5.1 DEVELOPING IMPROVED TOOLS FOR ELECTRIC UTILITY PEAK LOAD FORECASTING

David J. Sailor* and Paul B. Breslow
Tulane University, New Orleans Louisiana

1. INTRODUCTION

As power generation costs increase and market competition intensifies, precise demand forecasting information is becoming more valuable (Gellings, 1996; Jameson, 1999). While weather variability is a major contributor to variability in utility peak loads, commonly available meteorological data are not necessarily well-suited to use in load forecasting. Weather forecast data used as input for load forecasting are often regional in nature and not resolved at the scale of the city (or the utility's specific customer service area). Furthermore, variables available for model input (such as peak temperatures) while fairly highly correlated with load, may not be as useful as derived parameters such as heating and cooling degree-days. When degree-day variables are used they are often based on coarse spatial (and even coarse temporal) resolution temperature data. Furthermore, the assumed base temperature and appropriate time period for integration for degree day calculations should be region-specific and can be optimized for the particular application domain (e.g., Sailor and Muñoz, 1997). Likewise, for certain regions the inclusion of Latent Enthalpy Days (analogous to cooling degree days, but for humidity) can be an important determinant of electricity loads.

With these issues as motivating factors we are developing a new load forecasting approach to be used as a tool in estimating peak loads. The load model will be trained using an enhanced set of weather parameters along with a Tree-Structured-Regression (TSR) modeling tool in place of the commonly used Neural Network approach (e.g. Chow and Leung, 1996; Rahman and Hazim, 1996; Caciotta et al., 1997; Dash et al., 1997). Finally, to provide improved weather input for load forecasting we are developing an urbanized version of the MM5 mesoscale model from the National Center for Atmospheric Research. This model will provide more detailed representation of weather in and around urban areas, and will become an integrated component of the load forecasting system. This system, while still under development, will be described in this presentation.

* Corresponding author address: David J. Sailor, Tulane Univ., Dept. Mech. Engr., New Orleans, LA 70118; e-mail: sailor@tulane.edu

2. DATA

At the present time our focus is on city-scale short-term (e.g. 24-hour) forecasting of peak electric utility loads. The appropriate data for this analysis are summarized below.

2.1 Utility Load Data

We chose the metropolitan New Orleans (NO) area for our initial model development primarily because we have an existing relationship with the sole-source provider of electricity for metropolitan New Orleans (Entergy Corporation) and therefore had relatively easy access to their load data. The utility data that we obtained for NO included hourly load data (MWH) from Jan.1, 1986 through Apr. 30, 2001. Our initial analyses include only the first 10 years of the data set. Also, as our interest is in peak load forecasting we manipulated the hourly load data set to create a peak load data set that includes date, peak load, and time of peak load. We augmented these data set by defining a day-type variable (workday="W" or not-workday="N").

2.2 Meteorological Data

While the methodology being developed will eventually include enhanced meteorological model output for forecasting, initial model development (training) is being pursued using hourly meteorological data from the single nearby airport station (MSY). These data include hourly values of temperature, wind speed/direction, humidity, and cloud cover. We have calculated daily averages, maxima, and minima. We have also created the derived variables of cooling and heating degree days (using hourly data with a base temperature of 18.2 C). We also intend to investigate using Latent Enthalpy Days in place of raw humidity data:

$$LED = \frac{1}{24}\sum_{1}^{24}\alpha(E - E_o)\,, \qquad (1)$$

where E is the enthalpy and E_o the enthalpy at the measured temperature and a humidity ratio of 0.0116. The constant α takes on a value of 1 if the temperature is above 25.6 ^{o}C and a value of 0 if the temperature is below this value or if the enthalpy difference is negative. Conceptualy, LED represents just the amount of energy required to lower the humidity to ASHRAE comfort levels without reducing air temperature.

2.3 Combined Analysis Data Set

The daily data sets for load and weather were merged to create an analysis data set with complete information regarding peak loads and the corresponding weather parameters. In the present study we attempt to remove all non-weather factors by defining a one-day forecast Peak Load Ratio as follows:

$$PLR(n) = \frac{PL(n)}{PL(n-1)} \quad , \qquad (2)$$

where PL(n) is the peak load for day n. As can be seen in Fig. 1, the day-to-day variability in peak loads is commonly on the order of 10 to 20%.

