Modeling Regional Electricity Generation

1

Phillip Tseng

The U.S. Energy Information Administration
Email: Phillip.Tseng@EIA.DOE.Gov

April 2007

*This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matters from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's spring 2007 meeting with the Committee to be held April 19 and 20, 2007.

Modeling Regional Electricity Generation

Abstract

In recent years, natural gas use in the electric power sector has been on the rise. As a result, impacts of power generation from natural gas plants on the very tight U.S. natural gas market are becoming more pronounced. The ability to predict gas plant dispatching decisions can help analysts understand natural gas market conditions and the direction of price changes. Theoretically, dispatching decisions should be based on variable costs. However, considerations such as costs and accessibility of transmission lines, transmission and distribution losses, long-term contracts, distance to load centers, availability of fuels, ability to quickly ramp up or down generation, and cost and efficiency of available technologies, all play very important roles in power dispatching.

This paper has two objectives: 1. Present a model built on historical generation, capacity, and sales data to predict monthly dispatching of coal and natural gas power plants. 2. Demonstrate that the same modeling framework can be used to simulate the effects of capacity loss in a sub-region on regional generation patterns and trade flow. Preliminary model results show that the model can produce good projections. However, the model can be further improved to capture fuel switching for dual fired oil- gas generators.

1. Introduction

The United States Energy Information Administration (EIA) publishes a Monthly Short Term Energy Outlook. The publication, as its name suggests, covers EIA's view of short term energy markets and likely development in demand and supply of oil, gas, and electric power. One of the challenges in this publication is making consistent projections of natural gas use in the electric sector. It is important because an increase in power sector demand for natural gas can be a burden to the already tight U.S. gas market. EIA needs to improve its capability in assessing the impacts of higher demand for power on gas usage, overall demand for natural gas, the amount of gas injected into storage, and gas prices.

In meeting this challenge, EIA has completed a regional electricity dispatching model for the projection of power generation from coal and natural gas plants. The model divides the U.S. into three major regions, East, West, and Texas. One major region, West, has three sub-regions, the other, East, has nine sub-regions. The model has explicit demand and supply representations developed from historical data; it is structured to serve as a short term forecasting tool as well as an analytical framework for performing "what-if" simulations.

Section 2 discusses the modeling methodology. Section 3 describes the data sources and data transformation performed for meeting modeling requirement. Section 4 reports the performance of the model. Section 5 demonstrates the capability of the model in answering what-if questions. Section 6 summarizes the findings.

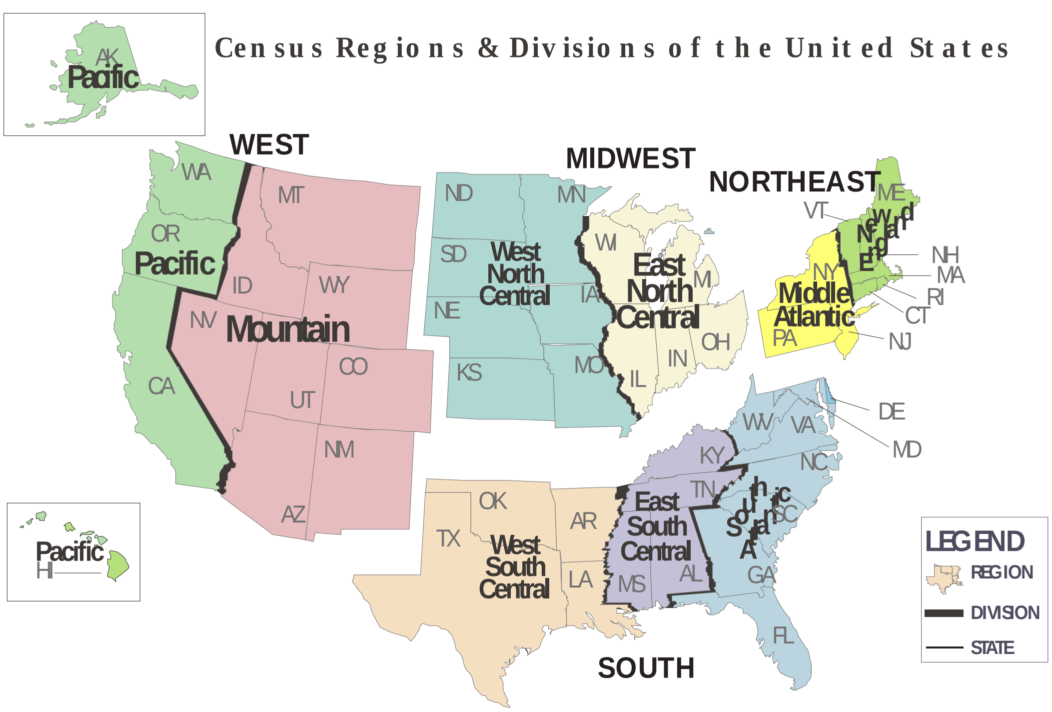
2. The EIA Regional Electricity Dispatching Model

