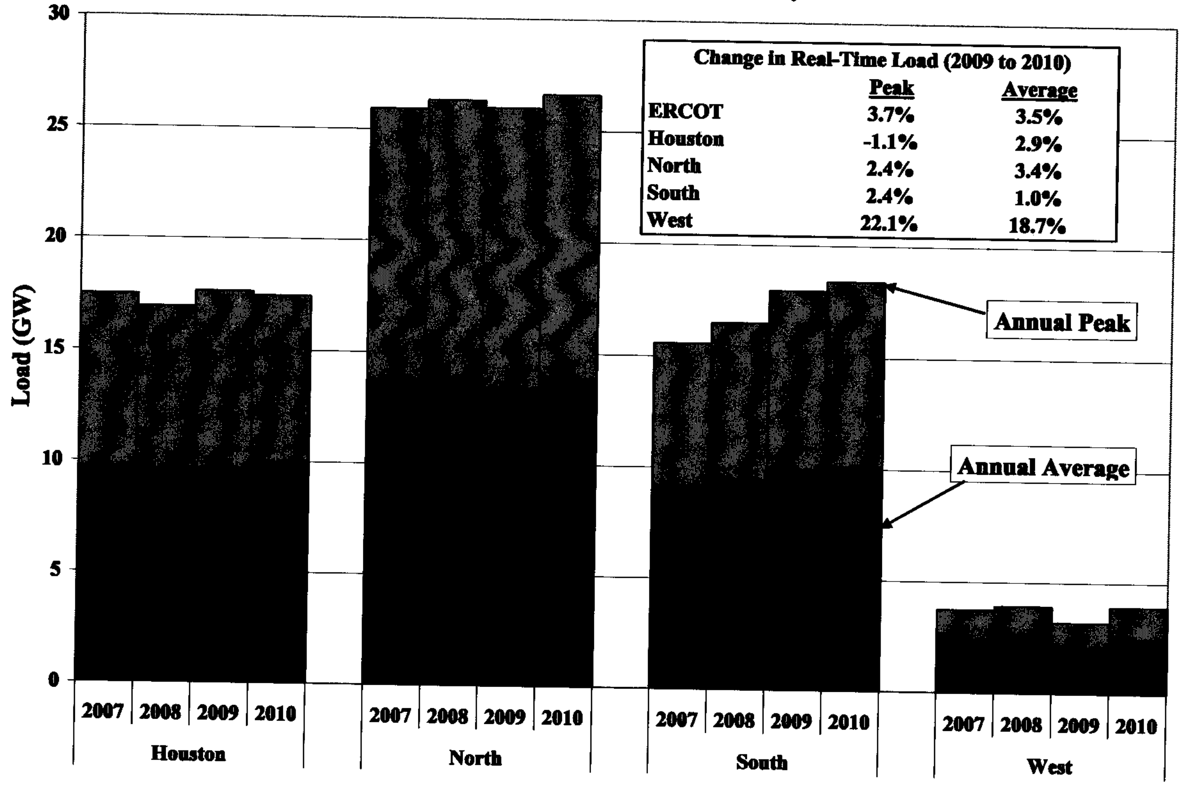
A. ERCOT Loads in 2010

There are two important dimensions of load that should be evaluated separately. First, the changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year. Second, it is important to separately evaluate the changes in the load during the highest-demand hours of the year. Significant changes in these peak demand levels have historically been very important and played a major role in assessing the need for new resources. The expectation in a regulated environment was that adequate resources would be acquired to serve all firm load, and this expectation remains in the competitive market. The expectation of resource adequacy is based on the value of electric service to customers and the damage and inconvenience to customers that can result from interruptions to that service. Additionally, significant changes in peak demand levels affect the probability and frequency of shortage conditions (*i.e.*, conditions where firm load is served but minimum operating reserves are not maintained). Hence, both of these dimensions of load during 2010 are examined in this subsection and summarized in Figure 18.

This figure shows peak load and average load in each of the ERCOT zones from 2007 to 2010. In each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North Zone is the largest zone (with about 38 percent of the total ERCOT load);[9] the South and Houston Zones are comparable (with about 27 percent) while the West Zone is the smallest (with about 7 percent of the total ERCOT load). Figure 18 shows the annual non-coincident peak load for each zone. This is the highest load that occurred in a particular zone for one hour during the year; however, the peak can occur in different hours for different zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

9 The Northeast Zone was integrated into the North Zone in 2007.

Figure 18: Annual Load Statistics by Zone

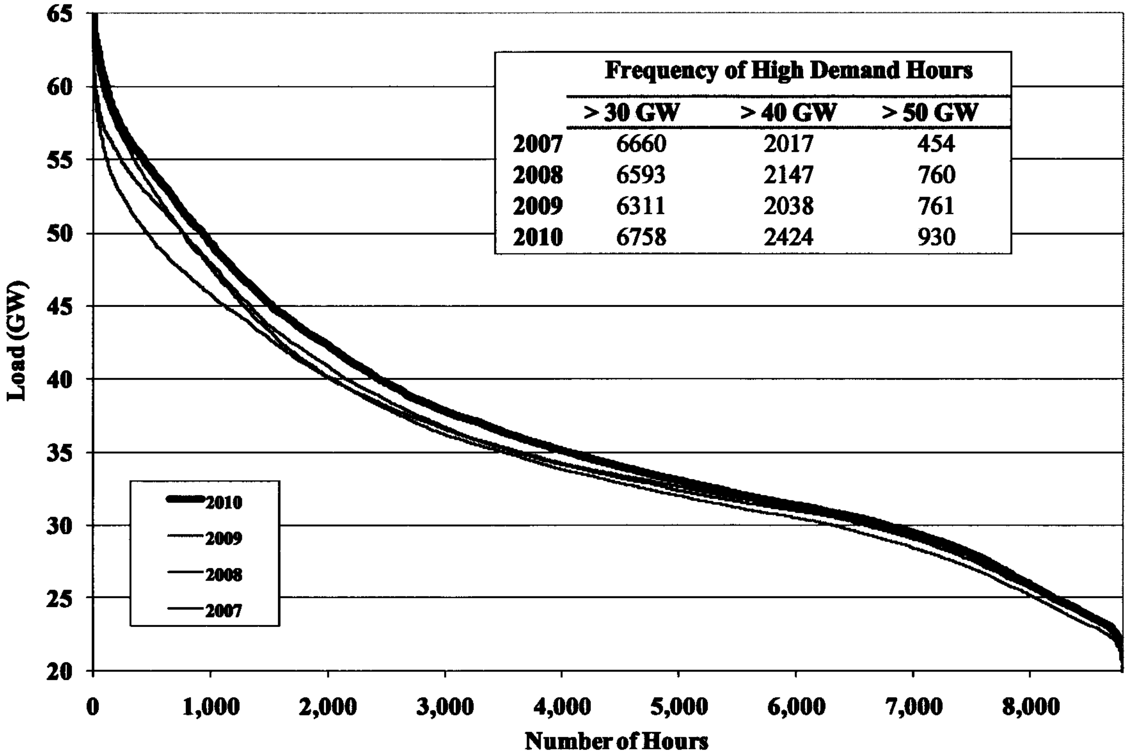


Overall, the ERCOT total load increased from 308,278 GWh in 2009 to 319,239 GWh, an increase of 3.5 percent, or an average of 1250 MW every hour. Similarly, the ERCOT coincident peak demand increased from 63,400 MW in 2009 to 65,776 MW in 2010, an increase of 3.7 percent.

To provide a more detailed analysis of load at the hourly level, Figure 19 compares load duration curves for each year from 2007 to 2010. A load duration curve shows the number of hours (shown on the horizontal axis) that load exceeds a particular level (shown on the vertical axis). ERCOT has a fairly smooth load duration curve, typical of most electricity markets, with low to moderate electricity demand in most hours, and peak demand usually occurring during the late afternoon and early evening hours of days with exceptionally high temperatures.

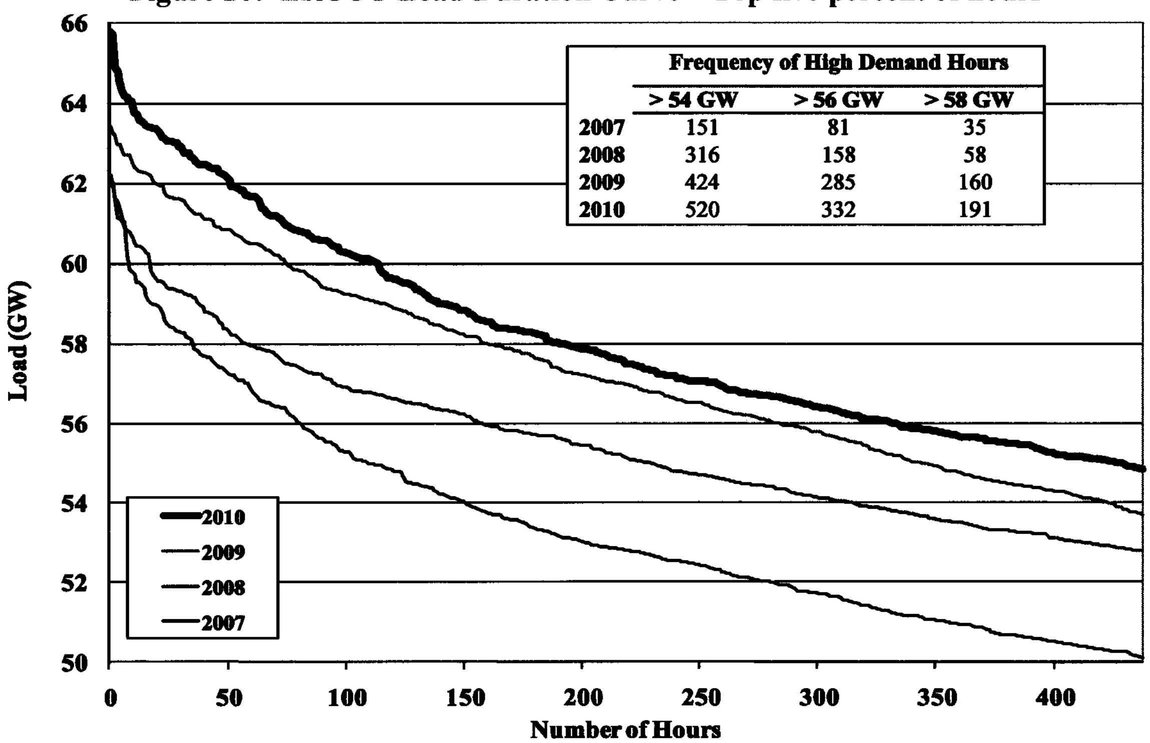
As shown in Figure 19, the load duration curve for 2010 is higher than in 2009 and is consistent with the load increase of 3.5 percent from 2009 to 2010.

Figure 19: ERCOT Load Duration Curve – All hours



	Frequency of High Demand Hours		
	> 30 GW	> 40 GW	> 50 GW
2007	6660	2017	454
2008	6593	2147	760
2009	6311	2038	761
2010	6758	2424	930

Figure 20: ERCOT Load Duration Curve – Top five percent of hours

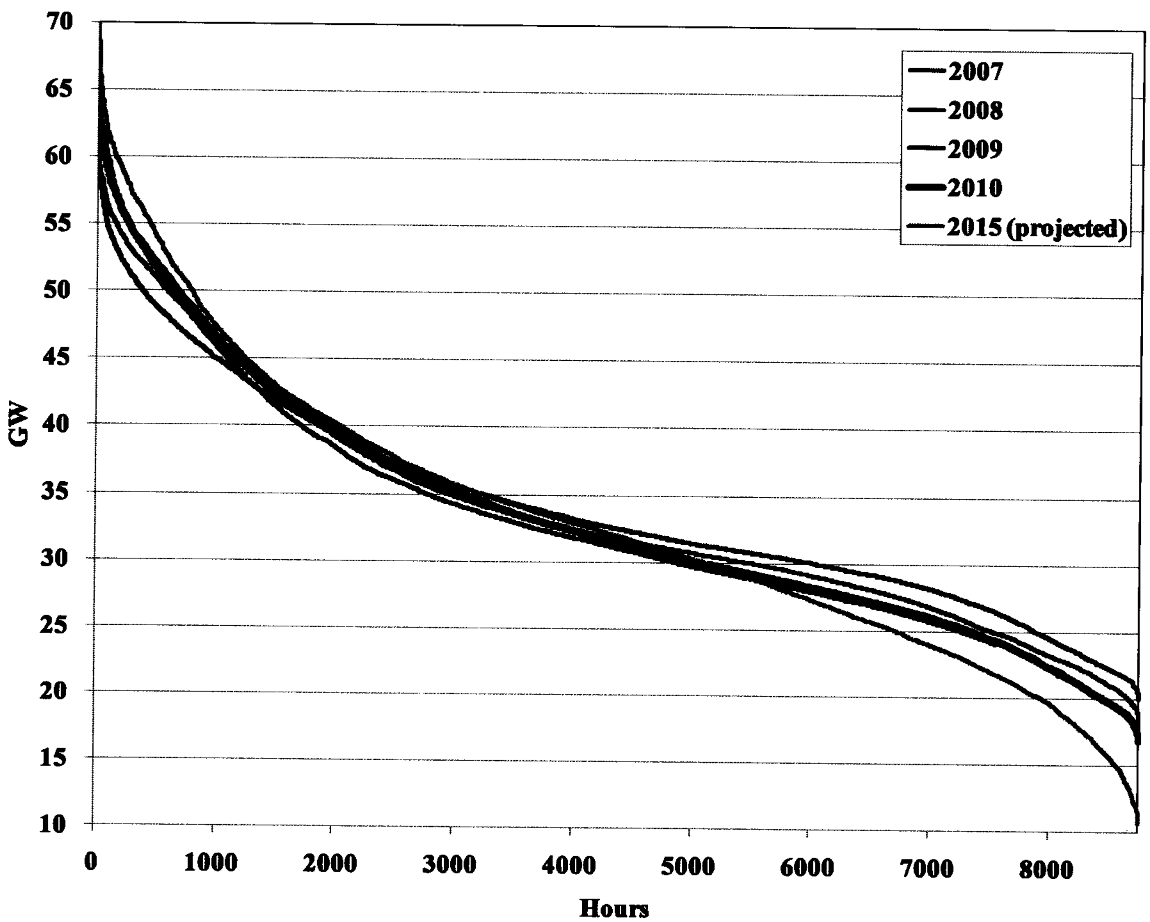


	Frequency of High Demand Hours		
	> 54 GW	> 56 GW	> 58 GW
2007	151	81	35
2008	316	158	58
2009	424	285	160
2010	520	332	191

To better illustrate the differences in the highest-demand periods between years, Figure 20 shows the load duration curve for the five percent of hours with the highest loads. This figure also shows that the peak load in each year is significantly greater than the load at the 95th percentile of hourly load. From 2007 to 2010, the peak load value averaged 20 percent greater than the load at the 95th percentile. These load characteristics imply that a substantial amount of capacity – over 10 GW – is needed to supply energy in less than 5 percent of the hours.

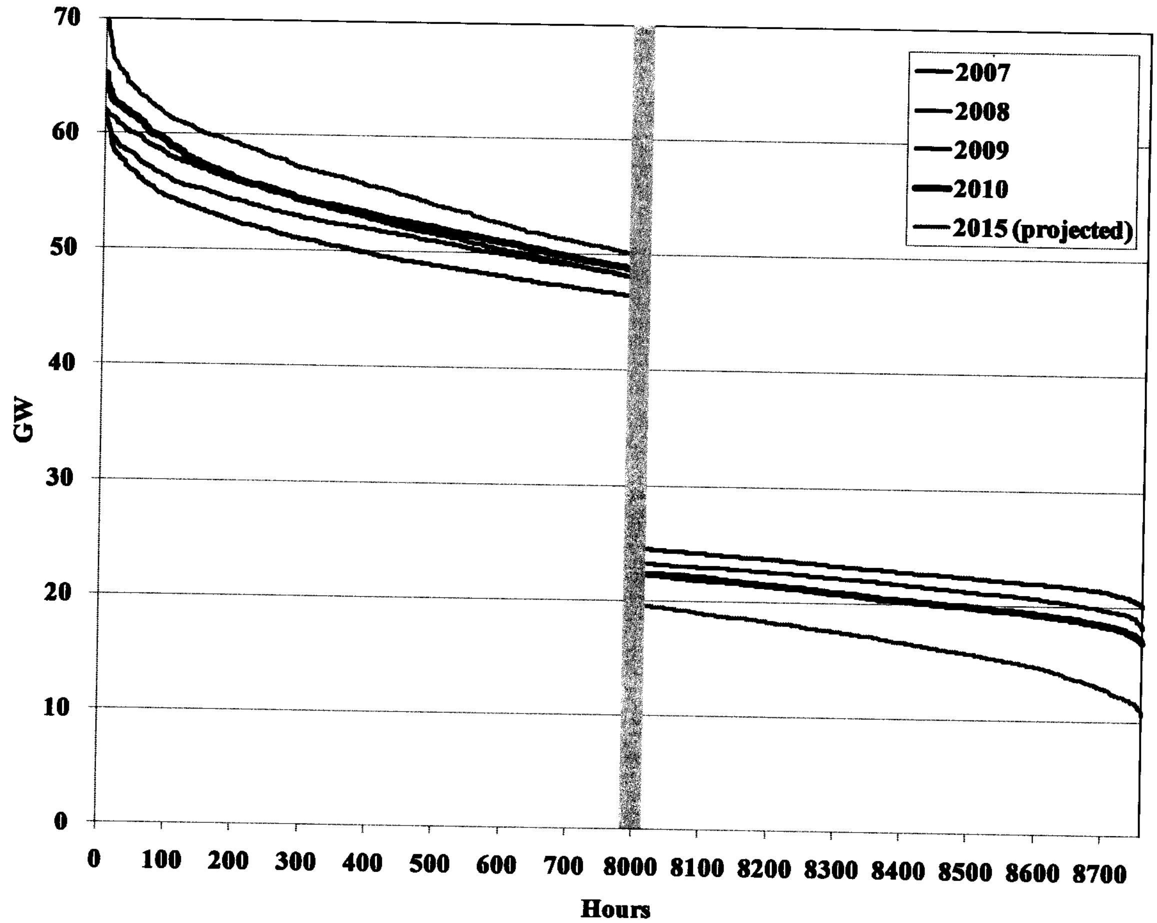
Increasing levels of wind resource in ERCOT also has important implications for the net load duration curve faced by the non-wind fleet of resources. Net load is defined as the system load minus wind production. Figure 21 shows the net load duration curves for 2007 through 2010, with projected values for 2015 based on ERCOT data from its Competitive Renewable Energy Zones assessment.