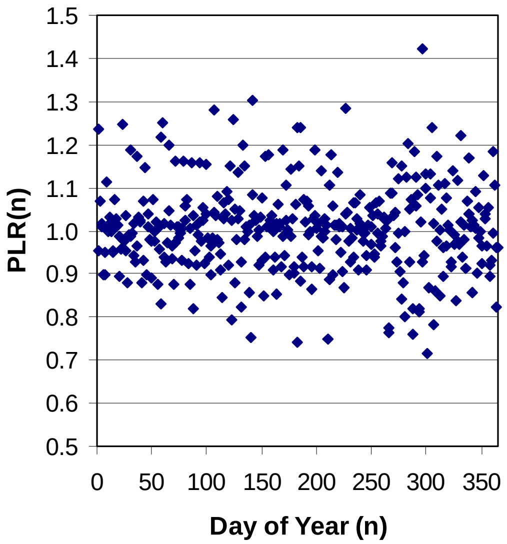

Fig. 1. Day-to-day variability of peak load for one year of data (1995).

To develop a suitable model for PLR it is important to include relevant weather parameters for day n and n-1. Alternatively one can define weather parameter ratios or deviations for the two days. The day-type variable can also be modified to include all 4 permutations of the day-type variable for day n and n-1. This new day-type parameter takes on values of 1, 2, 3, or 4 for Workday and Non-workday combinations of NW, NN, WN, and WW, respectively. Fig. 2 shows the expected relationship between the day-type parameter and the corresponding peak load ratio for the two days. Specifically, for workdays that follow non-workdays - typically Mondays – (DAYTYPE=3) the PLR is usually greater than unity (average of 1.17). Likewise for non-workdays that follow workdays (DAYTYPE=1) the PLR is typically less than unity (average of 0.89). Within each day-type category in Fig. 2, the variability in PLR is illustrative of the role of weather variability.

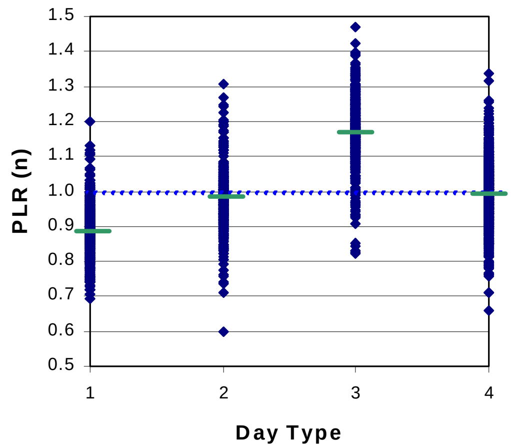

Fig. 2. Peak load ratios for the 4 day-type combinations.

3. MODELING APPROACH

The general approach is to use weather and peak load data for day n-1, along with a forecast of weather parameters for day n to estimate peak load for day n. This is accomplished indirectly by calculating the PLR(n), and can, in principle, be applied to forecasts ranging from 24 to 72 hours.

3.1 Mesoscale Model

While the initial focus of model development involves training a peak load model using observational data, the eventual goal is to incorporate mesoscale model forecast output into the modeling system. To make the meteorological model more accurate for urban applications we are currently developing an urbanized version of the MM5. While many potential modifications are under consideration, this model will, at a minimum, include the following components: a diurnally-varying albedo to represent the effect of solar zenith angle on the solar radiation absorbed by a city; a daily/seasonal anthropogenic heating profile to be incorporated as a perturbation to the surface air energy budget; and enhanced representation of urban surface characteristics by extending the single "urban" land use category from the USGS system to three or more sub-categories. Tradeoffs between accuracy and computational speed, while not an issue in the initial model development will be important for successful operational load forecasting.