2.1 Choice of Regions

The EIA regional electricity dispatching model divides the lower-48 states into three regions: Eastern, Western, and Texas. This division uses nine census regions plus four states: California, Florida, New York, and Texas. (See Figure 1) The selection of regions and sub-regions takes into consideration data availability, effects of trade within each region, and ease of use and maintenance for regular monthly dissemination. The general model structure can accommodate more regions for contingency analysis provided they are aggregated up from the state level.

The Eastern region includes nine sub-regions:

- New England
- Mid Atlantic minus New York
- East North Central
- West North Central
- South Atlantic minus Florida
- East South Central
- West South Central minus Texas
- Florida
- New York



The Western region includes three sub-regions:

- Mountain
- Pacific minus California
- California

Texas, separated out from West South Central, is a region by itself in the regional electricity dispatching model.

2.2 Model Structure and solution algorithm

This section describes the construction of supply curves, load curves, and the solution algorithm for the electricity generation model.

2.2.1 Regional supply curves

Conventional definition of a supply curve depicts the relationship between price and quantity supplied. Intuitively, dispatching decisions should be based on variable costs of power plants. A plant with the lowest variable cost should be dispatched first; as demand increases, the next lowest cost plant will supply power. In reality, dispatching decisions may depend on the least system cost instead of least variable cost because of spatial considerations and institutional factors; they may include costs and accessibility of transmission lines, supply contracts, distance to load centers, transmission and distribution losses, and availability of fuels. These considerations may play as important roles in actual power dispatching as the fuel cost and operating efficiency of available technologies.

Data on supply contracts, distance of a power plant to the load center, transmission distribution losses, and transmission tariff are not easily available. As a result, the construction of short run supply curves for coal, natural gas, residual fuel, and diesel fuel, relies on observed dispatching patterns instead of costs. The building block of supply curves is the power plant capacity utilization rate. For example, a plant with 1000 mega watts of coal capacity can generate 744,000 megawatt hours of electricity in July (1000*24*31). Reported net generation by the plant divided by 744,000 equals utilization rate. In each region, fossil plants are sorted by utilization rates and assigned to 10 bin numbers. Plants with utilization rates greater than or equal to 90% are assigned to bin number 1. Plants with utilization rate greater than or equal to 80% and less than 90% are assigned to bin number 2. All fossil plants are assigned one bin number. Table 1 shows the bin number, capacity in each bin, and cumulative capacity of coal and natural gas plants in Texas in July 2005.

The bin number and cumulative capacity are used to create supply curves, which embody capacity and historical dispatching order. We choose natural log of cumulative capacity as dependence variable to improve the fit and avoid negative fitted values. The supply curves for coal, gas, diesel fuel, and residual fuel take two functional forms.

Table 1: Coal and Gas Capacity (mega watt)

Bin Number	Coal	Gas	Cumulative coal	Cumulative gas
1	4632.320	2296.360	4632.320	2296.360
2	3707.360	5522.860	8339.680	7819.220
3	6269.800	5201.320	14609.48	13020.54
4	2598.160	10807.49	17207.64	23828.03
5	0.000000	6172.900	17207.64	30000.93
6	0.000000	4467.100	17207.64	34468.03
7	1767.200	5396.580	18974.84	39864.61
8	0.000000	9686.520	18974.84	49551.13
9	0.000000	5476.600	18974.84	55027.73
10	0.000000	14371.66	18974.84	69399.39

The coal supply curve takes the functional form:

Log(cumulative coal) = C(0) + C(1) * (1/bin)

Where C(0) and C(1) are estimated estimate coefficients

The gas, diesel, and residual fuel supply curves take the functional form:

Log(cumulative XXX) = C(0) + C(1) * bin + C(2) * bin^2

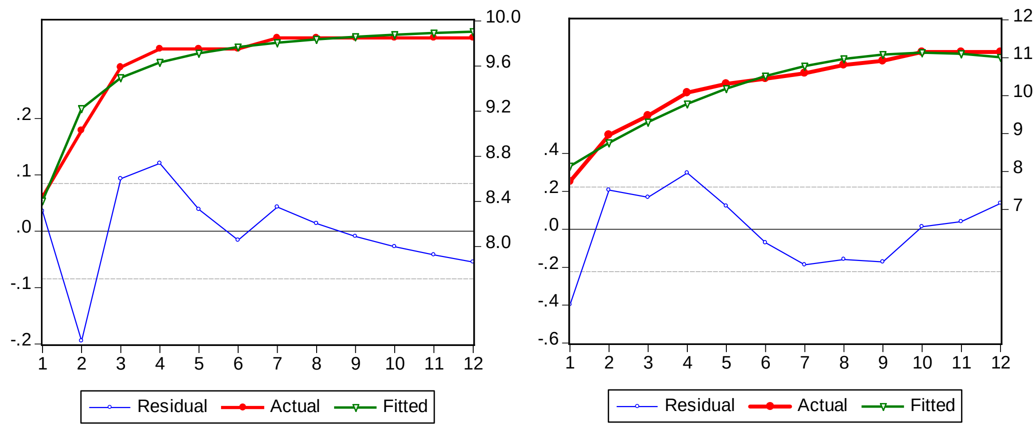
Where XXX=gas, diesel fuel, or residual fuel
C(0), C(1), and C(2) are estimated coefficients

Availability of coal power plants takes into consideration routine maintenance and planned outages. The data EIA collected show that most power plant operators reported use of small amounts of natural gas, diesel, or residual fuel; these observed data indicate that these coal power plants are not 100% available. An availability factor for each region is used to reflect this fact to improve model performance. The availability factor is 0.93, 0.94, and 0.96 for Easter Region, Western Region, and Texas, respectively.

Figure 2 shows the coal and gas supply curves. The right vertical axis is the natural log of cumulative capacity and the horizontal axis is the bin number corresponding to the cumulative capacity. For the coal equation, the adjusted R-squared is about 0.96. For the natural gas equation, the adjusted R-squared is about 0.95. Note that the horizontal axis shows 12 bin numbers. Number 11 and 12 are added to extend the flat portion of the supply curves. It is intended to enhance the fitted curves to conform to actual available capacity.

For the model, there are thirteen sets of supply curves for fossil plants for each month; one set for each sub-region. These monthly supply curves depict what plant operators will dispatch in response to variable demand levels in a typical day.

Figure 2: Texas July 2005 Coal and Gas Supply Curve



These historical monthly supply curves derived from monthly data for a single year capture only seasonal dispatching patterns. Dispatching gas power plants can be influenced by the prices of residual and diesel fuels. There are two aspects in fuel switching. First, if the relative price of gas is significantly below prices of residual fuel and diesel, utilization rate of natural gas plants could rise if the savings in "system cost" is positive. Second, for dual fueled gas/resid or gas/diesel plants, plant operators can switch if fuel cost savings from switching is positive. To capture the fuel choice over time, we pool the bin numbers and capacity data from 2002 through 2005 and introduce a relative price variable to the supply curve of gas and oil.

2.2.2 Regional demand for electric power

Demand for electricity changes by hour, by day, and by month. This fluctuation in demand determines dispatching needs and shapes dispatching patterns. Figure 3 shows the average 24-hour load shape of power demand in California in July 2005. Demand is lowest around 4 AM, ramps up to peak around 4 PM, and then ramps down in the evening. Changes in hourly demand determine hourly dispatching decisions. A sub-region without cost effective generating capacity during the peak hour would normally import from other sub-regions if transmission links are available. It is clear from the load curve that an importing region such as California does not have to import power every hour of each day. In fact, in the early morning hours in July, it put excess supply into pump storage. Figure 4 shows the seasonal pattern of demand from January 2004 through December 2005. In California, demand in April and May was the lowest, and July and August was the highest. This seasonal demand patterns can affect both maintenance decisions as well as seasonal dispatching patterns due to variations in plant availability.

Demand for power in a typical day depends on economic activities and daily temperature. Its components include demand from residential, commercial, industrial, and transportation sectors. As a result, weekday activities could be different from weekend activities and the level and shape of a 24-hour load curve can also be different. Ideally, two sets of 24-hour load curves should be used to capture the effects of peak load demand on dispatching decisions; one for weekdays and the other for weekends. However, the level of efforts in getting the more detailed data may outweigh the benefits of marginal improvements in out-of-sample projections. It is decided that for each sub-region one 24-hour load curve will be used to represent an average day for each month.