Figure 21: Net Load Duration Curves



Over 90 percent of the wind resources in the ERCOT region are located in West Texas, and the wind profiles in this area are such that most of the wind production occurs during off-peak hours or other times of relatively low system demand. This profile results in only modest reductions of the net load relative to the actual load during the hours of highest demand, but much more significant reductions in the net load relative to the actual load in the other hours of the year. The projection for 2015 indicates that the trend shown from 2007 to 2010 in Figure 21 is expected to continue and amplify with the addition of significant new wind resources and the reduction in the curtailment of existing wind resources after completion of new transmission facilities.

Figure 22: Top and Bottom Ten Percent of Net Load



Focusing on the left side of the net load duration curve shown in Figure 22, the average difference between peak net load and the 95th percentile of net load was 11.2 GW in 2007 to 2010, but this differential is projected to increase to over 15 GW by 2015. With an additional capacity requirement of more than 9 GW to meet the 13.75 percent reserve margin requirement, this means that over 24 GW of non-wind capacity will be required to exist on the system with an expectation of operating less than five percent of the hours in a year.

On the right side of the net load duration curve, the minimum net load was 17 GW in 2009 and 2010, but is projected to continue to decrease to less than 11 GW by 2015. These decreasing minimum load levels are expected to put operational pressure on the nearly 25 GW of nuclear and coal fuel generation currently installed in ERCOT.

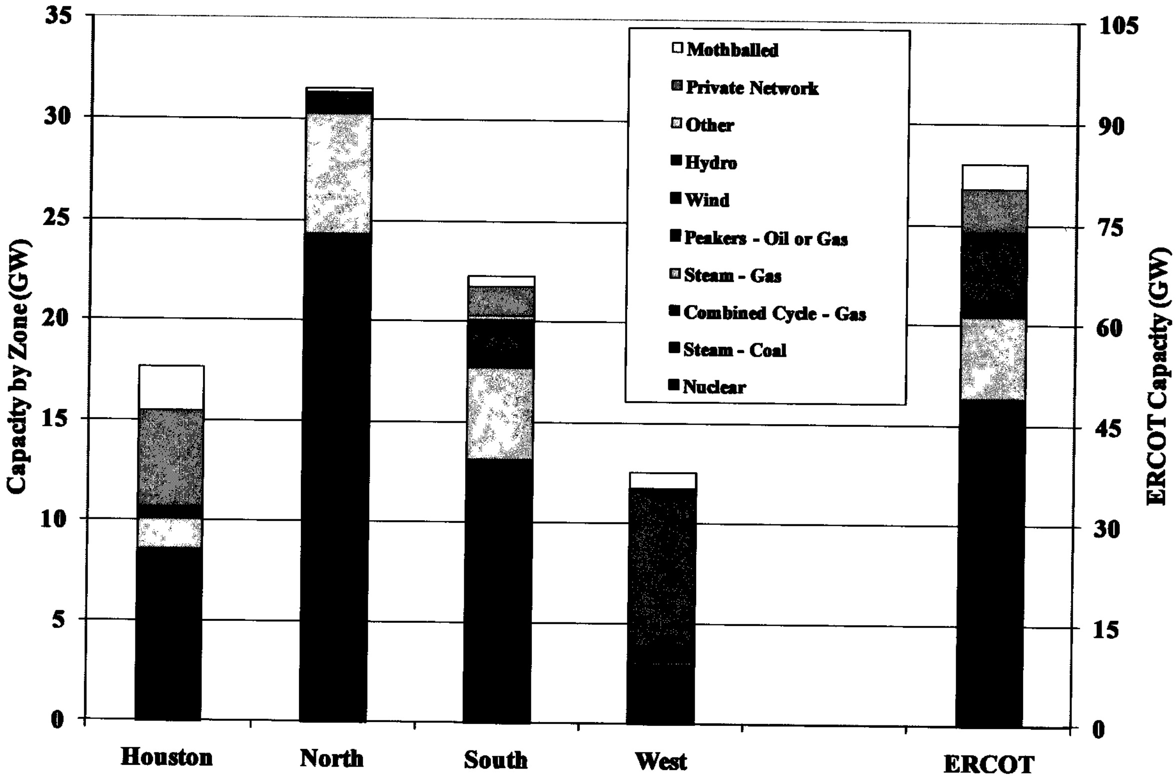
Thus, although the peak net load and reserve margin requirements are projected to continue to increase and create an increasing need for non-wind capacity to meet net load and reliability requirements, the non-wind fleet is expected to operate for fewer hours as wind penetration continues to increase. This outlook further reinforces the importance of efficient energy pricing during peak demand conditions and other times of system stress, particularly within the context of the ERCOT energy-only market design.

B. Generation Capacity in ERCOT

In this section we evaluate the generation mix in ERCOT. With the exception of the wind resources in the West Zone and the nuclear resources in the North and Houston Zones, the mix of generating capacity is relatively uniform in ERCOT. Figure 23 shows the installed generating capacity by type in each of the ERCOT zones.

The distribution of capacity among the ERCOT zones is similar to the distribution of demand with the exception of the large amount of wind capacity in the West Zone. The North Zone accounts for approximately 38 percent of capacity, the South Zone 27 percent, the Houston Zone 21 percent, and the West Zone 15 percent. The Houston Zone typically imports power, while the West and North Zones typically export power. Excluding mothballed resources and including only 8.7 percent of wind capacity as capacity available to reliably meet peak demand, the North Zone accounts for approximately 43 percent of capacity, the South Zone 29 percent, the Houston Zone 22 percent, and the West Zone 6 percent.

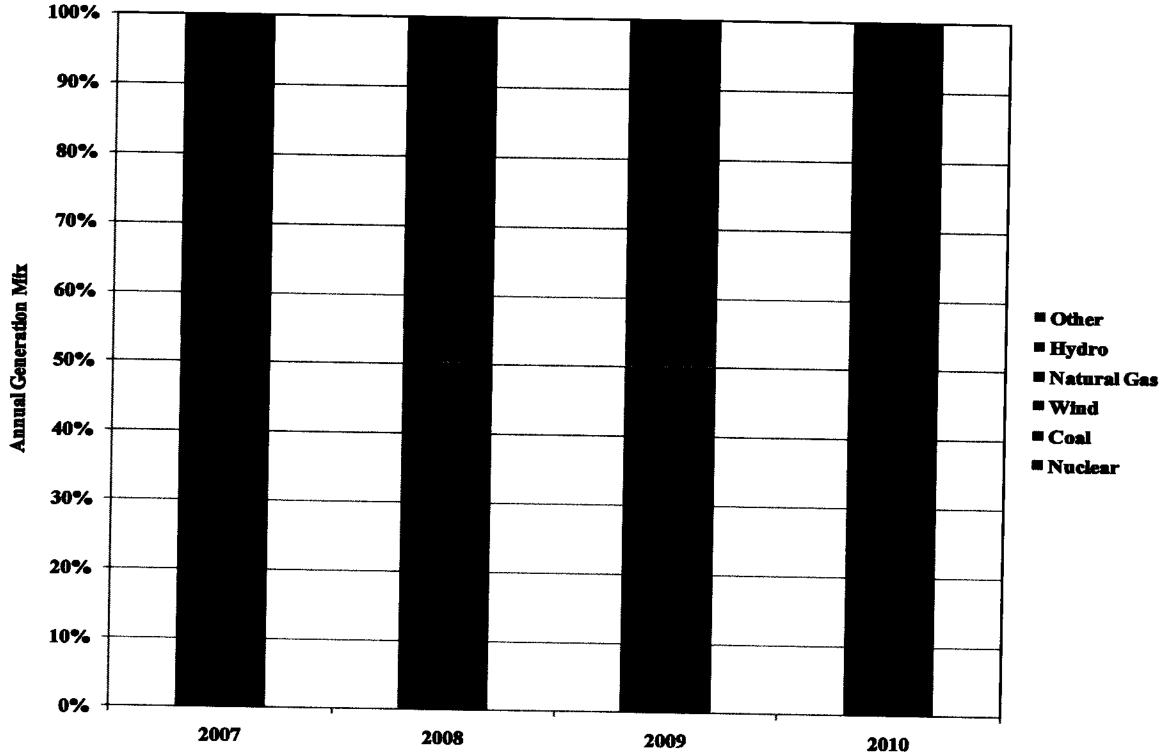
Figure 23: Installed Capacity by Technology for each Zone



Notable changes to ERCOT's installed generation during 2010 included new coal units and wind units increasing their percentage shares, while several less efficient natural gas fueled units were mothballed or retired. Even after these changes natural gas generation accounts for approximately 50 percent of the installed capacity in ERCOT.

The shifting contribution of coal and wind generation is evident in Figure 24, which shows the percent of annual generation from each fuel type for the years 2007 through 2010. In 2010 the percentage of generation produced by coal units increased from 37 percent to 40 percent. Wind generation provided 8 percent of the annual generation requirement in 2010, up from 3 percent in 2007. During the same period the percentage of generation provided by natural gas decreased from 45 percent to 38 percent.

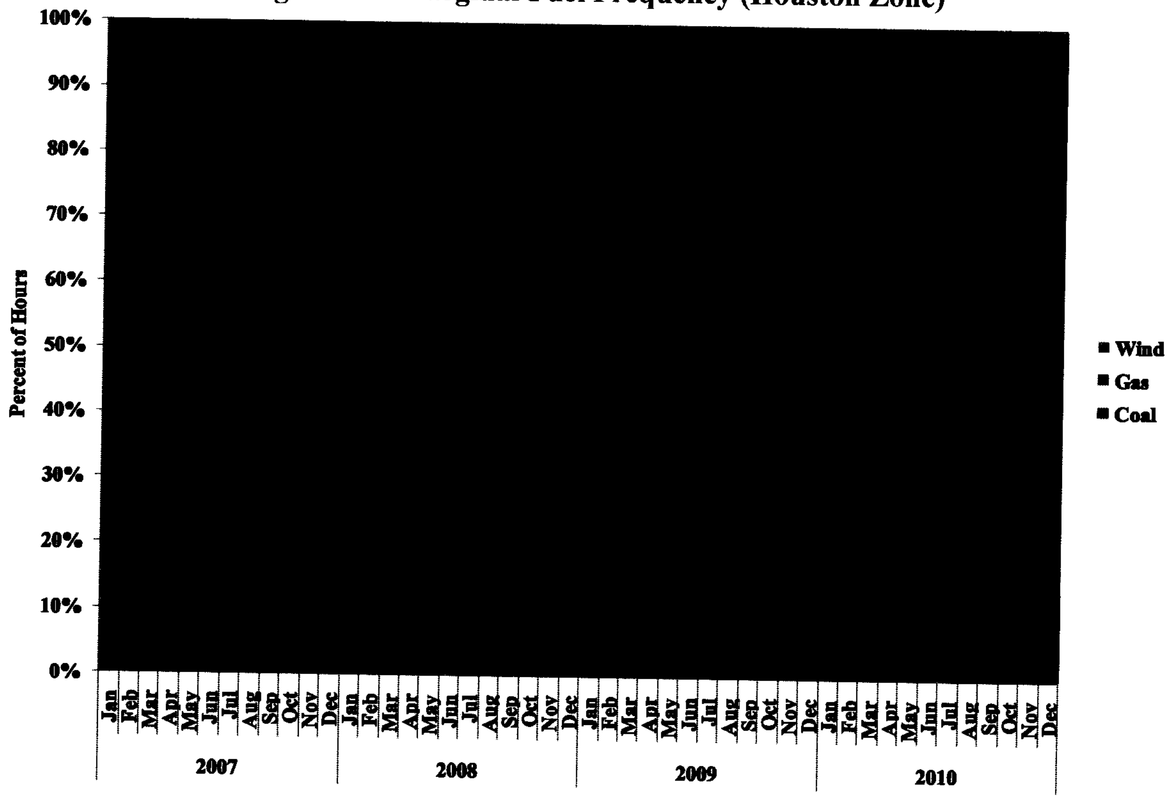
Figure 24: Annual Generation Mix



While ERCOT has coal/lignite and nuclear plants that operate primarily as base load units, its reliance on natural gas resources makes it vulnerable to natural gas price spikes. There is approximately 25 GW of coal and nuclear generation in ERCOT. Because there are few hours when ERCOT load is at this low level, natural gas resources will be dispatched and set the balancing energy spot price in most hours. Although coal-fired and nuclear units combined to produce more than half of the energy in ERCOT, they have historically played a much less significant role in setting spot electricity prices. However, with the recent additions of new coal generation combined with significant increases in wind capacity, with its lower marginal production cost than coal and lignite, the frequency at which coal and lignite are the marginal units in ERCOT is expected to increase in the future, particularly during the off-peak hours in the spring and fall, and even more as additional transmission capacity is added that will accommodate increased levels of wind production in the West Zone.

Figure 25 and Figure 26 show the marginal fuel frequency for the Houston and West Zones, respectively, for each month from 2007 through 2010.[10] The marginal fuel frequency is the percentage of hours that a generation fuel type is marginal and setting the price at a particular location.

Figure 25: Marginal Fuel Frequency (Houston Zone)

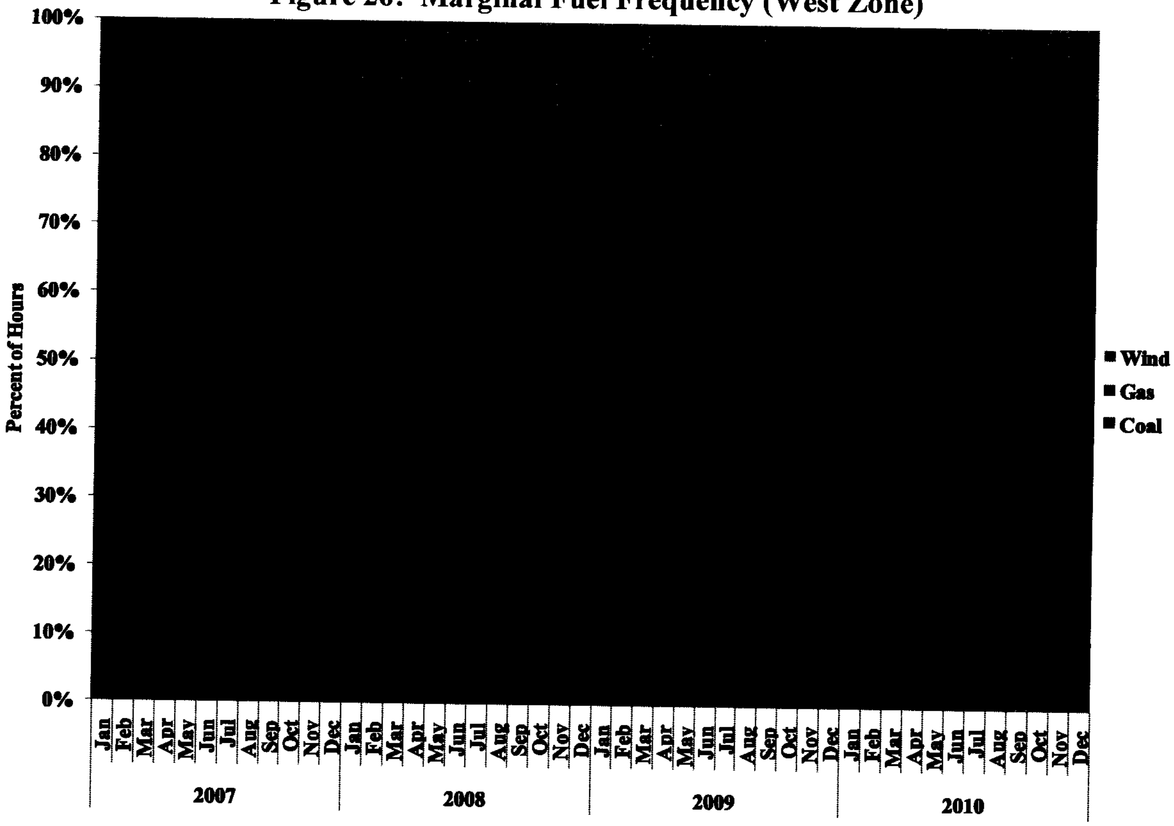


As shown in Figure 25, the frequency at which coal was the price setting fuel for the Houston Zone experienced a significant and sustained increase beginning in September 2008. This increase can be attributed to (1) increased wind resource production; (2) a slight reduction in demand in 2009 due to the economic downturn; and (3) periods when natural gas prices were very low thereby making coal and combined-cycle natural gas resources competitive from an economic dispatch standpoint. As significant additional wind, coal and potentially nuclear resources are added to the ERCOT region and transmission constraints that limit existing wind production are alleviated, it is likely that the frequency of coal as the marginal fuel will increase in coming years.

