ELSEVIER

International Journal of Forecasting 13 (1997) 161–174

international journal of forecasting

Short-run forecasts of electricity loads and peaks

Ramu Ramanathan[a,*], Robert Engle[a], Clive W.J. Granger[a], Farshid Vahid-Araghi[b], Casey Brace[c]

[a]University of California, Department of Economics, San Diego, La Jolla, CA 92093-0508, USA
[b]Texas A&M University, Department of Economics, College Station, TX 77483-4228, USA
[c]Puget Sound Power and Light Company, P.O. Box 97034, Bellevue, WA 98009-9734, USA

Abstract

This paper reports on the design and implementation of a short-run forecasting model of hourly system loads and an evaluation of the forecast performance. The model was applied to historical data for the Puget Sound Power and Light Company, who did a comparative evaluation of various approaches to forecasting hourly loads, for two years in a row. The results of that evaluation are also presented here. The approach is a multiple regression model, one for each hour of the day (with weekends modelled separately), with a dynamic error structure as well as adaptive adjustments to correct for forecast errors of previous hours. The results show that it has performed extremely well in tightly controlled experiments against a wide range of alternative models. Even when the participants were allowed to revise their models after the first year, many of their models were still unable to equal the performance of the authors' models. © 1997 Elsevier Science B.V.

Keywords: Comparative methods; Energy forecasting; Forecasting competitions; Regression methods; Exponential smoothing

1. Introduction

Electric utilities have always forecast the hourly system loads as well as peak loads to schedule generator maintenance and to choose an optimal mix of on-line capacity. As some facilities are less efficient than others, it is natural to bring them into service only during hours when the load is predicted to be high. Nowadays however, the need for accurate hourly load forecasts is even greater. There are several reasons for this that reflect the newer technologies available for generation and transmission of power. As utilities add smaller generating equipment, it becomes easier to adjust capacity flexibly in response to demand. Furthermore, utilities are now able to adjust capacity and demand through short term purchases and sales of power. Today and increasingly in the future, excess power can be sold and shortfalls can be made up by purchases. A careful calculation of the expected demand and supply can lead to contracts that enhance the profitability of the utility. All these lead to still more reliance on very short-run forecasts.

This paper reports on the design and implementation of a short-run forecasting model of hourly system loads and an evaluation of the forecast performance. The methodology was applied to historical data for the Puget Sound Power and Light Company, but the approach is more general and can be adapted to any utility. Puget Power did a comparative evaluation of various approaches to model-

*Corresponding author. Tel.: (619) 534-6787; fax (619) 534-7040.

PII S0169-2070(97)00015-0

ing hourly loads and those results are also presented here but without identifying the participants except by coded labels. Because of space limitations, many of the details (especially tables and charts) are omitted. For a copy of a more detailed paper, contact Professor Ramu Ramanathan at the Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, U.S.A.

2. Methodologies

There have been a number of approaches to modeling hourly loads [see Bunn and Farmer (1985); Electric Power Research Institute (1977) for a collection of several papers on the topic]. Ackerman (1985) compared three alternative models; (a) a twenty-fourth-order differencing of the raw data modeled as a first-order autoregressive (AR) process, (b) consecutive differencing modeled as AR(24) combined with a moving average (MA) of the first order, and (c) a simple AR model on a different data set that eliminated the input of the load dispatcher. They found the third specification to be more robust than the others across different data sets. Gupta (1985) proposed a model that not only took into account historical data and weather information but was adaptive in nature. Thus, model parameters were automatically corrected in response to changing load conditions. Schneider et al. (1985) formulated the total system load in terms of four components; basic load, weather sensitive load, error component, and a bias correction. The basic load is modeled by sine and cosine functions. Temperature effects were captured by a nonlinear function. The bias in forecasts from these terms is used to construct a bias correction that is added to the forecast to obtain a revised forecast. The authors found that the approach worked quite well for data from the San Diego Gas and Electricity Company. The approach adopted by Ramanathan et al. (1985) is a two-step method. First, a short-run model is formulated for a three month period that relates the hourly load to quickly changing variables such as the temperature, time of day, day of the week, and so on. The parameters of this model are then related in a second stage to slowly changing variables such as income, electricity price, demographic and industrial structure, and so on. The authors applied this technique to two very different data sets (a cross section of households observed over a full year and 30 regions of the United States observed for a 14 year period) and found the approach to approximate historical load shapes quite well. However, this is a load-shape model and does not forecast very easily.