In this model, we use twenty four time series data to represent twenty four demand slices for a typical day of a typical month. For example, hour one of the time series data represent hour one demand for all the months in the data base, which can cover time periods from January 2005 through the end of RSTEO forecasting period. The sum of these 24 demand slices times the number of days in the month will always equal to a pre-determined monthly demand, which can be either historical data or projected demand.

2.2.3 Solving for dispatching of fossil plants

This sub-section discusses the solution algorithm, adjustments to the load curve, model outputs, and the computing platform.

The solution algorithm for the electricity dispatching model is straightforward; for each 24-hour time period, regional supply is equated to regional demand. The equilibrium bin number, now a continuous variable, then determines generation by fuel type. Figure 5 illustrates the process. First, the model aggregates oil, gas, and coal supply curves to get a total fossil fuel supply curve for a region. Given a regional demand, the model equates supply to demand and solves for equilibrium bin numbers and net generation. The equilibrium bin number then determines generation from oil, gas, and coal plants.

Figure 3: An Example of California Load Curve in July 2005
(mwh)

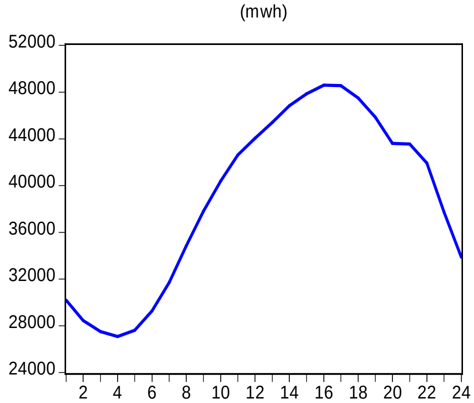
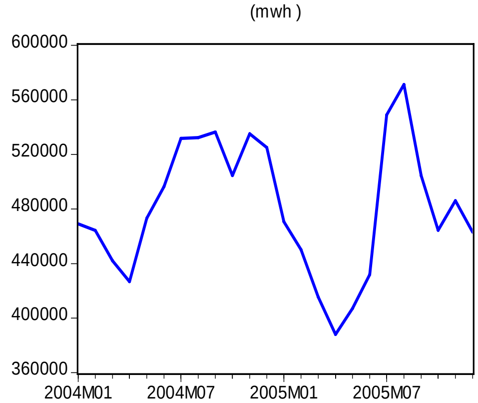


Figure 4: California Average Daily Demand for Electric Power
(mwh)



Mathematically, the following three equations show monthly supply, demand, and market clearing generations.

$$\text{Supply} = \sum S_{IJ}$$

Where I = coal, gas, diesel fuel, and residual fuel
J=sub-regions

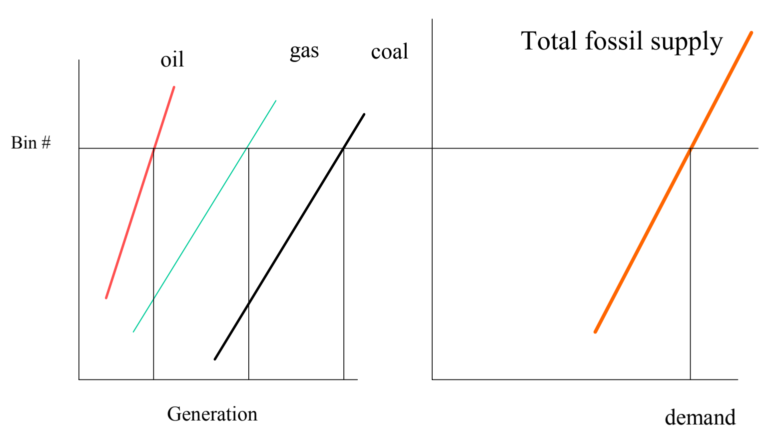
$$\text{Demand} = \sum D_J$$

$$\text{Supply} = \text{Demand}$$

For Texas, there is no sub-region. Monthly supply is simply the sum of four monthly supply curves: coal, gas, diesel fuel, and residual fuel. The January supply curves will be used, in conjunction with a 24-hour daily average January demand, to solve for 24 generation levels of coal, gas, diesel, and residual fuel. Daily coal generation is the sum of 24 hourly generations, and monthly generation equals daily generation times the number of days in the month.