10 The marginal fuel frequency for the North and South Zones are very similar to the Houston Zone.

Figure 26 shows that the frequency at which coal was the price setting fuel for the West Zone also experienced a significant and sustained increase beginning in September 2008. This figure also shows that beginning in late 2007 the frequency at which wind was the price setting fuel for the West Zone increased dramatically. This increase is attributable to the growth in installed wind capacity that far exceed the load in the West Zone combined with existing transmission capability that limits the export capability from the West Zone, as discussed in more detail in Section III.

Figure 26: Marginal Fuel Frequency (West Zone)



The average profile of West Zone wind production is negatively correlated with the load profile, with the highest wind production occurring during non-summer months, and predominately during off-peak hours. Figure 27 shows the average West Zone wind production for each month in 2010, with the average production in each month shown separately in four hour blocks.[11]

11 Figure 27 shows actual wind production, which was affected by curtailments at the higher production levels. Thus, the higher levels of actual wind production in Figure 27 are lower than the production levels that would have materialized absent transmission constraints.

Figure 27: Average West Zone Wind Production

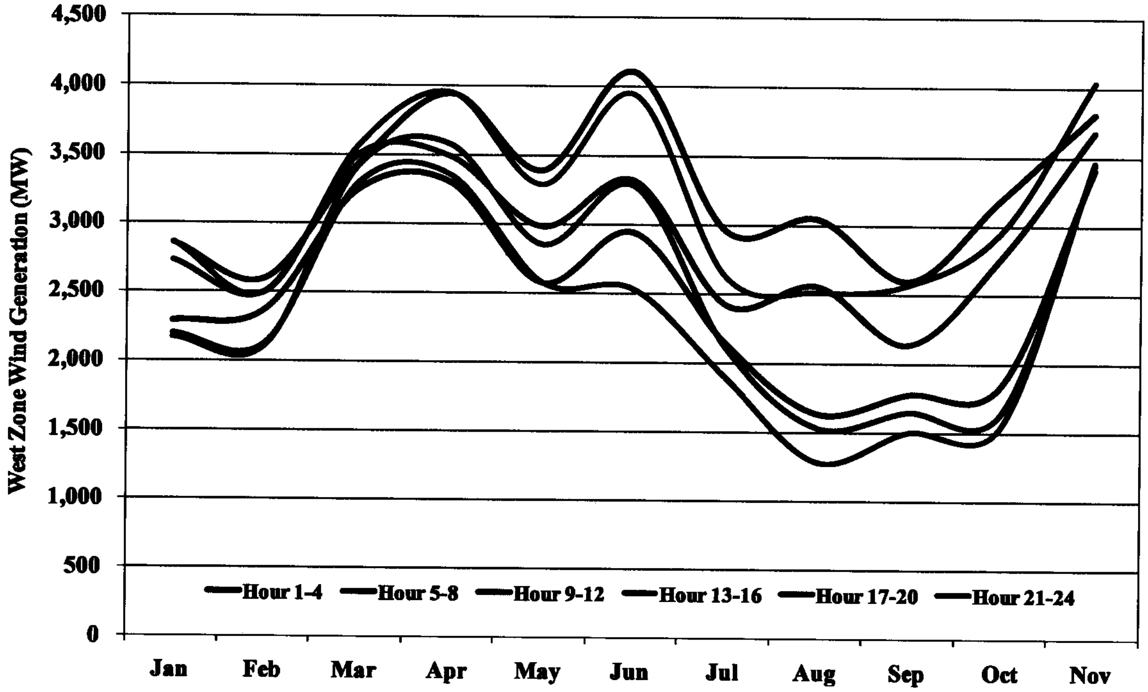
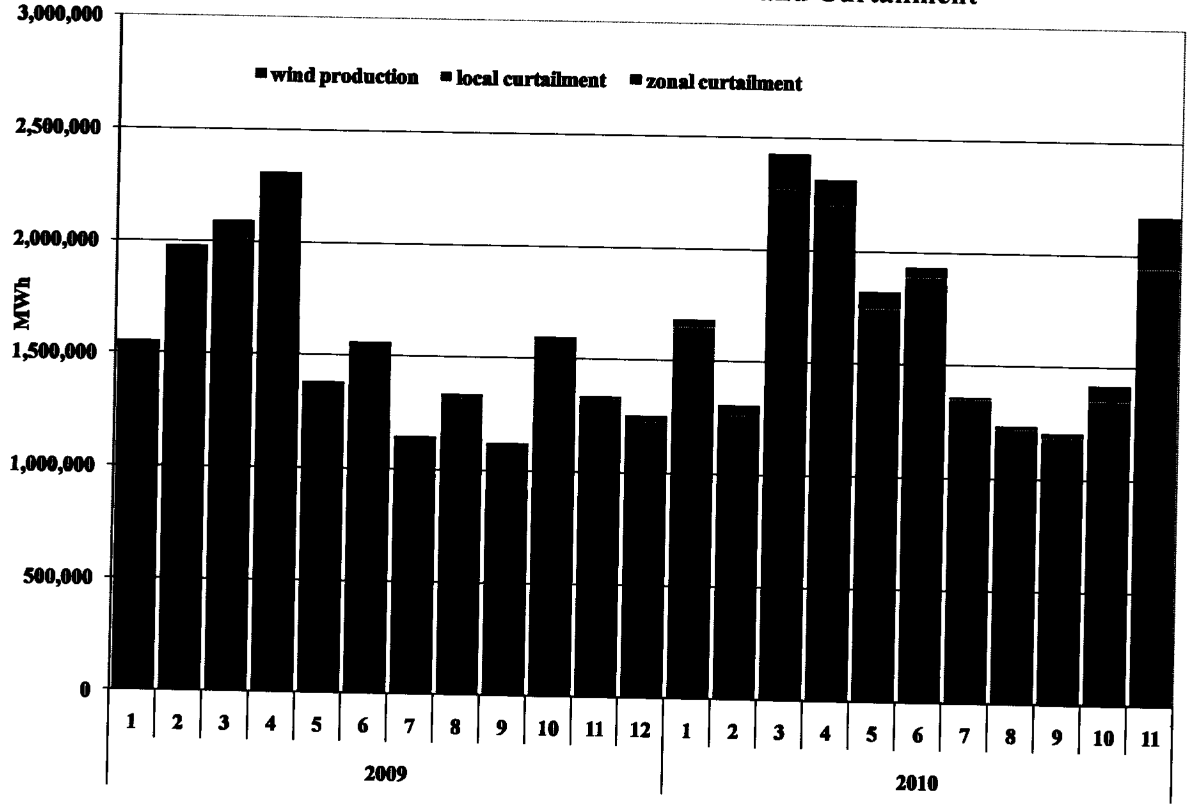


Figure 28 shows the wind production and local and zonal curtailment quantities for the West Zone for each month of 2009 and 2010. This figure reveals that the quantity of zonal curtailments for wind resources in the West Zone was increased from 442 GWh in 2009 to over 785 GWh in 2010, while the quantity of local curtailments decreased from over 3,400 GWh in 2009 to 1,068 GWh in 2010.

Figure 28: West Zone Wind Production and Curtailment

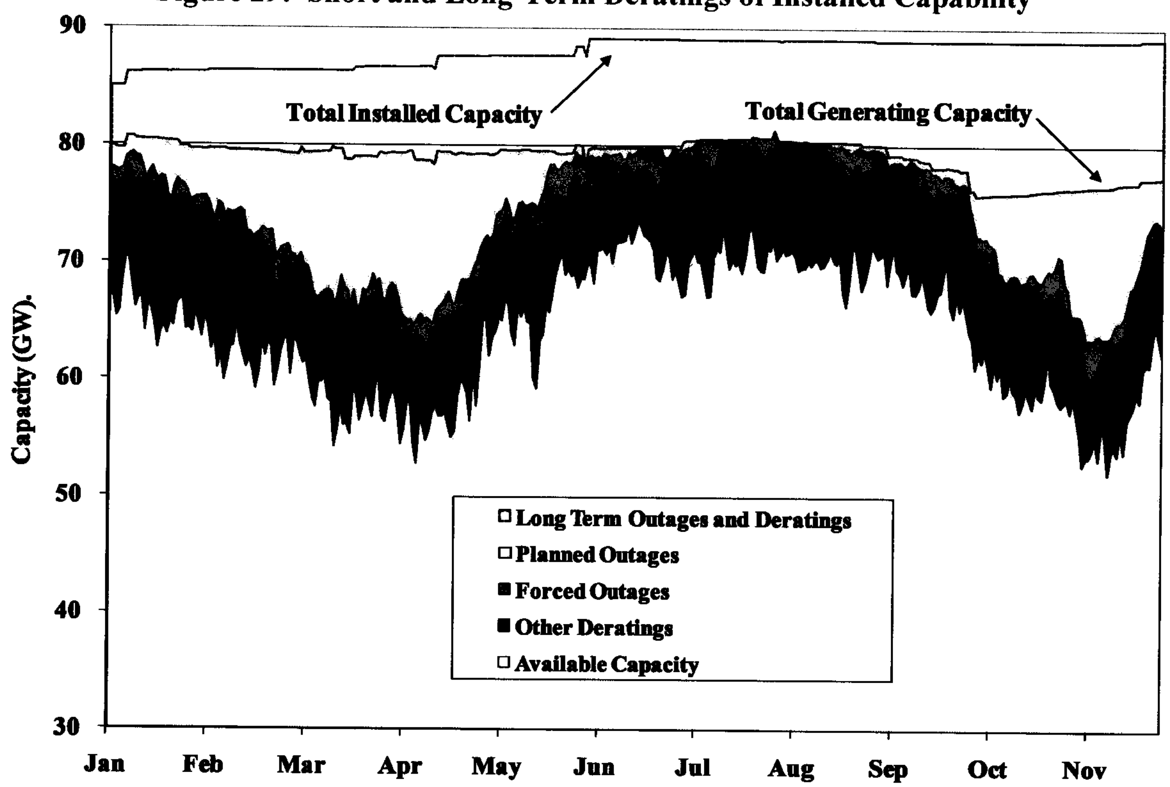


1. Generation Outages and Deratings

Figure 23 in the prior subsection shows that installed capacity is approximately 86 GW including mothballed units and wind units at nameplate capacity, and approximately 72 GW excluding mothballed capacity and including only the 8.7 percent of wind capacity assumed to be dependably available during peak load. Hence, the installed capacity exceeds the capacity required to meet annual peak load plus ancillary services requirements of 67 GW. This might suggest that the adequacy of resources is not a concern for ERCOT in the near-term. However, resource adequacy must be evaluated in light of the resources that are actually available on a daily basis to satisfy the energy and operating reserve requirements in ERCOT. A substantial portion of the installed capability is frequently unavailable due to generator deratings. A derating is the difference between the maximum installed capability of a generating resource and its actual capability (or "rating") in a given hour. Generators may be fully derated (rating equals 0) due to a forced or planned outage. It is also very common for generating capacity to be partially derated (*e.g.*, by 5 to 10 percent) because the resource cannot achieve its installed capability level due to technical or environmental factors (*e.g.*, component equipment failures or ambient temperature conditions).

In this subsection, we evaluate long-term and short-term deratings to inform our evaluation of ERCOT capacity levels. Figure 29 shows a breakdown of total installed capability for ERCOT on a daily basis during 2010. This analysis includes all in-service and switchable capacity. The capacity in this analysis is separated into five categories: (a) long-term outages and deratings, (b) short-term planned outages, (c) short-term forced outages, (d) other short-term deratings, and (e) available and in-service capability.

Figure 29: Short and Long-Term Deratings of Installed Capability*



* Includes all outages and deratings lasting greater than 60 days and all mothballed units.
* Switchable capacity is included under installed capacity in this figure.

Figure 29 shows that long-term outages and other deratings fluctuated between 13 and 24 GW. These outages and deratings reduce the effective resource margins in ERCOT from the levels reported above. A large component of the "other deratings" is associated with limited wind resources resulting in generating resources that are not capable of producing up to the full installed capability. Other causes of these deratings reflect:

- Cogeneration resources unavailable to serve market load because they are being used to serve self-serve load;
- Resources out-of-service for economic reasons (*e.g.*, mothballed units); or
- Resources out-of-service for extended periods due to maintenance requirements.

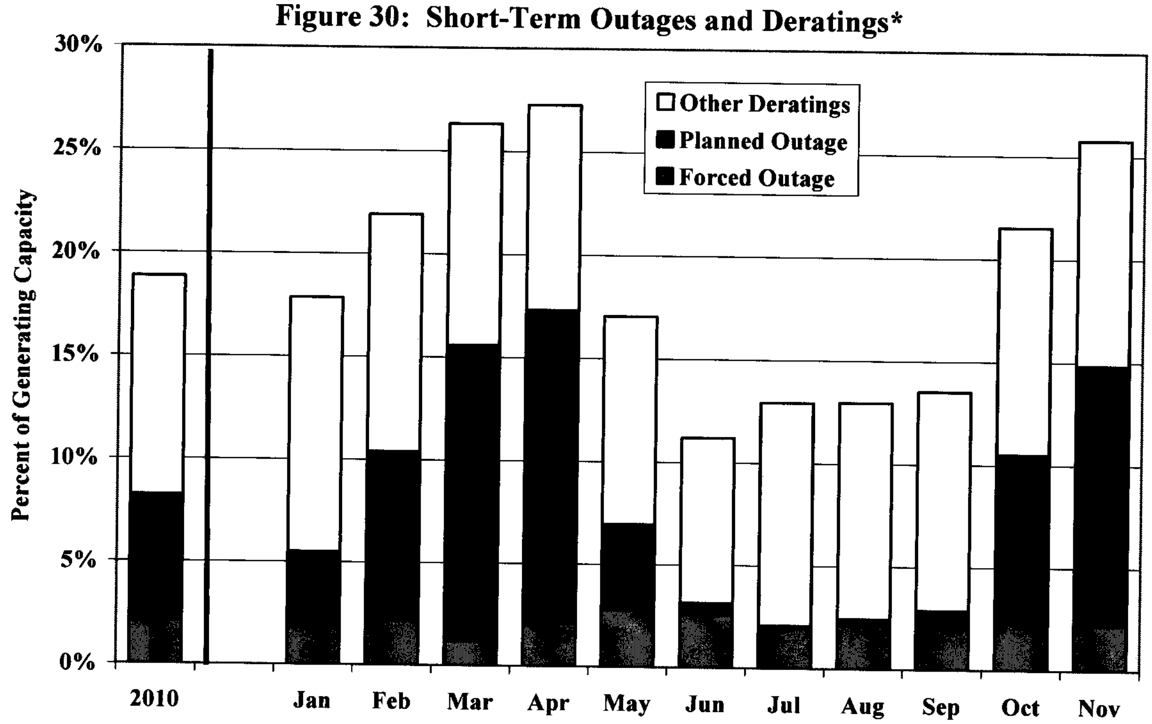
With regard to short-term deratings and outages, the patterns of planned outages and forced outages were consistent with expectations:

- Forced outages occurred randomly over the year and the forced outage rates were relatively low (although all forced outages may not be reported to ERCOT).
- Planned outages were relatively large in the spring and fall and extremely small during the summer.

The increase in the annual average of daily available capacity in 2010 from 2009 was approximately the same as the average daily load increase.