3. Description of the present project

The main goal of the present project was to develop a number of models to produce very short run forecasts of hourly system loads. Several teams of researchers were invited to participate in a real-time forecasting competition (labeled the "shoot-out") using historical data from the Puget Sound Power and Light Company (Puget) which acted as the host utility. This paper presents a detailed specification of the models developed by Engle, Granger, Ramanathan, and Vahid–Arraghi (EGRV) whose study was sponsored by the Electric Power Research Institute (EPRI), and a comparative evaluation of the forecast performance of the various models. The EGRV study also modeled the distributions around the forecast values so that a measure of error could be ascribed to each forecast, but those results are presented elsewhere [see the EPRI project final report Electric Power Research Institute (1993)] and were not part of the shoot-out.

At Puget, the daily forecasts were made at 8:00 A.M. by Mr. Casey Brace. Forecasts were produced for the entire next day, starting at midnight and through the following midnight and hence they were made from 16 to 40 hours into the future. On Friday, he forecast the entire weekend as well as Monday (16 to 88 hours). Before holidays, he forecast until the next working day. At the time the forecasts are produced, the information available are the weather service forecasts of temperatures at each hour as well as current and past weather and load information.

In order to formulate a statistical model to carry out the same task, Puget developed a data base of hourly observations for the fall and winter months from 1983–1990. Included with the data set were the hourly loads and the hourly weather observations. Each potential forecasting method was coded into a software program that would run on a PC. At 8:00

A.M. the program would be run and the requested data on past weather, past loads, and weather service forecasts would be entered. The official experiment began on November 1, 1990, and continued for five months, through to March 31, 1991. Altogether ten participants entered one or more forecasting models for this comparison. These included two EGRV models, several neural network models, several state space models, two pattern recognition models, a naive model, and most importantly, a model by Puget Power's expert Mr. Lloyd Reed. Because the models are proprietary, we do not have permission to describe in detail the forecasting models used by the other researchers. However, Harvey and Koopman (1993) have published one model using time-varying splines. Also, Brace et al. (1993) presented a paper at the Second International Forum on the Applications of Neural Networks of Power Systems. The techniques used range from the fairly conventional to very modern approaches. The results are interesting and are discussed below. Because of the importance of the problem and the enthusiasm of the participants, the experiment was continued for a second year, from 1991–1992. In this case, the participants were somewhat different and all model builders were given information on the other models and on the forecast results of the previous year. The second time around, the models were therefore more sophisticated than the first time.

4. The EGRV model

4.1. The basic strategy

All the models discussed in Section 2 used the data set as a single consecutive series, that is, they were arranged chronologically. After considerable research investigating several different approaches to hourly load forecasting, we decided on an hour by hour modeling strategy. In preliminary tests, this strategy proved superior in both fitting the data within the sample period and in forecasting 1989–90 without using this data in the analysis. We thereby rejected approaches using nonlinear and nonparametric models, time-varying parameter models, and general dynamic regression models. The basic strategy of the hour by hour model is very simple. In fact, it is the simplest strategy we considered. The forecasting problem of each hour is treated separately. That is, we estimate the load for the first hour of the day with one equation and the load for the second hour of the day from a different equation, and so on. Since we are forecasting for a full day, we have 24 separate forecasting equations. Furthermore, weekends and weekdays are separated so that there are actually 48 separate models. Each of these equations is very similar and is based only on information known at 8:00 A.M. on the day the forecast is made. In each case, there are four types of variables: deterministic, temperature, load, and past errors. A generic equation for HOUR1 can be written as

$$\begin{aligned}\mathrm{HOUR1} = {} & a\mathrm{DETERMINISTIC} \\ & + b\mathrm{TEMPERATURE} + c\mathrm{LOAD} \\ & + d\mathrm{PASTERRORS} + e\end{aligned} \tag{1}$$