3.2 Tree Structured Regression

There are a number of weather parameters that influence electric utility loads. Also, some of the parameters that are key to determining loads are categorical in nature (e.g., day of the week). The Tree Structured Regression (TSR) approach is uniquely suited to this type of problem. The basic premise is that some partitioning of the data pool is required before strong models can be developed for forecasting load. For example, each weather factor affects load differently depending upon whether the day is a weekday or weekend. Similarly the role of each weather factor in affecting load also depends on the general magnitude of other weather parameters.

Tree Structured Regression (TSR) is a subset of the statistical tools collectively known as Classification and Regression Trees (CART) and was pioneered by Breiman et al (1984). TSR is a guided (supervised) classification scheme that uses a predictand (such as peak load ratio) as guidance when generating clusters in the data pool. The "distance" between clusters is measured by the value of the predictand variable.

In any regression analysis, the condition that the learning and testing data sets should come from the same population must be met. Due to the inter-annual variability of utility loads, an arbitrary sampling procedure may result in an imbalance between learning and testing data sets. In our analyses, the learning and testing data sets are created using a random sampling procedure with 67% of data points going to the learning data set. By putting all the data points in the same pool, the learning set can be well mixed (can include datum points from any month and any year).

Tree-Structured Regression generates a regression tree using a succession of binary rules (questions). In generating the tree all training data initially reside in a root node. This node (and each subsequent node) is split into two descendent nodes based on a binary question. The question is posed based on the value of an independent variable. In the case where peak load ratio is being predicted it is likely that ΔCDD will be a split variable and a representative question might be of the form: $\Delta CDD > \beta$?, where β is some constant. Depending upon the answer to this binary question each training datum point travels to either the left or right descendent node. The choice of split variable and the actual value, β, for the binary question are determined in an optimization routine with the goal being to separate the initial node into two descendent nodes that are characteristic of distinctly different magnitudes of the dependent variable – peak load ratio in this case. Specifically, the optimal split is the one that minimizes the resubstitution estimate (RE):

$$RE = \sum_i \left[L_i - \overline{L}\right]^n + \sum_i \left[R_i - \overline{R}\right]^n \qquad (3)$$

where L_i and R_i are the values of the independent variable for each point in the "L"eft and "R"ight descendent nodes, respectively, and the over-bar represents a mean over the node. The value of n is generally taken as 2 (for a least squares approach).

The extent of tree growth can be controlled as desired. The terminal nodes of the tree then represent categories of datum points. Typically, in-node multiple linear regressions are performed in the terminal nodes and the binary rules combined with these regression equations represent the TSR model for the dependent variable. A sample tree with 2700 points of training data is shown in Fig. 3. For further details on the TSR approach see Breiman et al. (1984).

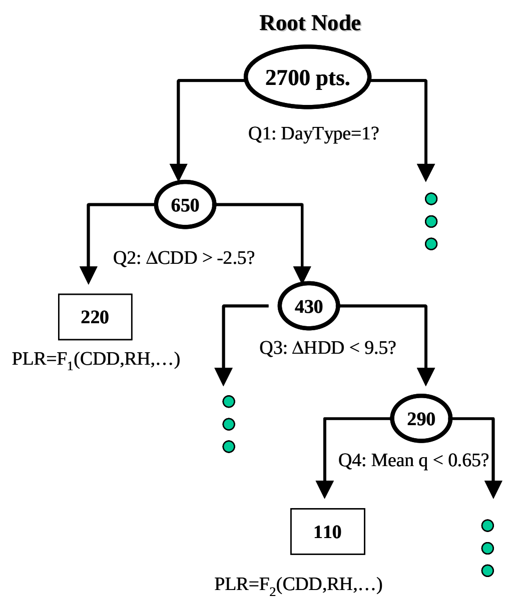

Fig. 3. A portion of a hypothetical regression tree relating the Peak Load Ratio (PLR) to various weather-based rules.

4. PRELIMINARY RESULTS

Initial analysis of our load data indicate strong correlations between PLR and the 1 and 2-day lagged degree day deviation variables ΔCDD_1 and

ΔCDD_2. The corresponding correlation coefficients are 0.46, and 0.37 respectively. As expected these correlations are significantly higher (double) than those of the corresponding raw temperature parameters.