The next analysis focuses specifically on the short-term forced outages and other short-term deratings. Figure 30 shows the average magnitude of the outages and deratings lasting less than 60 days for the year and for each month during 2010.

Figure 30: Short-Term Outages and Deratings*



* Excludes all outages and deratings lasting greater than 60 days and all mothballed units.

Figure 30 shows that total short-term deratings and outages were as large as 27 percent of installed capacity in the spring and fall, dropping to as low as 11 percent for the summer. Most of this fluctuation was due to anticipated planned outages, which ranged from 8 to 15 percent of installed capacity during spring and fall. Short-term forced outages occurred more randomly, as would be expected, ranging between one and three percent of total capacity on a monthly average basis during 2010. These rates are relatively low in comparison to other operating markets for two reasons. First, this measure of outages includes only full outages (*i.e.*, where the resource's rating equals zero). In contrast, other markets frequently report an equivalent forced outage rate, which includes both full and partial outages. Hence, the forced outage rate shown in Figure 30 is expected to be lower than equivalent forced outage rates reported for other markets. Second, because forced outage information is self-reported by generators, we are not confident that the available data includes all forced outages that actually occurred.

The largest category of short-term deratings was the "other deratings" that occur for a variety of reasons. The other deratings would include any short-term forced or planned outage that was not reported or correctly logged by ERCOT. This category also includes deratings due to ambient temperature conditions, cogeneration uses, wind deratings due to variable wind conditions and other factors described above. Furthermore, suppliers may delay maintenance on components such as boiler tubes, resulting in reduced capability. Because these deratings can fluctuate day to day or by season, some of the deratings are included in the "long-term outages and deratings" category while others are included in this category. The other deratings were approximately 10 percent on average during the summer in 2010 and as high as 12 percent in other months. These outage patterns are analyzed in more detail in Section IV of this report.

2. Daily Generator Commitments

One of the important characteristics of any electricity market is the extent to which it results in the efficient commitment of generating resources. Under-commitment can cause apparent shortages in real-time and inefficiently high energy prices while over-commitment can result in excessive start-up costs, uplift charges, and inefficiently-low energy prices.

This subsection evaluates the commitment patterns in ERCOT by examining the levels of excess capacity. Excess capacity is defined as the total online capacity plus quick-start[12] units minus the demand for energy, responsive reserve, up regulation and non-spinning reserve provided from online capacity or quick-start units. To evaluate the commitment of resources in ERCOT, Figure 31 plots the excess capacity in ERCOT during 2010. The figure shows the excess capacity in only the peak hour of each weekday because the largest generation commitment usually occurs at the peak hour. Hence, one would expect larger quantities of excess capacity in other hours.

Figure 31: Excess On-Line and Quick Start Capacity During Weekday Daily Peaks

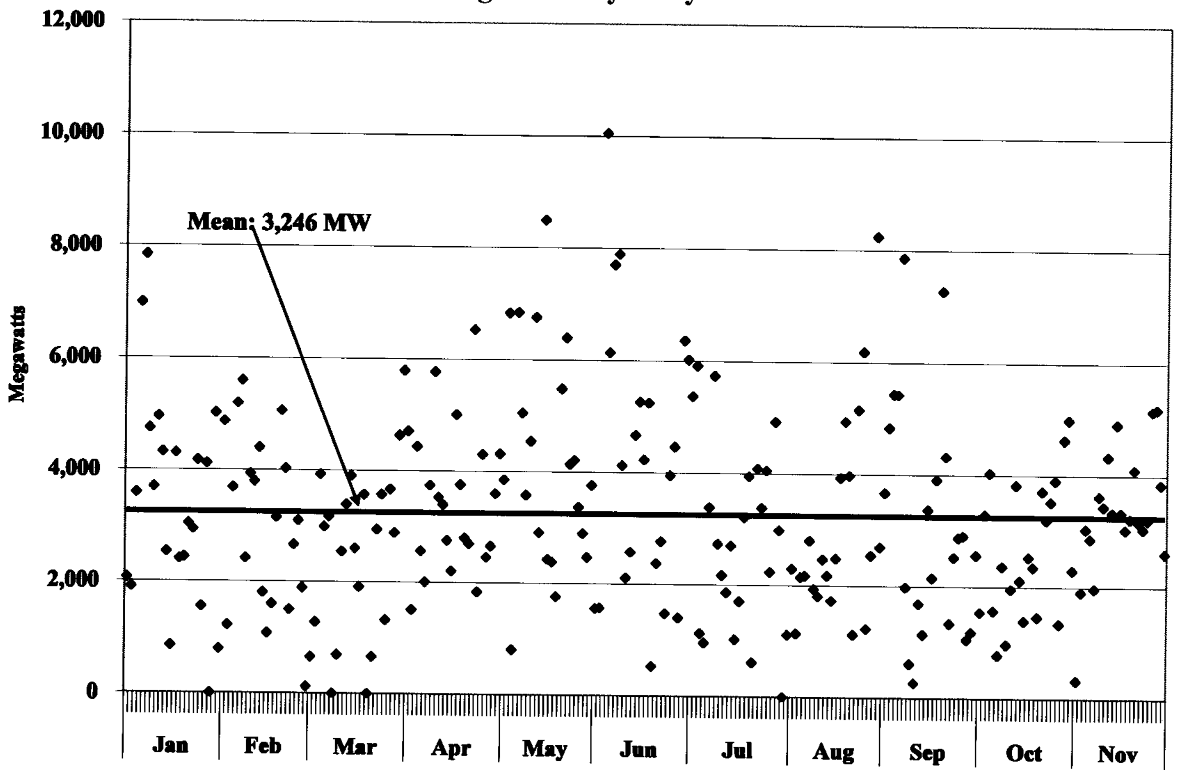


Figure 31 shows that the excess on-line capacity during daily peak hours on weekdays averaged 3,246 MW in 2010, which is approximately 8.9 percent of the average load in ERCOT. This is a decrease of more than 400 MW from the prior year. The overall trend in excess on-line capacity in recent years indicates a movement toward more efficient unit commitment across the ERCOT market; however, the zonal market structure is still based primarily upon a decentralized unit

12 For the purposes of this analysis, "quick-start" includes simple cycle gas turbines that are qualified to provide balancing energy.

commitment process whereby each participant makes independent generator commitment decisions that are not likely to be optimal. Further contributing to the suboptimal results of the current unit commitment process is that the decentralized unit commitment is comprised of non-binding resource plans that form the basis for ERCOT's day-ahead planning decisions. However, these non-binding plans can be modified by market participants after ERCOT's day-ahead planning process has concluded causing ERCOT to take additional actions that may be more costly and less efficient. We expect the introduction of a day-ahead energy market with a financially binding, centralized unit commitment under the nodal market design to result in substantial efficiency improvements in the commitment of generating resources.

C. Demand Response Capability

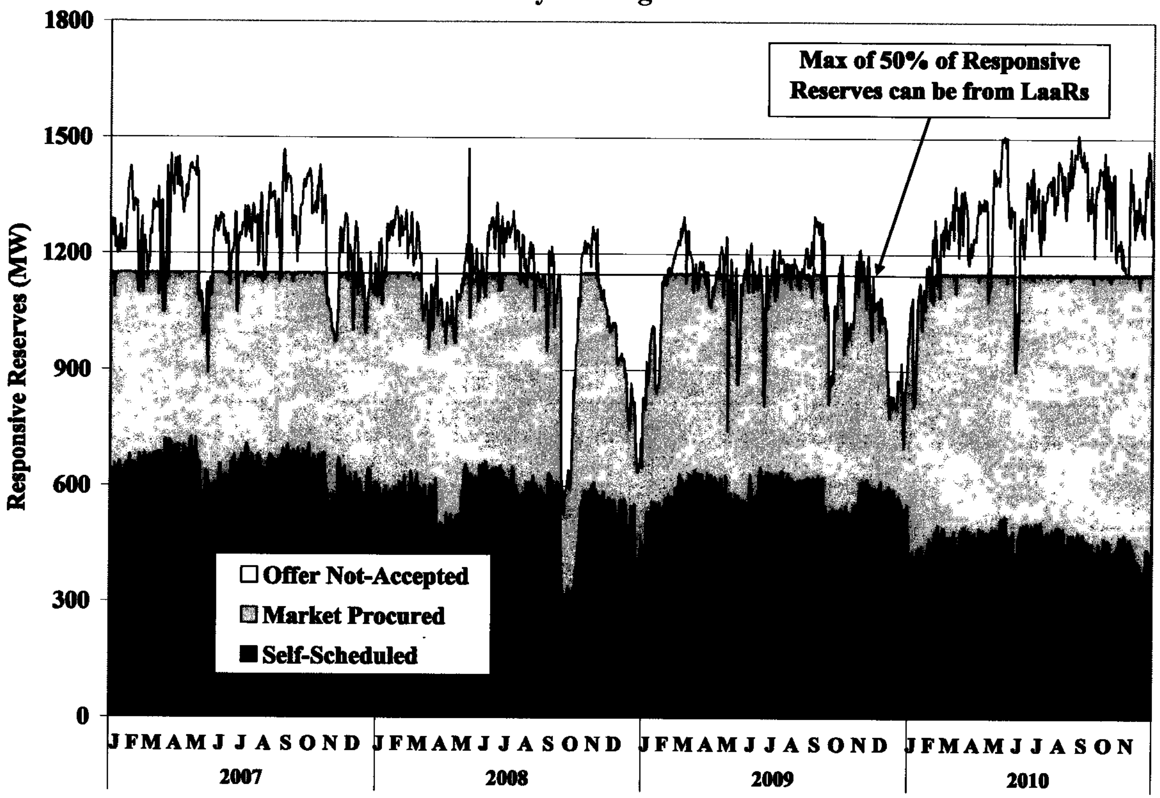
Demand response is a term that broadly refers to actions that can be taken by end users of electricity to reduce load in response to instructions from ERCOT or in response to certain market or system conditions. The ERCOT market allows participants with demand-response capability to provide energy and reserves in a manner similar to a generating resource. The ERCOT Protocols allow for loads to actively participate in the ERCOT administered markets as Loads acting as Resources ("LaaRs"). Additionally, loads may participate passively in the market by simply adjusting consumption in response to observed prices. Unlike active participation in ERCOT administered markets, passive demand response is not directly tracked by ERCOT.

ERCOT allows qualified LaaRs to offer responsive reserves and non-spinning reserves into the day-ahead ancillary services markets. Qualified LaaRs can also offer blocks of energy in the balancing energy market. LaaRs providing up balancing energy must have telemetry and must be capable of responding to ERCOT energy dispatch instructions in a manner comparable to generation resources. Those providing responsive reserves must have high set under-frequency relay ("UFR") equipment. A load with UFR equipment is automatically tripped when the frequency falls below 59.7 Hz, which will typically occur only a few times in each year.

As of December 2010, over 2,200 MW of capability were qualified as LaaRs. These resources regularly provided reserves in the responsive reserves market, but never participated in the balancing energy market and only a very small portion participated in the non-spinning reserves market. Figure 32 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2010.

Figure 32: Provision of Responsive Reserves by LaaRs

Daily Average



The high level of participation by demand response in the ancillary service markets sets ERCOT apart from other operating electricity markets. Figure 32 shows that the amount of offers by LaaRs to provide responsive reserves routinely exceeds 1,150 MW. For reliability reasons, 1,150 MW is the maximum amount of responsive reserves that can be reliably provided by LaaRs. Notable exceptions include a decrease in September of 2008 corresponding to the Texas landfall of Hurricane Ike and a more prolonged reduction from November 2008 through January 2009 that was likely a product of the economic downturn and its effect on industrial operations. Another seasonal reduction was observed during late 2009.

Although LaaRs are active participants in the responsive reserves market, they did not offer into the balancing energy, regulation, or non-spinning reserve services markets in 2010. This lack of participation is not surprising because the value of curtailed load tends to be very high, and providing responsive reserves offers substantial revenue with very little probability of being deployed. In contrast, providing non-spinning reserves introduces a much higher probability of being curtailed. Participation in the regulation services market requires technical abilities that most LaaRs cannot meet at this point.

D. Net Revenue Analysis

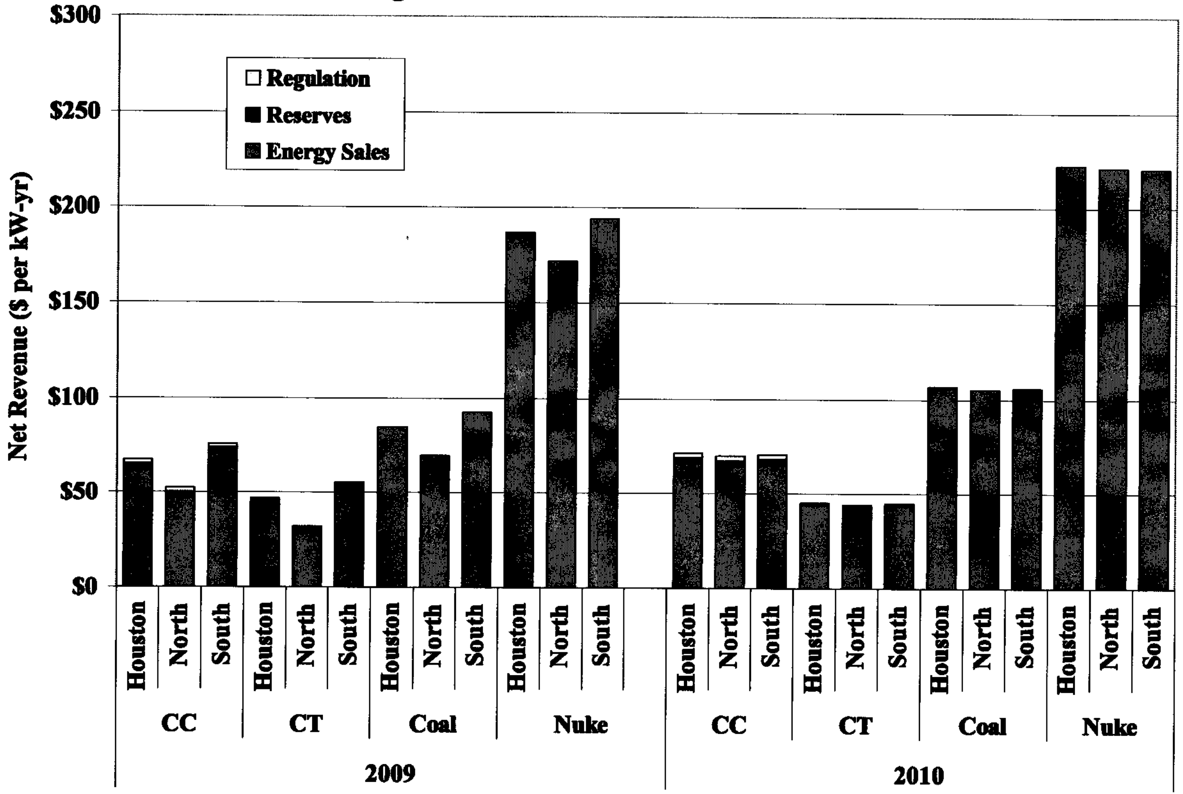
Net revenue is defined as the total revenue that can be earned by a generating unit less its variable production costs. Hence, it is the revenue in excess of short-run operating costs and is available to recover a unit's fixed and capital costs. Net revenues from the energy and ancillary services markets together provide the economic signals that inform suppliers' decisions to invest in new generation or retire existing generation. In a long-run equilibrium, markets should provide sufficient net revenue to allow an investor to break-even on an investment in a new generating unit. In the short-run, if the net short-run revenues produced by the market are not sufficient to justify entry, then one or more of three conditions exist:

- New capacity is not needed because there is sufficient generation already available;
- Load levels, and thus energy prices, are temporarily low due to mild weather or economic conditions; or
- Market rules are causing revenues to be reduced inefficiently.

Likewise, the opposite would be true if the markets provide excessive net revenues in the short-run. The persistence of excessive net revenues in the presence of a capacity surplus is an indication of competitive issues or market design flaws. In this section, we analyze the net revenues that would have been received by various types of generators in each zone.

Figure 33 shows the results of the net revenue analysis for four types of units in 2009 and 2010. These are: (a) a gas combined-cycle, (b) a combustion turbine, (c) a coal unit, and (d) a nuclear unit. For the gas-fired technologies, net revenue is calculated by assuming the unit will produce energy in any hour for which it is profitable and by assuming it will be available to sell reserves and regulation in other hours that it is available (*i.e.*, when it is not experiencing a planned or forced outage). For coal and nuclear technologies, net revenue is calculated by assuming that the unit will produce at full output. The energy net revenues are computed based on the balancing energy price in each hour.[13] Although most suppliers would receive the bulk of their revenues through bilateral contracts, the spot prices produced in the balancing energy market should drive the bilateral energy prices over time and are appropriate to use for this evaluation.