For the Western region, there are three sub-regions: Mountain, Pacific minus California, and California. The monthly regional supply curve is the sum of three sub-regions supply of coal, gas, diesel, and residual fuel. The regional 24 hourly demands is the sum of demand in these three sub-regions. For the Eastern region, there are nine sub-regions and the calculation is the same as the Western region except the number of sub-regions is nine instead of three.

Figure 5: An Illustration of Demand determines Dispatching



Adjustment to the load curves to remove supply of power from hydro, nuclear, renewable, and combined heat and power (CHP) is necessary to simplify the dispatching algorithm. An ideal approach is to remove the generation of these technologies from the load curve according to their dispatching pattern if we know when these technologies are dispatched and how much. For example, we know nuclear power plants probably run continuously 24 hours a day so we subtract hourly nuclear generation from the 24-hourly load curve. For hydro power, we should allocate more generation to peak hours. For solar power, we should convert monthly to daily average and allocate generation between late morning to late afternoon. The level of effort required to make the adjustments to fine tune the shape of adjusted load curve outweighs improvements in projecting dispatching fossil plants because solar generation is very small and wind generation cannot be determined by power plant operators. To simplify the adjustment, we convert monthly non-fossil generation to a 24-hour daily average generation curve and subtracting the generation from the 24-hour load curve. The dispatching model uses the adjusted load curves to solve for generation of coal, gas, diesel, and residual fuel.

The dispatching model outputs include regional and sub-regional power generation by coal, natural gas, diesel fuel, and residual fuel for the forecasting time horizon of 24 to 36 months. Fuel use can be computed based on net generation and average heat rate of power plants. In addition, criteria pollutants such as NOx, SOx, and CO2 can be derived from fuel consumption.

The model uses EViews software to estimate the coefficients of the supply curves and solve the model for generation by fuel type. It takes about 60 seconds to solve over a period of 48 months.

3. Data Sources

The model is built upon three major data sets collected and published by the U.S. Energy Information Administration. These include annual generation capacity, monthly net electricity generation, and monthly electricity sales. Following is a brief description of the data sources used to construct the model. It is important to note that EIA relied on a private consulting firm to convert the monthly sales data to 24-hout average daily load curves.

- EIA-826: Monthly electric utility sales and revenues reported by states. Sales data is based on billing cycle, which may not coincide with the calendar month.
- EIA-906: Monthly data on generation and fuel use at the generator or plant level by states.
- EIA-860: Annual generating capacity by technology and fuel at plant level

4. Performance of the Model

The current regional electricity model used 2005 annual capacity and monthly generations to create monthly electric power supply curves of coal, natural gas, diesel, and residual fuel for all 13 regions. Monthly sales of electricity for 2004, 2005, and 2006 at the state level are aggregated to the sub-regional level. For each sub-region, monthly sales data were then converted to 24-hour load curve. The model uses hourly load to solve for generation by coal, natural gas, diesel fuel, and residual fuel.

Figures 6, 7, and 8 show the model results; they compare projected average daily generation from coal and natural gas to their historical counterparts. Overall, the model is capable of capturing the seasonality of electricity markets. In general, the model has the tendency to use more coal and fewer gas and oil fired units.

While projected generations from coal and gas track historical patterns well, projected generations from diesel and residual fuel are not very satisfactory. Several factors may contribute to this deficiency. First, dispatching decisions oil based generators may be different from that of coal and natural gas; it is very likely that these so-called peak units may also be used as transition units to smooth out the changes in load because running a large generator as spinning reserve may be more costly if load shape and dispatching options can be well defined. Second, the load curves in each sub-region were aggregated from different sub-regions or states. As a result, these curves may have flattened load shapes and prevent the model from dispatching correctly. Third, generators may have reported inconsistent data to EIA because the relatively small shares of oil-based units in the whole generation mix. Fourth, EIA may have adopted a methodology in the estimation of generation for plants which are not in the monthly sample. An experienced model user could make adjustments to the bin numbers assigned to diesel and residual fuel units and try to match the historical generation pattern. However, the fact that oil-based units are a very small fraction of total generation makes it less attractive to spend too much effort to make the improvement.

There are many different measures to evaluate the performance of a model. The selection of measurements, therefore, depends on the intended use of the model. EIA will use the model to provide monthly forecasts and gain insights in the use of natural gas by the power generation sector. EIA will also use the model to simulate the effects of unplanned plant outage on generation patterns and trade flows; scenario analysis could include the effects of above normal cooling degree days on power generation and use of natural gas and petroleum based products.