DETERMINISTIC refers to variables that are perfectly predictable, such as the day of the week, the month of the year, and the year. In most cases, these effects are modeled by dummy variables. The monthly binary variables are important because the load changes over the winter not simply because the temperature changes. There are other annual patterns which must be picked up. The year variables are important because they model any long term growth in the regional economy and its demand for electricity. Although these could be modeled by binary variables, that would give no ability to forecast into the future. Instead, we use two trends; one is linear with the year and the other is its reciprocal. Clearly, an economic forecast of the regional economy could be much more accurate. The temperature variables are very important in modeling the short-run fluctuations in demand. Perhaps unsurprisingly, this is the most complex part of the model. Since the weather service forecasts temperature for each hour of the next day, the forecasting model can be estimated with the actual temperature even though it would not be known by the forecaster. The forecast is constructed simply by inserting the forecast of the weather into the model. The most important variable is the current temperature. Since Puget Power used the model only in the winter, there is no cooling load, so the relation between temperature and load is monotonic. The higher the temperature, the lower the demand. However, the relation is slightly nonlinear as the square of the temperature is also useful, giving a quadratic relation which has a steeper slope at very low temperatures. The temperature effect is not constant across months either. For many of the hours, an interaction between the monthly binary variables and the temperature is significant. In addition to the temperature at the hour for which the load is being forecast, the daily maximum temperature is important for both the forecast day and the previous day. Both the level and the square of this temperature are used. Finally, a seven-day moving average of past-midnight temperatures is used to allow for the effects of long cold spells which appear to have very significant augmenting effects on the loads. Altogether, there are twelve variables that depend upon temperature. These allow a rich specification of the weather component of the model. The variable LOAD is simply the load at 8:00 A.M. This is measuring the state of the system at the time of the forecast. The higher the load is at 8, the higher it is expected to be a day later. The size of this coefficient is, however, only about 10% so that most of the forecast is due to other factors. One naive model would be to forecast that the load tomorrow would be what it was today so that this coefficient would be close to 1 at least for HOUR8. Our results clearly reject this model. The load is allowed to have different impacts after a weekend or after a holiday. The lagged errors are an innovative feature of the model. At 8:00 A.M. it is known how accurate were the previous day's predictions of the model for the first 8 hours. It is also known how accurate were the predictions two days ago for all 24 hours. A wise judgemental forecaster would surely look at these and consider adjusting the forecast somewhat as a result. We approach this problem statistically by introducing the last five lagged errors into the model. We would generally expect that a positive coefficient would be appropriate, as factors which were missed yesterday would be expected to be partially recurring today. The fifth lag corresponds to the error from the same day the previous week since weekends are ignored in the weekday model. The coefficients of the most recent error vary from a high of 50% to a low of about 20%. Because the errors are partly caused by errors in the temperature forecast, which should not be corrected by the utility (unless they are trying to beat the weather service at its own forecast), the residuals to be used must be based upon the model forecast errors using the true temperatures. This feature of the model, allows it to recover relatively quickly from short lived changes in the load shape.

A criticism of the EGRV approach is that it is "only" a multiple regression model and, in particular, one with numerous correlated variables. The model aims to capture the very short-run dynamics of household behavior as the hours of the day change and as the household behavior interacts with temperature. People get up, go to school or work, return home, and so forth, and at each time different appliances are used. The model is not built using basic economic theory that suggests including the price of electricity and appliances or the stock of appliances because, in the very short run, these do not change. Thus, slowly changing variables such as increases in population, industrial growth, global warming, and so on are of little or no relevance in a model that is attempting to forecast from one day to the next. An adaptive model, introduced below, will adjust for any biases that result from the exclusion of these variables. We see from the results discussed later that the regression model with dynamic error structure has outperformed most of the complicated models. Utilities typically have personnel who can carry out the kind of analysis performed by EGRV. As for the second objection, it should be noted this is a forecasting model and hence we are not particularly interested in interpreting the signs of individual coefficients. It is well known that multicollinearity does not seriously jeopardize the forecasting ability of a model and often enhances it.

4.2. An example for HOUR1

Explanatory variables for HOUR1 include binary variables for the day of the week, month of the year, and yearly trends. They also include 1 A.M. temperatures, temperature squared, max of the day and max of the previous day and their squares, seven day average, and interactions of temperature and monthly binary variables. Finally they include the previous day's 8 A.M. load and its interactions with Monday and the day after a holiday. This model is estimated by ordinary least squares with a fifth-order Cochrane–Orcutt autoregressive error structure. Any version of such a Generalized Least Squares program will do and probably a simple iterative procedure would be adequate. Such an iteration would first estimate the model by OLS and then introduce the lagged residuals into the equation and re-estimate it by OLS again. The general model specified above has 31 coefficients and about 1000 observations. Such a model is not obviously overparameterized especially since many of the variables are orthogonal to each other. However, it is sensible to set coefficients to zero which are insignificant and then respecify the model in order to improve the precision of the remaining coefficients and to increase the powers of hypothesis tests. In the present case, 13 variables were eliminated. The final model estimated for 1983–1990 and used for the Fall/Winter 1990–91 forecast comparison is:

$$\begin{aligned}\mathrm{HOUR1} = {} & 1678.9 + 40.6\mathrm{YEAR} + 32.4\mathrm{DEC} + 0.5\mathrm{FEB} \\ & + 71.7\mathrm{MAR} - 193.9\mathrm{MONDAY} \\ & + 223.9\mathrm{DAYAFTERHOLIDAY} - 2.86\mathrm{TEMP} \\ & + 0.39\mathrm{TEMP}^2 \\ & - 7.68\mathrm{SEVENDAYAVGMIDTEMP} \\ & - 3.62\mathrm{YESTERDAYMAXTEMP} \\ & + 0.08\mathrm{LOAD8A.M.} \\ & + 0.07\mathrm{MONDAY{*}LOAD8A.M.} \\ & - 0.10\mathrm{DAYAFTERHOLIDAY{*}LOAD8A.M.} \\ & + 0.52e[t-1] + 0.15e[t-2] + 0.07e[t-3] \\ & + 0.14e[t-5] \end{aligned} \tag{2}$$

This model shows the typical form of all of the equations. Temperature, the seven day average of temperature, previous day's and maximum temperature are all negative reflecting the heating load. The lagged load contributes 8% to the forecast except on Mondays when it is $0.08+0.07-0.10$ or 5%, and after holidays when it is -2%, indicating a drop in load at 1 A.M. after a holiday. When the model was re-estimated for the 1991–92 forecasting experiment with the addition of the data from 1990–91, the same specification was used but the coefficients turned out to be slightly different. Generally the changes are small enough so that the model properties are unaltered.