Figure 33: Estimated Net Revenue



For purposes of this analysis, we assume heat rates of 7 MMBtu per MWh for a combined cycle unit, 10.5 MMBtu per MWh for a combustion turbine, and 9.5 MMBtu per MWh for a new coal unit. We assume variable operating and maintenance costs of $4 per MWh for the gas units and $5 per MWh for the coal unit. We assume fuel and variable operating and maintenance costs of $8 per MWh for the nuclear unit. For each technology, we assumed a total outage rate (planned and forced) of 10 percent.

Some units, generally those in unique locations that are used to resolve local transmission constraints, also receive a substantial amount of revenue through uplift payments (*i.e.*, Out-of-

[13] The December 2010 net revenue calculation uses generation weighted settlement point prices. The generation units are mapped into 2010 CSC zones.

Merit Energy, Out-of-Merit Capacity, and Reliability Must Run payments). This source of revenue is not considered in this analysis. The analysis also includes simplifying assumptions that can lead to over-estimates of the profitability of operating in the wholesale market. The following factors are not explicitly accounted for in the net revenue analysis: (i) start-up costs, which can be significant; and (ii) minimum running times and ramp restriction, which can prevent the natural gas generators from profiting during brief price spikes. Despite these limitations, the net revenue analysis provides a useful summary of signals for investment in the wholesale market.

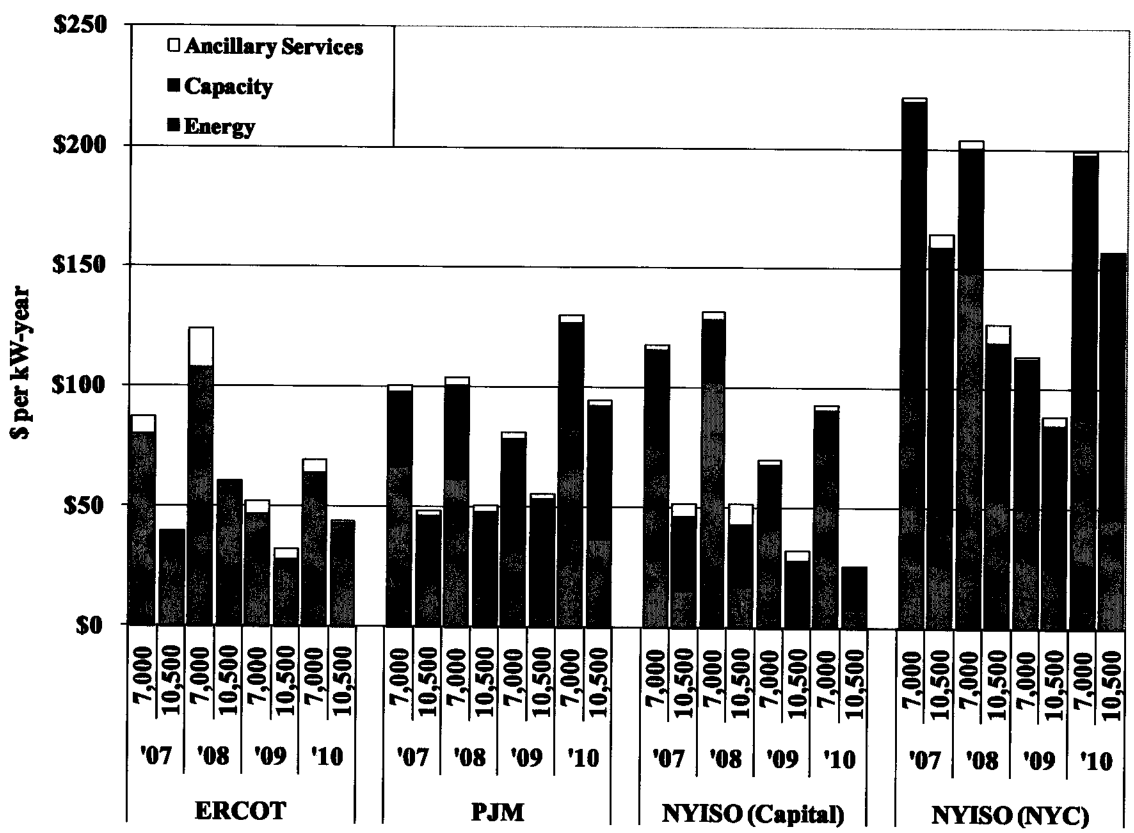
Figure 33 shows that the net revenue generally increased in 2010 compared to each zone in 2009. Based on our estimates of investment costs for new units, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from $80 to $105 per kW-year. The estimated net revenue in 2010 for a new gas turbine was approximately $45 per kW-year. For a new combined cycle unit, the estimated net revenue requirement is approximately $105 to $135 per kW-year. The estimated net revenue in 2010 for a new combined cycle unit was approximately $70 per kW-year. These values indicate that the estimated net revenue in 2010 was well below the levels required to support new entry for a new gas turbine or a combined cycle unit in the ERCOT region.

Prior to 2005, net revenues were well below the levels necessary to justify new investment in coal and nuclear generation. However, high natural gas prices through 2008 allowed energy prices to remain at levels high enough to support new entry for these technologies. The production costs of coal and nuclear units did not change significantly over this period, leading to a dramatic rise in net revenues. That view has now changed with the relatively lower natural gas prices experienced in 2009 and 2010. For a new coal unit, the estimated net revenue requirement is approximately $210 to $270 per kW-year. The estimated net revenue in 2010 for a new coal unit was approximately $105 per kW-year. For a new nuclear unit, the estimated net revenue requirement is approximately $280 to $390 per kW-year. The estimated net revenue in 2010 for a new nuclear unit was approximately $221 per kW-year. These values indicate that the estimated net revenue for either a new coal or a nuclear unit in ERCOT was well below the levels required to support new entry in 2010.

Although estimated net revenue once again is below levels to support new investment, the nodal market design will have an effect on the profitability of new resources. In a particular location, nodal prices could be higher or lower than the prices in the current market depending on the pattern of congestion.

To provide additional context for the net revenue results presented in this section, we also compared the net revenue for two types of natural gas-fired technologies in the ERCOT market with the net revenue in other wholesale markets with centrally cleared capacity markets. Figure 34 compares estimates of net revenue for the ERCOT North Zone, PJM, and two locations within the New York ISO. The figure includes estimates of net revenue from energy, reserves and regulation, and capacity. ERCOT does not have a capacity market, and thus, does not have any net revenue from capacity sales.

Figure 34: Comparison of Net Revenue of Gas-Fired Generation between Markets



The figure above shows that net revenues increased from 2009 to 2010 for both technologies in all markets. The only exception was gas peaking units in the NYISO Capital zone, where net revenue decreased slightly. In the figure above, net revenues are calculated for central locations. However, there are load pockets within each market where net revenue and the cost of new investment may be higher. Thus, even if new investment is not generally profitable in a market, it may be economic in certain areas. Finally, resource investments are driven primarily by forward price expectations, so historical net revenue analyses do not provide a complete picture of the future pricing expectations that will spur new investment.

E. Effectiveness of the Scarcity Pricing Mechanism

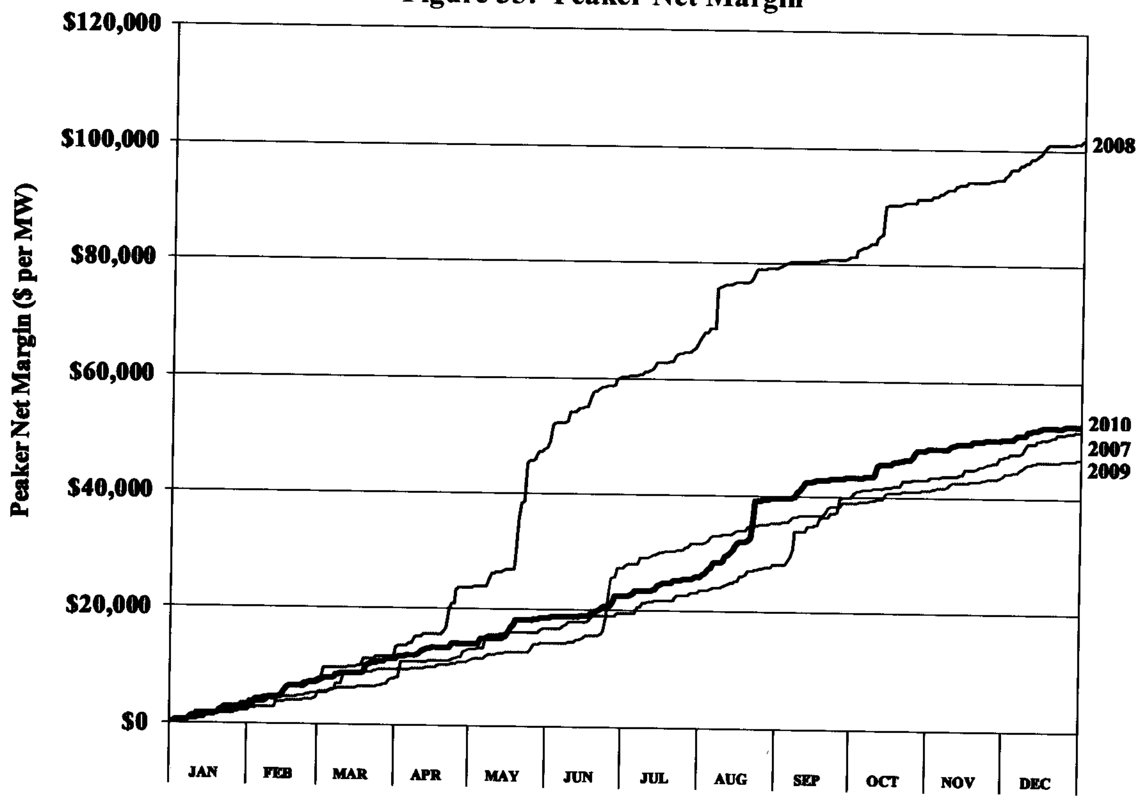
The Public Utility Commission of Texas ("PUCT") adopted rules in 2006 that define the parameters of an energy-only market. These rules include a Scarcity Pricing Mechanism ("SPM") that relaxed the existing system-wide offer cap by gradually increasing it to $1,500 per MWh on March 1, 2007, $2,250 per MWh on March 1, 2008, and to $3,000 per MWh shortly after the implementation of the nodal market. Additionally, market participants controlling less than five percent of the capacity in ERCOT by definition do not possess market power under the PUCT rules. Hence, these participants can submit very high-priced offers that, per the PUCT rule, will not be deemed to be an exercise of market power. However, because of the competition faced by the small market participants, the quantity offered at such high prices – if any – is very small.

PUCT SUBST. R. 25.505 provides that the IMM may conduct an annual review of the effectiveness of the SPM. This subsection provides an assessment of the effectiveness of the SPM in 2010 under ERCOT's energy-only market structure.

Unlike markets with a long-term capacity market where fixed capacity payments are made to resources across the entire year regardless of the relationship of supply and demand, the objective of the energy-only market design is to allow energy prices to rise significantly higher during legitimate shortage conditions (*i.e.*, when the available supply is insufficient to simultaneously meet both energy and minimum operating reserve requirements) such that the appropriate price signal is provided for demand response and new investment when required. During non-shortage conditions (*i.e.*, most of the time), the expectation of competitive energy market outcomes is no different in energy-only than in capacity markets.

Hence, in an energy-only market, expectations about both the magnitude of the energy price during shortage conditions and the frequency of shortage conditions are what will attract new investment when required. In other words, the higher the price during shortage conditions, the fewer shortage conditions that are required to provide the investment signal, and vice versa. While the magnitude of price expectations is determined by the PUCT energy-only market rules, it remains an empirical question whether the frequency of shortage conditions over time will be optimal such that the market equilibrium produces results that satisfy the reliability planning requirements (*i.e.*, the maintenance of a minimum 13.75 percent planning reserve margin).

Figure 35: Peaker Net Margin



The SPM includes a provision termed the Peaker Net Margin ("PNM") that is designed to measure the annual net revenue of a hypothetical peaking unit. Under the rule, if the PNM for a year reaches a cumulative total of $175,000 per MW, the system-wide offer cap is then reduced to the higher of $500 per MWh or 50 times the daily gas price index. Figure 35 shows the cumulative PNM results for each year from 2007 through 2010.[14] As previously noted, the net

14 The proxy combustion turbine in the Peaker Net Margin calculation assumes a heat rate of 10 MMBtu per

revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit is approximately $80 to $105 per kW-year (i.e., $80,000 to $105,000 per MW-year).

Thus, as shown in Figure 35 and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in only one of the last four years. In 2008, the peaker net margin and net revenue values rose substantially, surpassing the level required to support new peaker entry. However, a significant portion of the net revenue increase in 2008 was associated with extremely inefficient transmission congestion management and inefficient pricing mechanisms associated with the deployment of non-spinning reserves.[15] Both of these issues were corrected in the zonal market and were further improved with the implementation of the nodal market in late 2010. With these issues addressed, the peaker net margin dropped substantially in 2009 and 2010.

In our review of the effectiveness of the SPM in 2010 improvement was observed in two areas of concern raised last year: (1) ERCOT's day-ahead load forecast bias and (2) appropriate energy pricing during the deployment of non-spinning reserves. However, we found the dependence upon market participants to submit offers at or near the offer cap to not be a reliable means of producing scarcity pricing during shortage conditions under the zonal market design.

1. ERCOT Day-Ahead Load Forecast Error

ERCOT's zonal market includes the operation of a day-ahead Replacement Reserve Service ("RPRS") market that is designed to ensure that adequate capacity is available on the system to meet reliability criteria for each hour of the following operating day. This includes an assessment of the capacity necessary to meet forecast demand and operating reserve requirements, as well as capacity required to resolve transmission constraints.

A key input to the RPRS market is the day-ahead load forecast developed by ERCOT. If the day-ahead load forecast is significantly below actual load and no subsequent actions are taken, ERCOT may run the risk of there not being enough generating capacity online to meet reliability

MWh and includes no other variable operating costs or startup costs.

15 *See* 2008 ERCOT SOM Report at 81-87.

criteria in real-time. In contrast, if the day-ahead load forecast is significantly higher than actual load, the outcome may be an inefficient commitment of excess online capacity in real-time.