The goodness of fit of the model can be measured by the mean absolute percentage error (MAPE). Equation 1 shows the computation of MAPE.

$$100 * \sum_{t=T+1}^{T+h} \left| (\hat{y}_t - y_t) / y_t \right| / h \qquad (1)$$

Where $\hat{y}_t$ is projected value

y_t is historical value

h is number of observations

Figure 6: Eastern Region Coal and Gas Generation
(mwh/day)

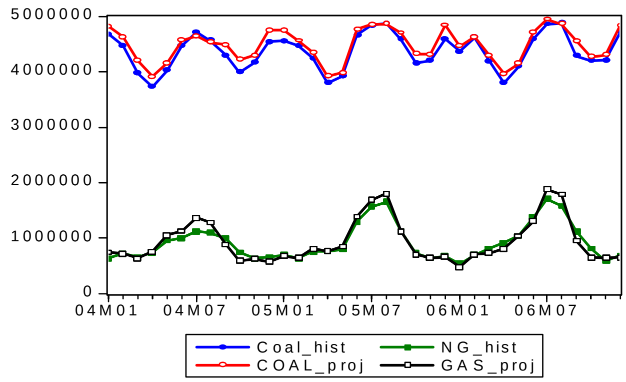


Figure 7: Western Region Coal and Gas Generation
(mwh/day)

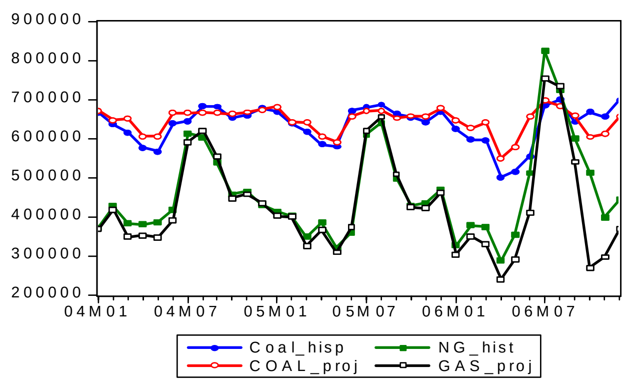


Figure 8: Texas Coal and Gas Generation
(mwh/day)

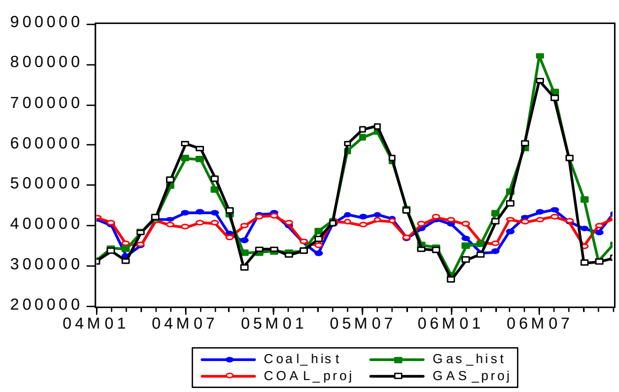


Table 2 shows the mean absolute percentage error for the Eastern Region, Western Region, and Texas. The MAPE calculated from the 12 months of 2005 for gas are 2.10%, 2.88%, and 4.14% for Texas, Western Region, and Eastern Region respectively. MAPE increased when we include the twelve months of 2004 out of the sample forecast. The 24 month MAPE calculated from 2004 and 2005 projections are 3.04%, 3.51% and 7.44%. Western Region has the smallest MAPE for coal. The performance of the Eastern region is not as good as the other two regions because of the concentration of diesel and residual fuels in the East and the inability of the model to select the dispatching patterns satisfactorily.

Another measure of the performance of the model is the cumulative percentage difference within the forecasting time horizon. It is relevant especially for the natural gas market. If natural gas use is consistently above normal, injections into underground storage would be below normal. As a result, natural gas prices in the spot and future market may increase in anticipation of potential shortages. In addition to the MAPE, the other measure can be twelve month cumulative percentage error. Equation 2 reports the calculation.

$$\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t) / y_t \tag{2}$$

Table 3 shows the 2005 12-month cumulative percentage difference (CPD). Texas has the smallest CDP. Both Texas and West have absolute CPD smaller than 1%. However, East shows 2.35% for coal and 4.06% for gas. For the East, CPD for coal and gas are positive and it reflects the model under project generation from diesel and residual fuel.