Initial development of regression trees for PLR has resulted in strong models that have correlation coefficients in excess of 0.90, and are capable of explaining more than 80% of the variance in peak load ratios. The dominant (split) parameters in these trees are the day-type and heating/cooling degree-day variables. At the present time humidity is only incorporated in the models through the raw humidity variable and its daily differences. As a result humidity does not factor into the split variables, and is only incorporated in the multiple linear regression equations within the terminal nodes. It is expected that when Latent Enthalpy Day variables are included in the analysis they will also figure prominently in the list of split variables and will result in improved predictive capability. Fig. 4 shows the general prediction capabilities in terms of the predicted PLR vs. the actual PLR.

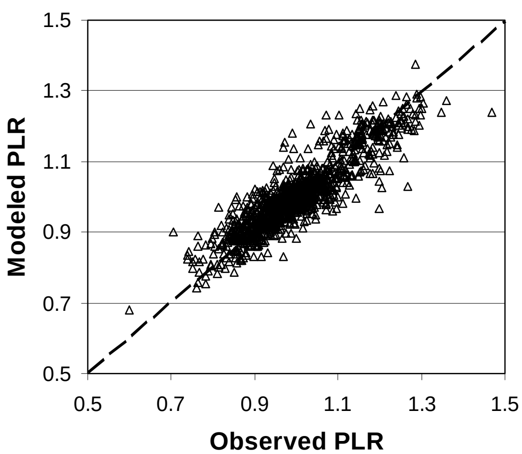

Fig. 4. Sample plot showing quality of TSR results for a preliminary model that includes only traditional CDD and HDD variables (R=0.90).

5. CONCLUSIONS

The methodology under development represents enhancements to traditional load forecasting approaches in several respects. First, the traditional variables used in load forecasting are expanded to include additional derived variables that are more closely tied with the weather-related causes of load variability. A second enhancement is the use of tree-structured regression for building the empirical relationships between load and weather variables. A final improvement, still under development is the urbanization of a mesoscale model for providing improved meteorological conditions for use in the load forecasting stage. Eventually these tools will be coupled in a single system for predicting peak loads. While 24-hour forecasts are the focus of the present work the approach is easily extended to longer forecast periods.

Acknowledgements

The authors wish to thank Dr. Xiangshang Li for his ongoing assistance and insightful comments. Neel Sus assisted with the data manipulation. We also wish to acknowledge Jim Kenney and Doug Heinson at Entergy Corporation for providing the necessary load data. One author, PBB, wishes to acknowledge financial support of a graduate fellowship from the Louisiana Board of Regents.

REFERENCES

Breiman, L., J.H. Friedman, R.A. Olsen, and J.C. Stone, 1984: Classification and Regression Trees, Wadsworth.

Caciotta, M., et al., 1997: Application of Artificial Neural Networks to Historical Data Analysis for Short-Term Electric Load Forecasting, *European Transactions on Electrical Power*, **7** (1), 49-56.

Chow, T.W.S., and C.T. Leung, 1996: Neural-Network-Based Short-Term Load Forecasting of Anomalous Load Periods, *IEEE Transactions on Power Systems*, **11** (4), 1736-1742.

Dash, P.K. et al., 1997: A Real-Time Short-Term Load Forecasting System Using Functional Link Network, *IEEE Transactions on Power Systems*, **12** (2), 675-680.

Gellings, C.W., 1996: Demand Forecasting in the Electric Utility Industry. Tulsa, Ok: Penn Well.

Jameson, R., 1999: Managing Energy Price Risk, 2nd Ed., London: Risk Books, 324pp.

Rahman, S., and O. Hazim, 1996: Load Forecasting for Multiple Sites: Development of an Expert System-Based Technique, *Electric Power Systems Research*, **39** (3), 161-169.

Sailor, D.J., and R. Muñoz, 1997: Sensitivity of Electricity and Natural Gas Consumption to Climate in the U.S. - Methodology and Results for Eight States, *Energy, the International Journal*, **22** (10), pp. 987-998.