Figure 36: Average Day-Ahead Load Forecast Error by Month and Hour Blocks

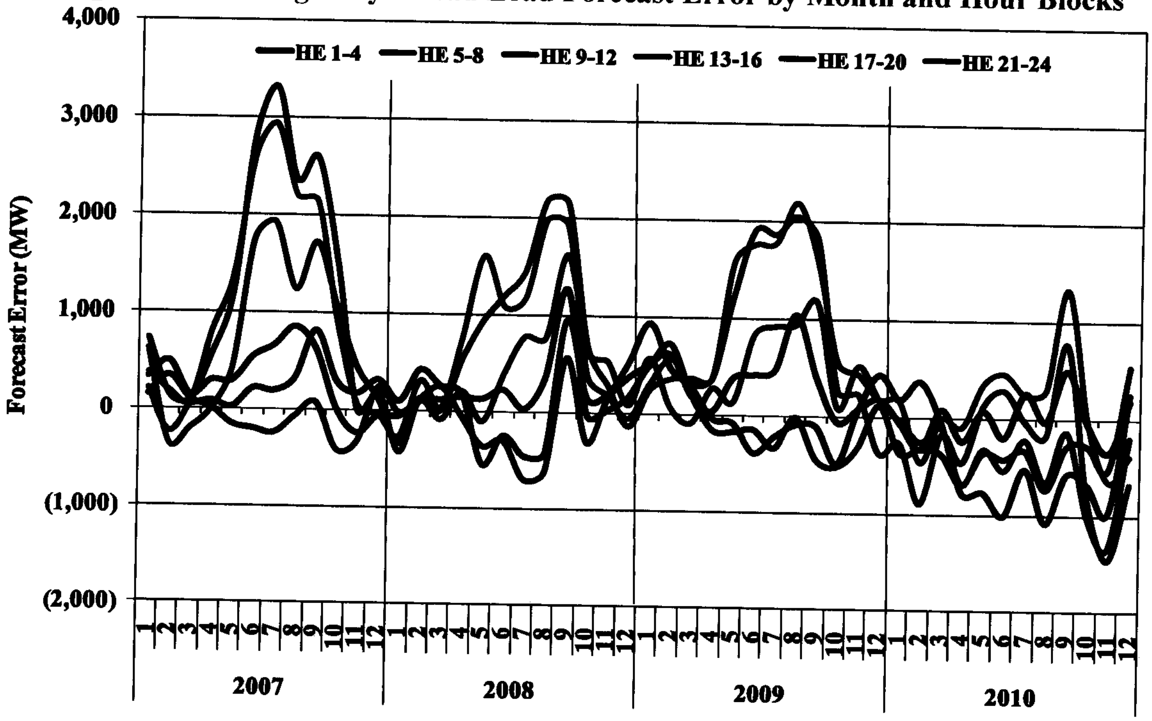
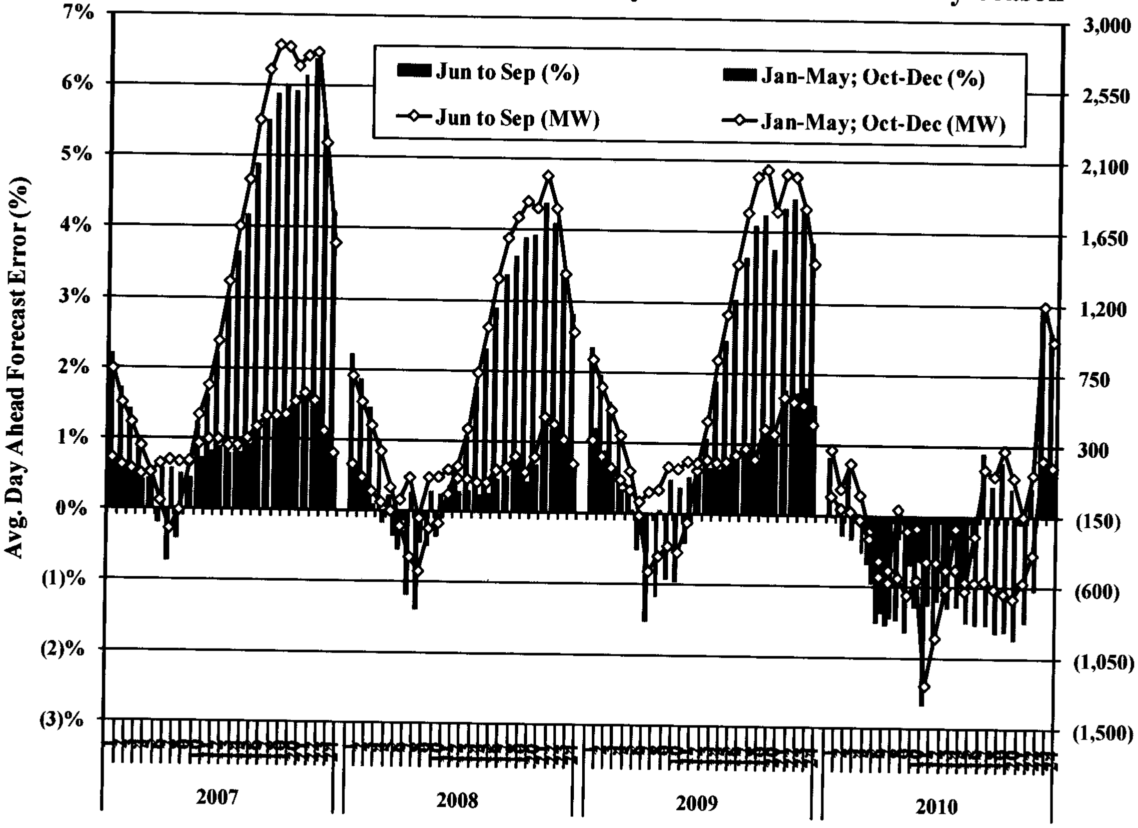


Figure 36 shows the day-ahead load forecast error data for 2007 through 2010 with the average megawatt error displayed for each month in four hour blocks (hours ending). This figure shows a change of pattern from significant over-forecasting during peak load hours to much smaller errors in 2010.

Figure 37 shows the average hourly day-ahead load forecast error for the summer months of June through September, and also for the months of January through May and October through December for 2007 through 2010. In this figure, positive values indicate a day-ahead load forecast that was greater than the actual real-time load. These data indicate a much reduced positive bias (*i.e.*, over-forecast) in the day-ahead load forecast over most of the house hours in 2010 when compared to previous years.

Figure 37: Average Day-Ahead Hourly Load Forecast Error by Season



The 2010 ERCOT ancillary service procurement methodologies was modified to adjust the ERCOT day-ahead load forecast to account for the historically measured net load forecast bias, and to compensate for this adjustment by increasing the quantity of non-spinning reserves procured. The revised procedures went into effect in January 2010 and the effect is clearly observable in the data shown in Figure 37 when comparing 2010 to the three prior years.

2. Improved energy pricing during deployment of Non-Spin

The following improvements related to the deployment and pricing of non-spinning reserves were implemented in May 2009:[16]

- Eliminate the previous *ex post* re-pricing provisions to provide for *ex ante* pricing during non-spinning reserve deployments, thereby providing more pricing certainty for resources and loads and significantly reducing the probability of *ex post* scarcity level prices during non-scarcity conditions;

[16] See Protocol Revision Request 776.

- Allow quick start units providing non-spinning reserves to offer in the balancing energy market at a market-based price reflecting the cost and risks of starting and deploying these resources; and
- Reduce the probability of transitional shortages by providing more timely access to these reserves through the balancing energy market instead of manual operator deployments.

With the increased quantity of non-spinning reserves being procured in 2010, the increased efficiencies in market operations and pricing during the deployment of non-spinning reserves became even more important.

Non-spinning reserves are comprised mostly of off-line natural gas fired combustion turbines capable of starting in 30 minutes or less. Although the implementation of the nodal market has significantly increased market efficiencies in a number of areas, including the move to a five minute rather than 15 minute energy dispatch, the initial implementation lacked an efficient economic commitment mechanism for resources such as offline gas turbines and other resources that are not immediately dispatchable in the five minute energy dispatch. The current mechanisms result in prices that are inefficiently low because they are not representative of the costs associated with starting and running the gas turbines that are being deployed to meet demand.

As previously recommended, this deficiency in ERCOT's nodal market design should be addressed by implementing an additional energy market "look ahead" dispatch functionality to produce a projected unit dispatch with energy and ancillary services co-optimized.[17] This additional functionality represents a major change to ERCOT systems, which may not be able to be implemented for several years. However, because the market inefficiencies associated with the current mechanism are significant, we recommend that an interim solution be pursued that can be implemented in the near term that will more reasonably reflect the marginal costs of the actions being taken when non-spinning reserves are deployed and necessary to meet demand.

3. Dependence on High-Priced Offers by Market Participants

As a general principle, competitive and efficient market prices should be consistent with the marginal cost of the marginal action taken to satisfy the market's demand. In the vast majority

[17] See Direct Testimony of David B. Patton, PUCT Docket No. 31540 at pages 35-41.

of hours, the marginal cost of the marginal action is that associated with the dispatch of the last generator required to meet demand. It is appropriate and efficient in these hours for this generator to "set the price." However, this is not true under shortage conditions. When the system is in shortage, the demand for energy and minimum operating reserves cannot be satisfied with the available resources, which will cause the system operator to take one or more of the following actions:

- Sacrifice a portion of the operating reserves by dispatching them for energy;
- Voluntarily curtail load through emergency demand response programs;
- Curtail exports or make emergency imports; or
- Involuntarily curtail load.

A market design that adheres to the pricing principles stated above will set prices that reflect each of these actions. When the market is in shortage, the marginal action taken by the system operator is generally to not satisfy operating reserves requirements (*i.e.*, dispatching reserves for energy). Diminished operating reserves results in diminished reliability, which has a real cost to electricity consumers. In this case, the value of the foregone reserves – which is much higher than the marginal cost of the most expensive online generator – should be reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response and transmission.

Under the PUCT rules governing the energy-only market, the mechanism that allows for such pricing during shortage conditions relies upon the submission of high-priced offers by small market participants.

Figure 38: Balancing Energy Market Prices during Shortage Intervals

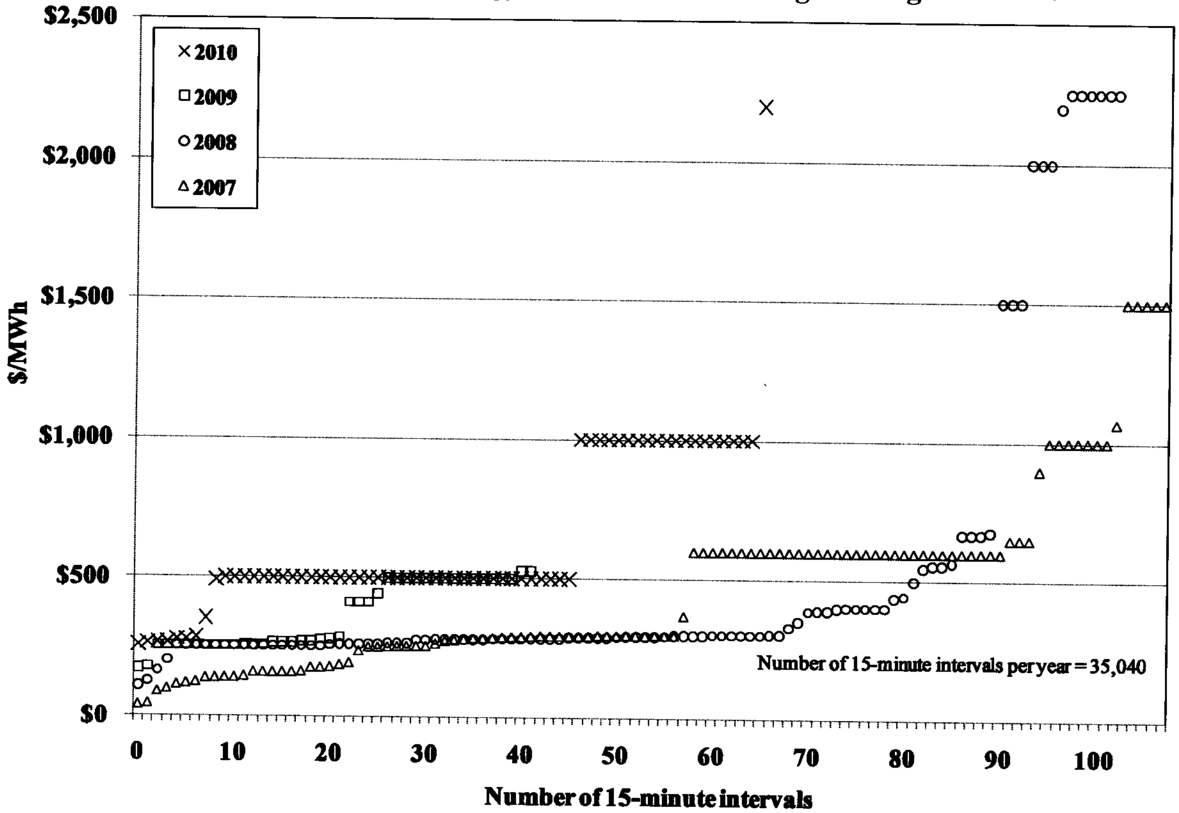
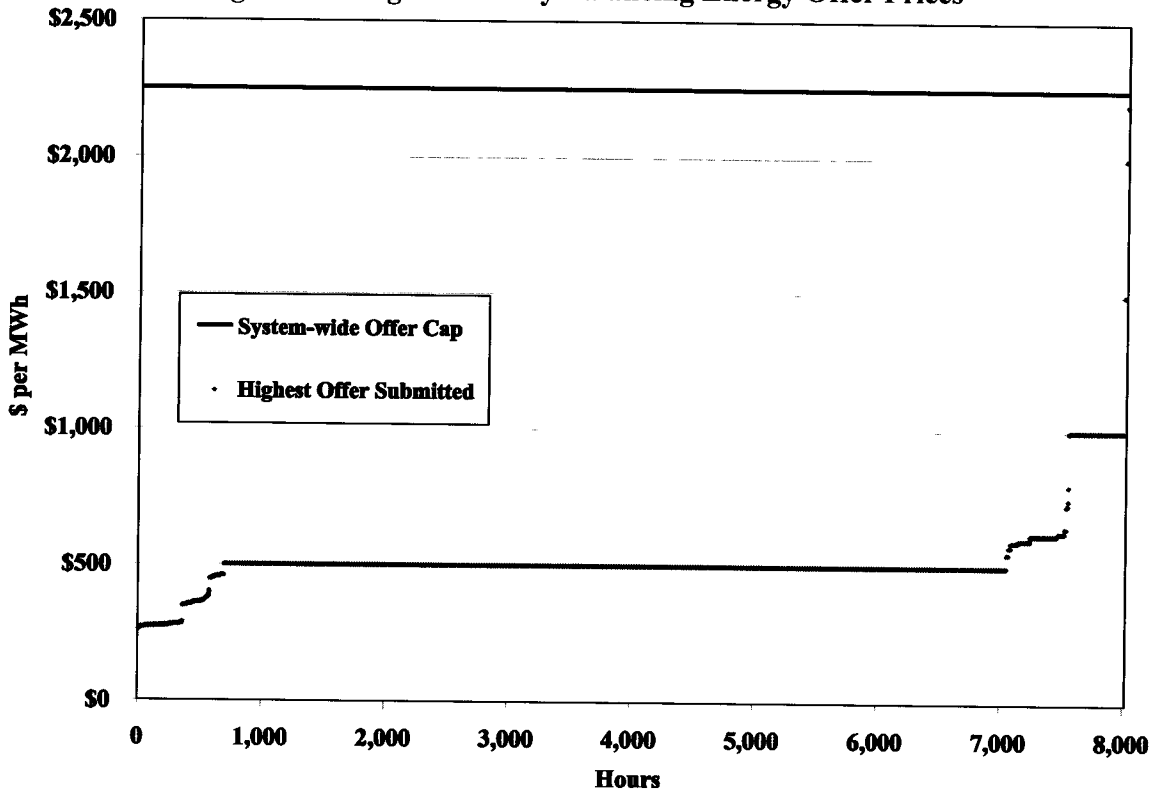


Figure 38 shows the balancing market clearing prices during the 15-minute shortage intervals from 2007 through November 2010. The 66 shortage intervals for the first eleven months of 2010 are significantly fewer than the 108 and 103 shortage intervals that occurred in 2007 and 2008, respectively, but more than the 42 intervals that occurred in 2009. Although each of the data points in Figure 38 represents system conditions in which the market was in shortage, the pricing outcomes are widely varied, with the majority of prices reflecting the marginal offer of the most expensive generation resource dispatched as opposed to the value of foregone operating reserves. Had an offer been submitted that established the MCPE at the system-wide offer cap in each of the 66 shortage intervals of zonal market operations, the 2010 annual peaker net margin would have increased from $52,491 to $79,009 per MW-year, an increase of over 50 percent. The associated increase in the annual load-weighted average balancing energy price would have been less significant, increasing from $39.40 to $43.27 per MWh, a 9.8 percent increase.

These results indicate that relying exclusively upon the submission of high-priced offers by market participants was generally not a reliable means of producing efficient scarcity prices during shortage conditions under ERCOT's zonal market. In fact, although the system-wide offer cap was $2,250 per MWh, there was only one interval in 2010 when an offer submitted by a market participant set the clearing price at greater than $1000 per MWh. Figure 39 shows the highest balancing energy offer price submitted by all market participants in each hour of 2010 (through November), ranked from lowest to highest.

Figure 39: Highest Hourly Balancing Energy Offer Prices



This figure shows that there were only 451 hours (5.6 percent) with an offer that exceeded $1,000 per MWh, and the average of the highest offers submitted by any market participant across all hours in 2010 was $520 per MWh.

More reliable and efficient shortage pricing could be achieved by establishing pricing rules that automatically produce scarcity level prices when operating reserve shortages exist. Such an approach would be more reliable because it would not be dependent upon the submission of high-priced offers by small market participants to be effective. It would also be more efficient during the greater than 99 percent of time in which shortage conditions do not exist because it would not be necessary for market participants to effectively withhold lower cost resources by offering relatively small quantities at prices dramatically higher than their marginal cost.

This type of mechanism is part of the nodal market design. At times when there is insufficient capacity available to meet both energy and minimum operating reserve requirements, all available capacity will be dispatched and the clearing price will rise in a predetermined manner to a maximum of the system-wide offer cap. During December 2010 there were fifteen executions of the security constrained economic dispatch (SCED) algorithm that resulted in the system-wide energy clearing price being set at the system-wide offer cap. These fifteen SCED intervals represented thirty-two minutes spread over five settlement intervals.

III. Transmission and Congestion

One of the most important functions of any electricity market is to manage the flows of power over the transmission network by limiting additional power flows over transmission facilities when they reach their operating limits. Under ERCOT's zonal market design, constraints on the transmission network are managed in two ways. First, ERCOT is made up of zones with the constraints between the zones managed through the balancing energy market. The balancing energy market model increases energy production in one zone and reduces it in another zone to manage the flows between the two zones when the interface constraint is binding, *i.e.*, when there is interzonal congestion. Second, all other constraints not defined as zonal constraints (*i.e.*, local congestion) are managed through the redispatch of individual generating resources. In this section of the report we review the ERCOT transmission system usage and analyze the costs and frequency of transmission congestion.