Table 2: Mean Absolute Percentage Error

	East		West		Texas	
	Coal	Gas	Coal	Gas	Coal	Gas
2005	2.58%	4.15%	1.92%	2.88%	2.39%	2.10%
2004/5	3.06%	7.44%	2.44%	3.51%	3.32%	3.04%

5. Scenario Analysis

The electricity industry has been going through restructuring since the early 1990s. The aging transmission infrastructure and the shifts and changes of new load centers have prompt the need to understand the flow of electricity power. This model provides a consistent framework to analyze market demand, supply, and trade flows.

The model uses a bottom up approach to solve for power generations consistent with demand, generation capacity, and historical generation pattern at an aggregated level such as the Eastern and Western regions. The process is straightforward: first, aggregate the demand and supply from sub-regions to regional level and solve for generation and trade flows for each sub-regions that clear the aggregate market demand and supply. Next, use the model solution for the sub-regions to get generation and implied trade flow. Model results for a sub-region such as implied trade flow can be used to impose constraints on imports and simulate impacts of different demand and supply scenarios on generation patterns and use of fossil fuel in the sub-region.

Figures 9 and 10 demonstrate possible generation and trade patterns in the West in July 2005. The model solution is derived from the Western Region with three sub-regions. However, to simplify the analysis, we assume there are only two sub-regions in the West instead of three; IMP08_7 is imports of mountain region and IMP10_7 is imports of California. In this case, IMP08_7 is negative and shows hourly exports of power to California. The model results show that the level of California power imports varies by the hour and peaks at and around hour 17. Given an established base case, we can make several sensitivity model runs to examine the potential effects of various scenarios on generation pattern and fuel use. Specific scenarios could include the following:

- Raise peak load in California
- Lower imports
- Remove generation capacity

Table 3: 12 month cumulative percentage difference

	East		West		Texas	
	Coal	Gas	Coal	Gas	Coal	Gas
2005	2.35%	4.06%	0.62%	-0.83%	-0.23%	0.24%

These model runs will produce new generation patterns and fuel use. It could also identify if the implied constraint on trade flows causes any difficulties when the demand is raised to a level much higher than normal.

Another useful application of the modeling approach is in the analysis of power flows and transmission reliability. One of the most challenging tasks facing analysts of transmission reliability studies is the lack of consistent power flow data. Congestion on transmission lines is reported. However, it is very difficult to assess the severity of the problem. EIA published a report on transmission data needs a few years ago and it received stakeholders' recognition. Data collection, however, needs to be more focused and it can be done properly only if users of the data understand how the collected data will be used. The regional electricity model provides a consistent framework in linking sub-region demand, supply, and trade in a trade zone. Model solutions can be used to identify or impute trade flows at below sub-region level and they could help in the analysis of transmission issues.

The model imposes several conditions to the load curve and generation data that should provide consistent estimates of the implied power flows. Equations 3 and 4 specify the conditions that the model inputs and solution must meet. Equation 3 specifies that the sum of 24 hourly loads in a sub-region must equal an exogenously pre-determined monthly demand. Equation 4 says regional generation in each hour must equal regional hourly demand. It is the way the model solves for a market clearing demand and generation.

Figure 9: An Example of Hourly Load, Generation, and Imports for Mountain Region, July 2005
(mwh)

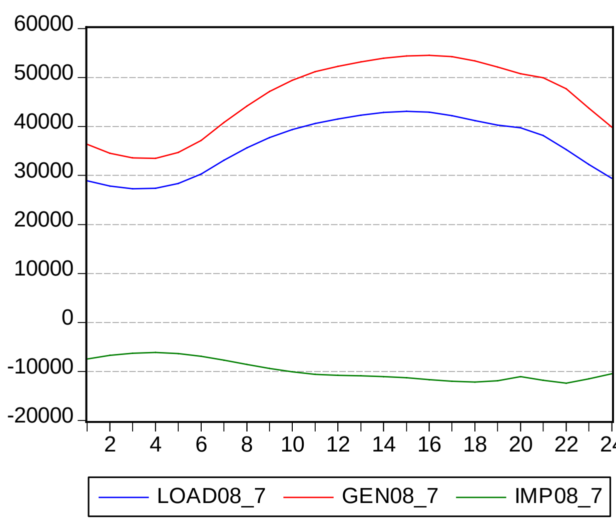
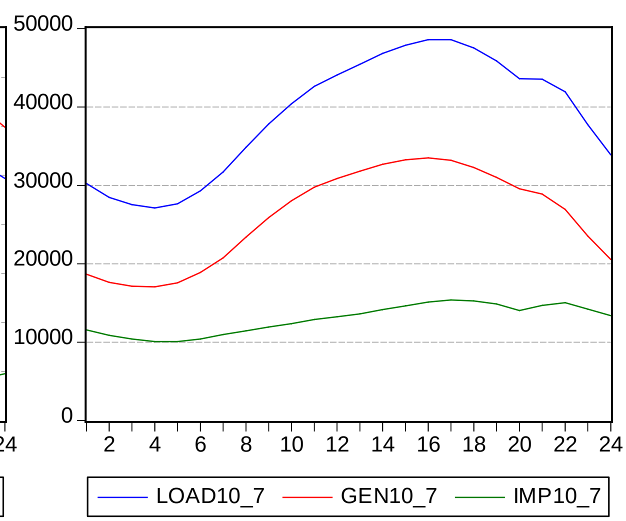


Figure 10: An Example of Hourly Load, Generation, and Imports for California, July 2005
(mwh)



$$\sum_{h=1}^{h=24} D_{hr} * \text{days in a month} \equiv \text{Monthly demand in a region} \qquad (3)$$

Where D=hourly demand
h=hour 1 through 24
r=regions

$$\sum_{r=1}^{r=n} G_{hr} = \sum_{r=1}^{r=n} D_{hr} \qquad (4)$$

Where G=hourly generation

Equations 5 and 6 show that the model solution can provide useful insights on maximum power flow if power generation in a sub-region matches well with the historical data. Equation 5 states that if generation in a sub-region matches closely to historical data, it should provide useful insights on the hourly dispatching pattern of power plants in the regions. As a result, the maximum difference between hourly load and generation could provide information on possible limits on transmission limits.

$$\sum_{h=1}^{h=24} G_{hr} * \text{days in a month} \approx \text{Monthly generation in a region} \qquad (5)$$

$$Max(abs|D_h - G_h|) \quad \text{for h=1 through 24} \qquad (6)$$

In the current model, sub-region 8 is the census Mountain region, which consists of eight states. In the previous example on the trade between California and Mountain region, one may ask about the trade flows within the Mountain regions and to California. Given the current model framework, we can use power exports from a model run from the Western Region to adjust the Mountain region hourly demand. We can use the state load data and state supply curves to solve for hourly generation. A calibrated model run for the Mountain region will report generation and power flows within the mountain region. These imputed power flow data can then be compared with transmission data collected by the EIA. Note that reported data on generation and sales in each state provide a glimpse of the possible import/export of power. However, the non-linear characteristics of load and supply curves make it difficult to pinpoint the level of imports during peak demand hours when potential bottlenecks in supply and the delivering system can be encountered. As a result, the simulation approach can provide more valuable insights on potential bottlenecks.

In the areas of studying constraints of criteria pollutants on power generation, the data base of this model can be linked directly to the capacity of individual power plants. The model can calculate cumulative emissions associated with each supply curve. Coal supply curves can be adjusted to reflect the constraints of emissions of criteria pollutants on generating capacity.

6. Summary

The modeling approach adopted in this paper provides consistent projections of generations from coal and natural gas. Using 2005 generation and capacity data, the model could project coal and natural use with good results. The mean absolute percentage error ranges from 1.92% to 4.15% for the 12 months of 2005. The MAPE ranges from 2.44% to 7.44% when we include the out of sample forecast of 2004 data. Twelve month cumulative percentage differences for 2005 are also good; they range from less than 1% to about 4%.

The model can be improved if the model can get insight into the use of residual and diesel fuel. A modeler can change bin assignments for generators burning residual fuel and diesel and raised the level of generation from these two fuels. As a result, the performance in the East can improve further. The model can be further improved if the effects of fuel costs on dual-fueled generations can be captured more appropriately.

In addition to forecasting, the modeling framework can be used to conduct scenario analysis and provide insights on the flows of power between states.

Questions for the ASA Energy Committee

A few sub-regions in the East have generators that can switch from gas to oil or vice versa. The adopted modeling methodology uses observed dispatching patterns as proxies to supply curves. There are needs to estimate a switching variable to capture the effects of gas and oil prices on gas and oil use if we want to improve the performance of the model.

Should we incorporate the fuel switching in the model structure? Alternatively, Is it appropriate to make adjustments to the model solution based on empirically estimated changes in oil-gas share?