A. Electricity Flows between Zones

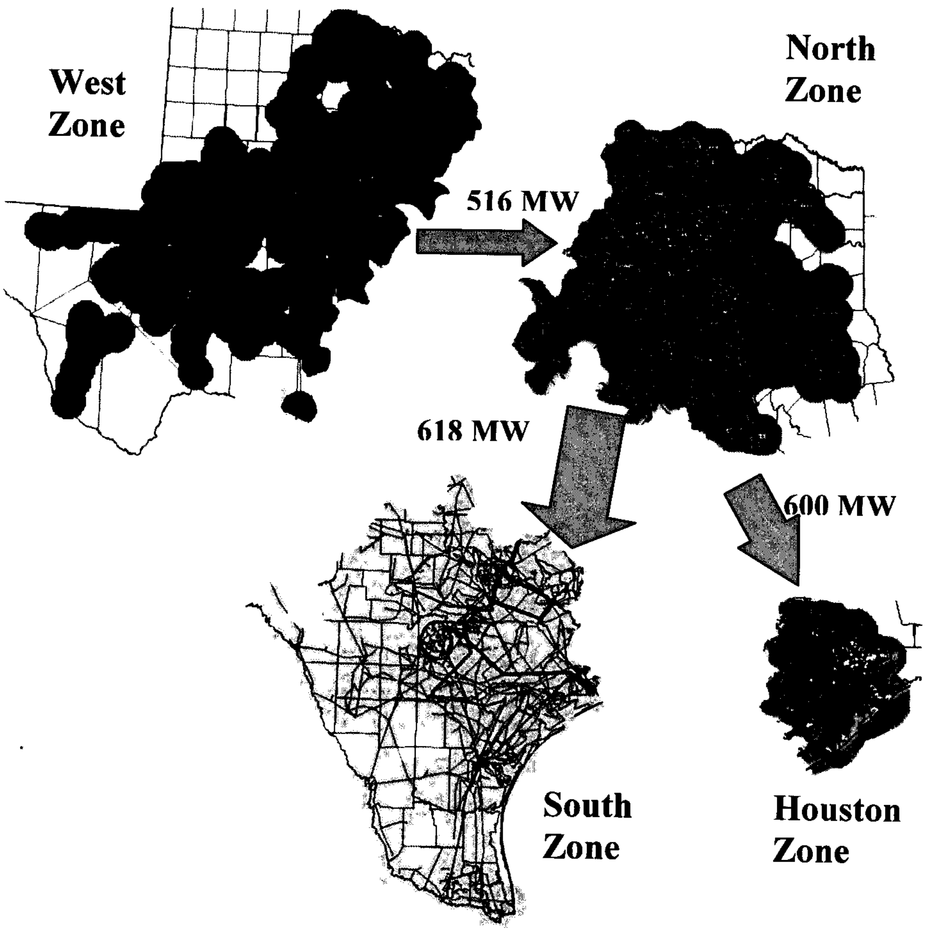
In 2010 there were four commercial pricing zones in ERCOT: (a) the North Zone, (b) the West Zone, (c) the South Zone, and (d) the Houston Zone. Under ERCOT's zonal market design, operators use the Scheduling, Pricing and Dispatch ("SPD") software to economically dispatch balancing energy in each zone to serve load and manage congestion between zones. The SPD model embodies the market rules and requirements documented in the ERCOT protocols.

To manage interzonal congestion, SPD uses a simplified network model with four zone-based locations and five transmission interfaces. These five transmission interfaces, referred to as Commercially Significant Constraints ("CSCs"), are simplified representations of groups of transmission elements. ERCOT operators use planning studies and real-time information to set limits for each CSC that are intended to utilize the total transfer capability of the CSC. In this subsection of the report, we describe the SPD model's simplified representations of flows between zones and analyze actual flows in 2010.

The SPD model uses zonal approximations to represent complex interactions between generators, loads, and transmission elements. Because the model flows are based on zonal approximations, the estimated flows can depart significantly from real-time physical flows.

Estimated flows that diverge significantly from actual flows are an indication of inaccurate congestion modeling leading to inefficient energy prices and other market costs. This subsection analyzes the impact of SPD transmission flows and constraints on market outcomes.

Figure 40: Average SPD-Modeled Flows on Commercially Significant Constraints During All Intervals in 2010



Note: In the figure above, CSC flows are averaged taking the direction into account. So one arrow shows the average flow for the West-to-North CSC was 516 MW, which is equivalent to saying that the average for the North-to-West CSC was *negative* 516 MW.

Figure 40 shows the four ERCOT geographic zones as well as the five CSCs that interconnect the zones: (a) the West to North interface, (b) the South to North interface, (c) the North to South interface, (d) the North to Houston interface, and (e) the North to West interface. A single arrow is shown for the modeled flows of both the North to West and West to North CSCs and the South to North and North to South CSCs. Based on average SPD modeled flows, the North Zone exports a significant amount of power.

The most important simplifying assumption underlying the zonal model is that all generators and loads in a zone have the same effect on the flows over the CSC, or the same shift factor in relation to the CSC.[18] In reality, the generators and loads within each zone can have widely differing effects on the flows over a CSC. To illustrate this, we compared the flows calculated by using actual generation and zonal average shift factors to the average actual flows that occurred over each CSC. The flows over the North to West and South to North CSCs are not shown separately in the table below since they are equal and opposite the flows for the West to North CSC and North to South CSCs, respectively.

Table 2: Average Calculated Flows on Commercially Significant Constraints
Zonal-Average vs. Nodal Shift Factors

CSC 2010	Flows Modeled by SPD (1)	Flows Calculated Using Actual Generation (2)	*Difference* = (2) - (1)	Actual Flows Using Nodal Shift Factors (3)	*Difference* = (3) - (2)
West - North	516	408	*-108*	419	*11*
North - Houston	600	479	*-121*	775	*296*
North - South	618	517	*-101*	189	*-328*

The first column in Table 2 shows the average flows over each CSC calculated by SPD. The second column shows the average flows over each CSC calculated by using zonal-average shift factors and actual real-time generation in each zone instead of the scheduled energy and balancing energy deployments used as an input in SPD. Although these flows are both calculated using the same zonal-average shift factors, they can differ when the actual generation varies from the SPD generation. This difference is shown in the third column (in italics). These differences indicate that the actual generation levels result in calculated flows on each CSC that were lower than the flows modeled by SPD.

18 For a generator, a shift factor indicates the portion of the incremental output of a unit that will flow over a particular transmission facility. For example, a shift factor of 0.5 would indicate that half of any incremental increase in output from a generator would flow over the interface. A negative shift factor would indicate a decrease in flow on an interface resulting from an increase in generation.

The fourth column in Table 2 reports the actual average flows over each CSC by using nodal shift factors applied to actual real-time generation and load. The difference in flows between columns (3) and (2) is attributable to using zonal average shift factors versus nodal shift factors for generation and load in each zone. These differences in flows are shown in the fifth column (in italics).

These results show that the heterogeneous effects of generators and load in a zone on the CSC flows can cause the actual flows to differ substantially from the SPD-calculated flows. Table 2 shows that by using nodal (actual) shift factors reduced the calculated flows on the North to South interface by 328 MW and increased the calculated flows on the North to Houston CSC by 296 MW.

The use of simplified generation-weighted shift factors limits the SPD model from efficiently resolving and assigning the costs of interzonal congestion. In the long run, the use of generation-weighted shift factors for loads systematically biases prices, so that buyers in some zones pay too much, and others pay too little. Further, the use of average zonal shift factors creates significant operational challenges for ERCOT in the real-time management of zonal congestion because the response to zonal dispatch instructions can often affect the actual flow on a CSC in a manner that is significantly different than that calculated by the simplified assumptions in the SPD model. In turn, ERCOT will tend to operate the system more conservatively to account for the operational uncertainties introduced by the simplified assumptions in the SPD model, the effect of which is discussed in more detail later in this section.

To provide additional understanding of the electricity flows between zones prior to discussing the details of interzonal congestion in the next subsection, Figure 41 shows the actual average imports of power for each zone in 2010. In this figure, positive values represent imports and negative values indicate exports.

Figure 41: Actual Zonal Net Imports

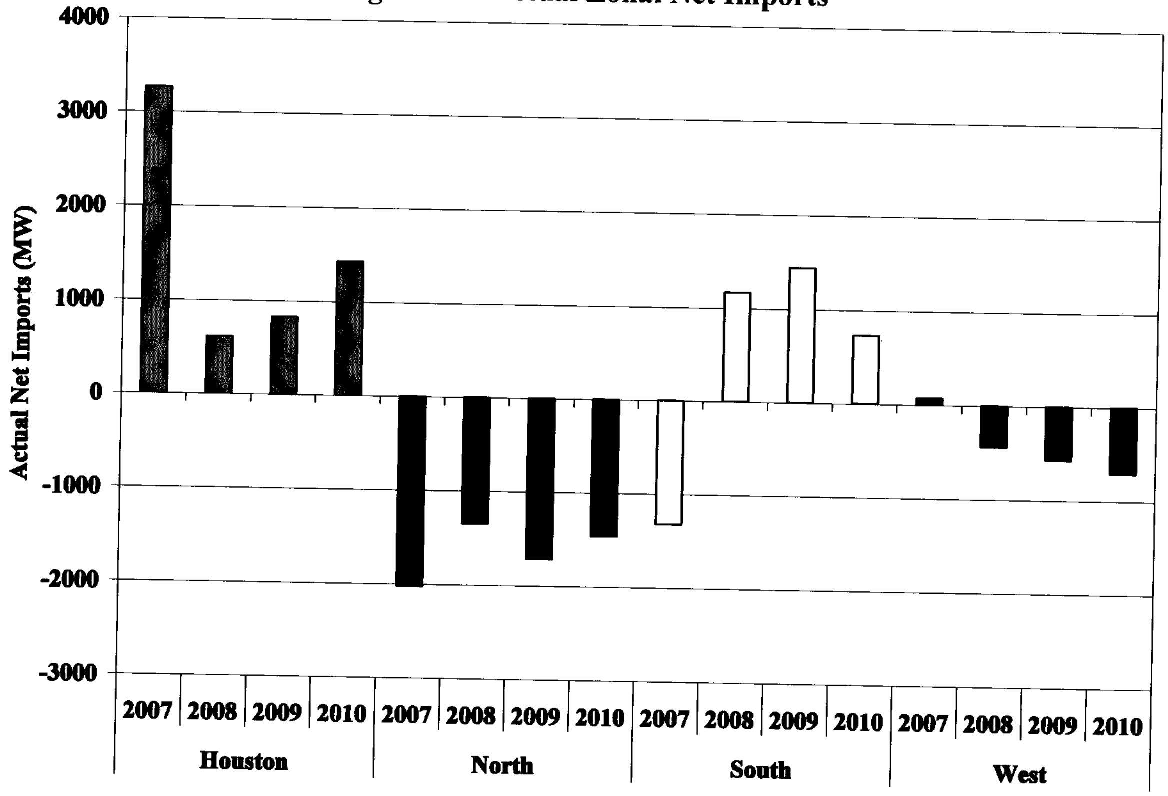


Figure 41 shows that the Houston Zone is a net importer of power, while the North Zone is a net exporter. The reduction in the Houston Zone imports in 2008, 2009 and 2010 and corresponding change in the South Zone from a net exporter to a net importer can be attributed to the movement of the 2,700 MW South Texas Nuclear Project from the South Zone to the Houston Zone in 2008. The West Zone transitioned from a net importer in 2007 to a net exporter in 2008, 2009 and 2010. This reflects the significant increases in the installed capacity of wind resources in the West Zone that occurred over this time period.

B. Interzonal Congestion

The prior subsection showed the average interzonal flows calculated by SPD compared to actual flows in all hours. This subsection focuses on those intervals when the interzonal constraints were binding. Although this excludes most intervals, it is during these constrained intervals that the performance of the market is most critical.

Figure 42 shows the average SPD-calculated flows between the four ERCOT zones during constrained periods for the five CSCs. The arrows show the average magnitude and direction of the SPD-calculated flows during constrained intervals. The frequency with which these constraints arise is shown in parentheses.

Figure 42: Average SPD-Modeled Flows on Commercially Significant Constraints During Transmission Constrained Intervals in 2010

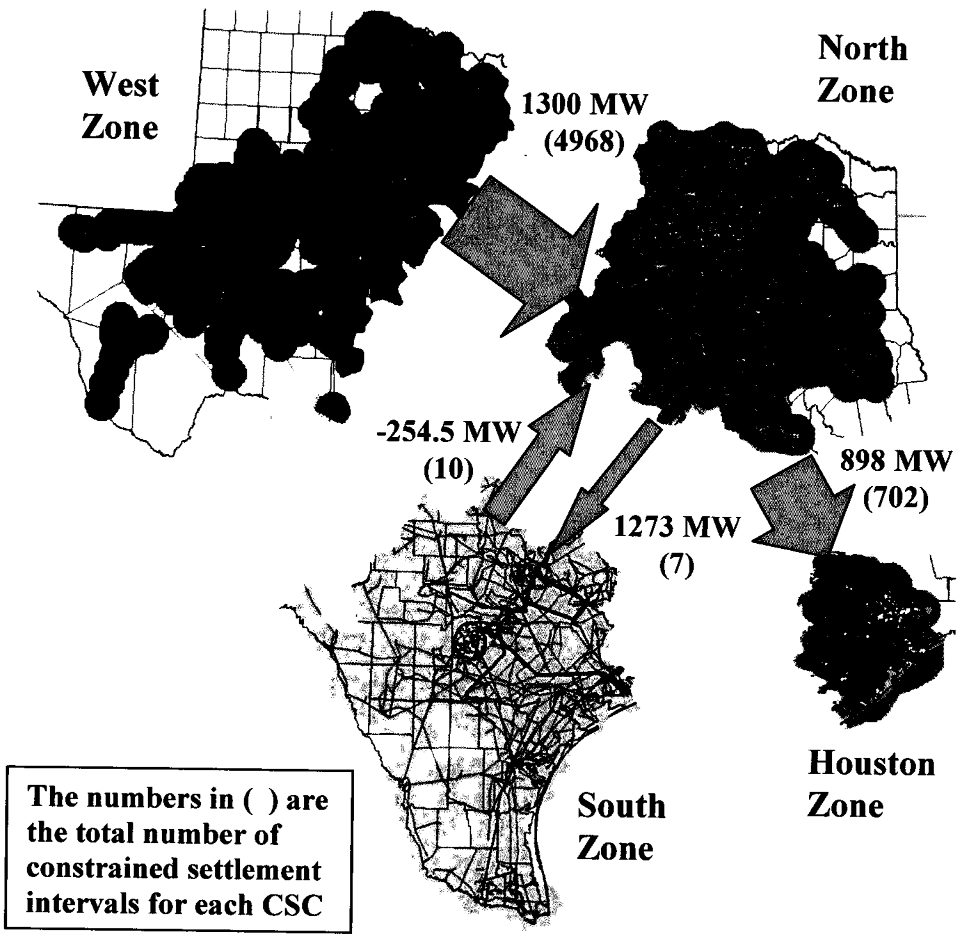


Figure 42 shows that inter-zonal congestion was most frequent in 2010 on the West to North and the North to Houston CSCs, followed by the South to North CSC. The West to North CSC exhibited SPD-calculated flows averaging 1300 MW during 4,968 constrained 15-minute intervals (15 percent of the totals intervals in the year). The SPD-calculated average flow for the

North to Houston CSC was 898 MW during 702 constrained intervals (2 percent of the total intervals). The North to South and South to North CSCs were constrained very infrequently.

Table 3: Average Calculated Flows on Commercially Significant Constraints during Transmission Constrained Intervals

Zonal-Average vs. Nodal Shift Factors

CSC 2010	Flows Modeled by SPD (1)	Flows Calculated Using Actual Generation (2)	*Difference = (2) - (1)*	Actual Flows Using Nodal Shift Factors (3)	*Difference = (3) - (2)*
West - North	1300	1352	*52*	1563	*211*
North - Houston	898	628	*-270*	1132	*504*
North - South	1273	1302	*29*	955	*-347*
South - North	-254	-330	*-76*	355	*685*

Table 3 shows data similar to that presented in Table 2, except that the data in Table 3 is limited to only those intervals in which the CSC transmission constraint was binding. The first column in Table 3 shows the average SPD-modeled flows for each CSC. The second column shows the average flows over each CSC calculated by using zonal-average shift factors and actual real-time generation in each zone instead of the scheduled energy and balancing energy deployments. Although these flows are both calculated using the same zonal-average shift factors, they can differ when the actual generation varies from the SPD generation. This difference is shown in the third column (in italics). These differences indicate that the actual generation levels result in calculated flows on each CSC that were lower than the flows modeled by SPD. The most notable deviation was across the North to Houston CSC.

The fourth column in Table 3 reports the actual average flows over each CSC when using nodal shift factors applied to actual real-time generation and load. The difference in flows between columns (3) and (2) is attributable to using zonal average shift factors versus nodal shift factors for generation and load in each zone. These differences in flows are shown in the fifth column (in italics).

The West to North CSC was relatively close to actual flows, whereas the average actual flows for the North to South, South to North and North to Houston CSCs varied significantly from the average flows modeled by SPD.

C. Zonal Congestion Management Challenges

As discussed in the first part of this section, differences that exist between the commercial SPD model representation and the physical reality create operational challenges for ERCOT to efficiently manage zonal transmission congestion while also maintaining reliable operations. Table 4 shows the average physical limit, actual flow and the difference between the average physical limit and the actual flow for each CSC during binding intervals in 2010.

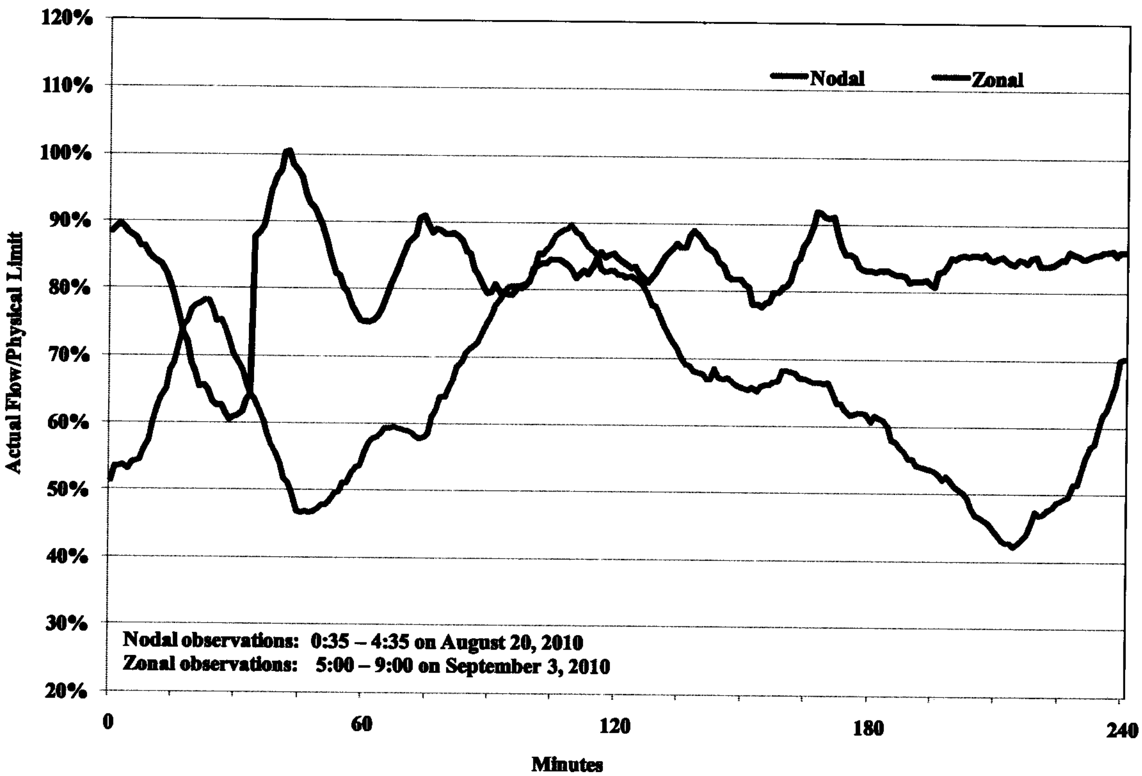
Table 4: CSC Average Physical Limits vs. Actual Flows during Constrained Intervals

CSC 2010	Average Physical Limit (MW)	Average Actual Flow (MW)	Avg. Physical Limit - Avg. Actual Flow (MW)
North to South	1155	955	200
North to Houston	2991	2497	494
South to North	580	355	225
West to North	2077	1563	514

Table 4 shows that, for all CSCs in 2010, the average actual flow was considerably less than the average physical limit. It is worth noting that the average physical limit for every CSC was higher in 2010 than last year, The largest increase being the average physical limit on West to North CSC that increased by more than 500 MW. However, even with the increase in physical limit, the average actual flow on the West to North CSC remained more than 500 MW less than the limit. For all CSCs combined, the average actual flow was 21 percent less than the average physical limit, a ratio similar to that experienced last year. To maximize the economic use of the scarce transmission capacity, the ideal outcome would be for the actual flows to reach, but not to exceed the physical limits to maintain reliable operations. However, primarily for the reasons discussed in the first part of this section, achieving such ideal outcomes is practically impossible in the context of the zonal market model. Further, the bifurcated process of resolving zonal and local congestion can at times lead to reliability conflicts that are difficult to resolve within the relatively inflexible framework of the zonal market design.

The nodal market provides many improvements, including unit-specific offers and shift factors, simultaneous resolution of all transmission congestion, actual output instead of schedule-based dispatch, and 5-minute instead of 15-minute dispatch, among others. These changes should help to increase the economic and reliable utilization of scarce transmission resources well beyond that experienced in the zonal market, and in so doing, also dispatch the most efficient resources available to reliably serve demand. Early indications of the improvement in constraint utilization expected under the nodal market are evident in Figure 43, which compares the utilization of the West to North constraint under the zonal and nodal markets. The difference in utilization across similar four hour periods is obvious. Using zonal market congestion management techniques the average utilization was 64 percent compared to 83 percent under the nodal market. Although this is much too small a sample to draw definitive conclusions, the improvement demonstrated bodes well for the improved transmission system utilization expected to occur under the nodal market design.

Figure 43: West to North Constraint Utilization



D. Congestion Rights Market

Interzonal congestion can be significant from an economic perspective, compelling the dispatch of higher-cost resources because power produced by lower-cost resources cannot be delivered over the constrained interfaces. When this constraint occurs market participants must compete to use the available transfer capability between zones. To allocate this capability efficiently, ERCOT establishes clearing prices for energy in each zone that will vary in the presence of congestion and charges the transactions between the zones the interzonal congestion price.

Under the zonal market design, one means by which ERCOT market participants can hedge congestion charges in the balancing energy market is by acquiring Transmission Congestion Rights ("TCRs") or Pre-assigned Congestion Rights ("PCRs"). Both TCRs and PCRs entitle the holder to payments corresponding to the interzonal congestion price. Hence, a participant holding TCRs or PCRs for a transaction between two zones would pay the interzonal congestion price associated with the transaction and receive TCR or PCR payments that offset the congestion charges. TCRs are acquired by annual and monthly auctions (as explained in more detail below) while PCRs are allocated to certain participants based on historical patterns of transmission usage. To analyze congestion rights in ERCOT, we first review the TCRs and PCRs that were auctioned or allocated for each CSC in 2010.

Figure 44: Transmission Rights vs. Real-Time SPD-Calculated Flows

Constrained Intervals - 2010

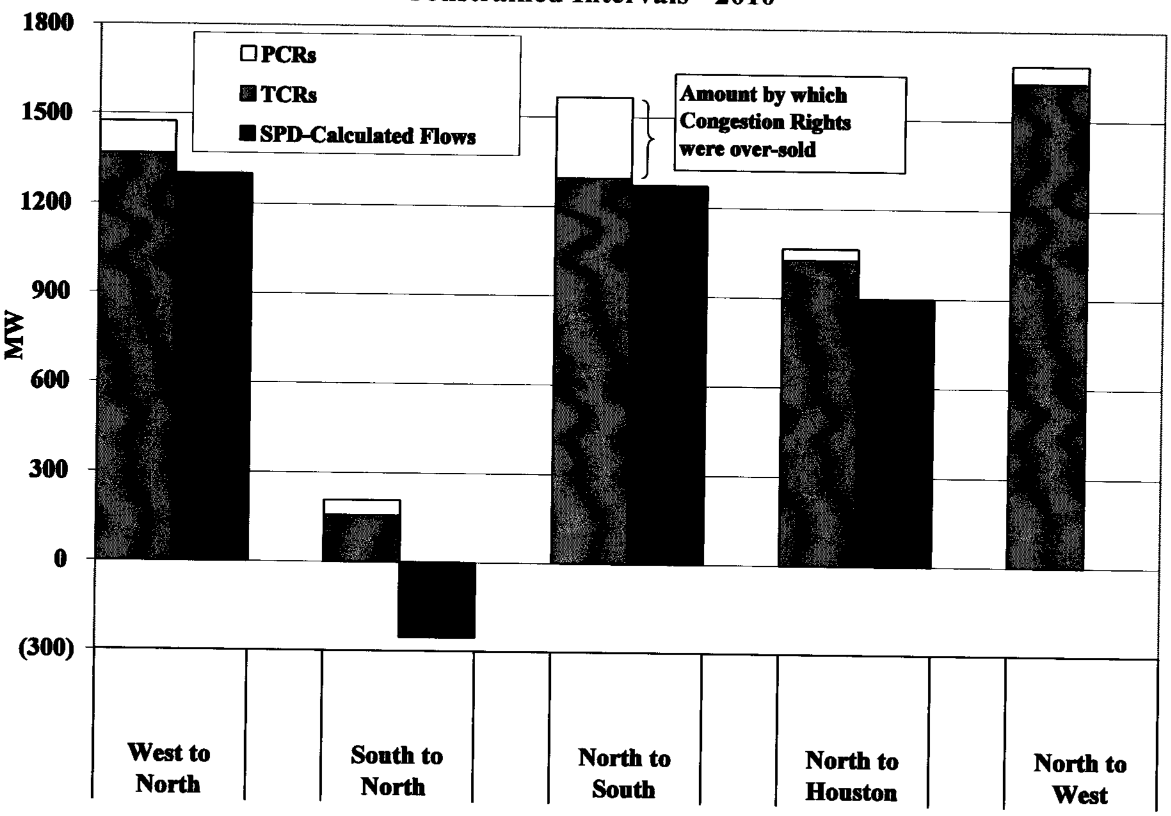


Figure 44 shows the average number of TCRs and PCRs awarded for each of the CSCs in 2010 compared to the average SPD-modeled flows during the constrained intervals. Figure 44 shows that total congestion rights (the sum of PCRs and TCRs) on all the interfaces exceeded the average real-time SPD-calculated flows during constrained intervals. For example, congestion rights for the North to Houston CSC were oversold by an average of 167 MW. These results indicate that the congestion rights were oversold in relation to the SPD-calculated limits. Although there were almost 1700 MW of PCRs and TCRs distributed for the North to West CSC in 2010, there were no instances of congestion across this interface.

As shown in Table 5, the average amount of TCRs awarded each month in 2010 is higher than in 2009 except the North to Houston CSC.

Table 5: Monthly Average Congestion Rights Awarded

	2009			2010		
	PCRs	TCRs	*Total*	PCRs	TCRs	*Total*
West - North	70	721	*791*	107	1367	*1474*
South - North	66	80	*146*	52	156	*208*
North - South	250	1091	*1341*	273	1294	*1567*
North - Houston	48	1197	*1245*	39	1027	*1066*
North - West	33	741	*774*	60	1626	*1686*

Ideally the financial obligations to holders of congestion rights would be satisfied with congestion revenues collected from participants scheduling over the interface and through the sale of balancing energy flowing over the interface. When the SPD-calculated flows are consistent with the quantity of congestion rights sold over the interface, the congestion revenues will be sufficient to satisfy payments to the holders of the congestion rights. Alternatively, when the quantity of congestion rights exceeds the SPD-calculated flow over an interface, congestion revenues from the balancing energy market will not be sufficient to meet the financial obligations to congestion rights holders.

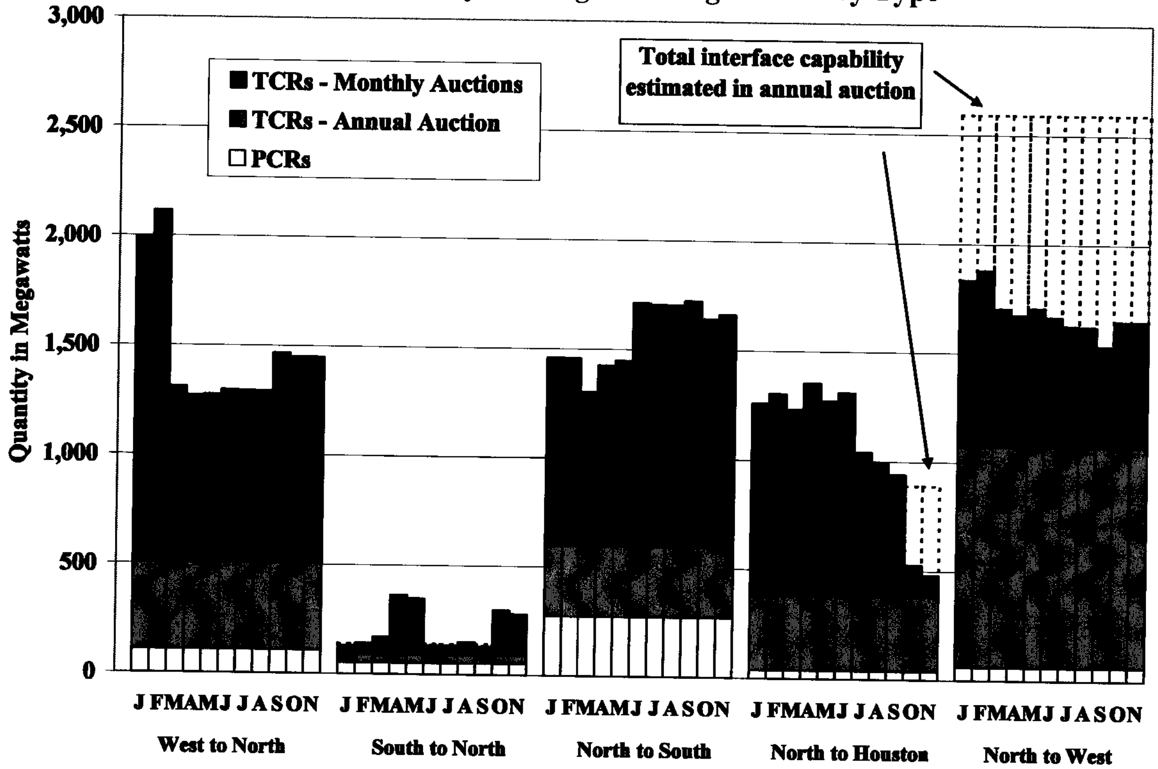
As an example, suppose the SPD-calculated flow limit is 300 MW for a particular CSC during a constrained interval and that holders of congestion rights own a total of 800 MW over the CSC. ERCOT will receive congestion rents from the balancing energy market to cover precisely 300 MW of the 800 MW worth of obligations. Thus, a revenue shortfall will result that is proportional to the shadow price of the constraint on the CSC in that interval (*i.e.*, proportional to the congestion price between the zones). In this case, the financial obligations to the congestion rights holders cannot be satisfied with the congestion revenue, so the shortfall is charged proportionately to all loads in ERCOT as part of the Balancing Energy Neutrality Adjustment ("BENA") charges.

To provide a better understanding of these relationships, we next review ERCOT's process to establish the quantity of congestion rights allocated or sold to participants. ERCOT performs studies to determine the capability of each interface under peak summer conditions. This summer planning study is the basis for offering 40 percent of the available TCRs for sale in the annual auction. These rights are auctioned during December for the coming year. Additional TCRs are offered for sale based on monthly updates of the summer study. Because the monthly studies tend to more accurately reflect conditions that will prevail in the coming month, the monthly designations tend to more closely reflect actual transmission limits.

However, the monthly studies used to designate the TCRs do not always accurately reflect real-time transmission conditions for two main reasons. First, transmission and generation outages can occur unexpectedly and can significantly reduce the transfer capability of a CSC. Even planned transmission outages may not be known to ERCOT when performing their studies. Second, conditions may arise causing the actual physical flow to be significantly different from the SPD modeled flow. As discussed above, ERCOT operators may need to respond by lowering the SPD-modeled flow limits to manage the actual physical flow. Accordingly, it is likely that the quantity of congestion rights awarded will be larger than available transmission capability in SPD.

To examine how these processes have together determined the total quantity of rights sold over each interface, Figure 45 shows the quantity of each category of congestion rights for each month during 2010. The quantities of PCRs and annual TCRs are constant across all months and were determined before the beginning of 2010, while monthly TCR quantities can be adjusted monthly.

Figure 45: Quantity of Congestion Rights Sold by Type



When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 45, the total quantity of PCRs and TCRs sold in the annual and monthly auctions equaled or exceeded the summer estimate and therefore no excess capability is shown.

Market participants who are active in congestion rights auctions are subject to substantial uncertainty. Outages and other contingencies occur randomly and can substantially change the market value of a congestion right. Real-time congestion prices reflect the cost of interzonal congestion and are the basis for congestion payments to congestion rights holders. In a perfectly efficient system with perfect forecasting by participants, the average congestion price should equal the auction price. However, we would not expect full convergence in the real-world, given uncertainties and imperfect information. To evaluate the results of the ERCOT congestion rights market, in Figure 46 we compare the annual auction price for congestion rights, the average