4.3. Models across the hours

We have examined the forecasting model for HOUR1 in detail. A similar process is carried out for all 24 hours. Although the same variables are considered in each case to begin with, different variables finally remain in the model as the significance varies. This, in fact, is quite sensible. Some portions of the load curve may vary more over the year than others and consequently the monthly binary variables are important for modeling them. At some times of the day weather sensitive loads are more important than at others and at some times of the year the load may be more responsive to temperature than at others. We did, however, carry out a SUR estimation with common variables, but the differences in forecasts were negligible. We therefore chose to omit insignificant variables in order to improve efficiency. Most of the variables which are dropped in the equations are the monthly binary variables and their interactions with temperature. Overall, the final number of variables in each equation goes from a low of 14 to a high of 20 with an average of 16. Table 1 gives the estimated values for three representative hours (1 A.M., 8 A.M., and 5 A.M.) using 1990–91 data. It is particularly interesting to examine the effect of one variable across the hours. For example, the temperature coefficient for HOUR1 is -2.86 while for HOUR8 it is -25.03, or almost ten times larger. The effect of the heating load is expected to be much more important at 8 in the morning than at midnight. Presumably many customers have set-back thermostats which are relatively unresponsive to external temperatures until the morning. The general impression from examining the coefficients is that they do vary across hours in a rather systematic fashion. This observation confirms our earlier conclusion that building a separate model for each hour was far superior to assuming that the same model could be used for all hours. It is natural to try to estimate one very complicated model rather than 24 simpler ones. However, there are 393 coefficients in the 24 hourly models and it seems unlikely that a very effective approximation can be found which is sufficiently parsimonious to allow estimation, testing, and forecasting. A general impression of the success of this approach can best be achieved by examining a series of plots of the daily

Table 1
Model A coefficients at selected hours

Deterministic components	1 A.M.	8 A.M.	5 P.M.
Constant	1678.9	3586.4	2904.5
Year	40.6	95.3	88.3
Year inverse	0.0	151.9	0.0
October	0.0	−157.2	0.0
November	0.0	−118.0	0.0
December	32.4	−128.7	0.0
February	70.5	0.0	247.6
March	71.7	0.0	344.6
Monday	−193.9	0.0	0.0
Day after holiday (including Monday)	223.9	196.8	0.0
Friday	0.0	31.3	−53.6
Temperatue components			
Temperature	−2.86	−25.03	0.00
October × temperature	0.00	0.00	0.00
November × temperature	0.00	0.00	0.00
December × temperature	0.00	0.00	0.00
February × temperature	0.00	0.00	−6.33
March × temperature	0.00	0.00	−8.30
Temperature squared	0.39	0.13	0.34
Moving AVG of last 7 days midnight temps	−7.68	−8.60	−5.44
Maximum temperature	0.00	0.00	−12.99
Yesterday's maximum temperature	−3.62	−2.90	−1.83
Maximum temperature squared	0.00	0.00	0.00
Yesterday's max. temp. squared	0.00	0.00	0.00
Lag load components			
Load at 8:00 A.M. yesterday	0.08	0.13	0.07
Monday load at 8 yesterday	0.07	0.00	0.00
After holiday load at 8 yesterday	−0.10	−0.08	0.00
Autoregressive parameters			
Lag 1	0.52	0.40	0.00
Lag 2	0.15	0.18	0.37
Lag 3	0.07	0.00	0.29
Lag 4	0.00	0.00	0.10
Lag 5	0.14	0.00	0.12

forecasts. In Appendix D of the EPRI report the actual and fitted values are plotted for each day from November 1, 1990 through to March 3, 1991. The fits are generally impressive. Analysis of the worst fitting episodes indicates that large components of the load forecast errors are due to errors in the weather forecasts.

4.4. Statistical analysis

It is worth discussing the statistical theory behind this model in order to better understand why it works. Clearly, it seems surprising that the best forecast of HOUR1 and HOUR2 should be estimated separately from different data sets. This discussion shows why this procedure is not only statistically sound but has led to a very natural parameterization.

Each equation can be formulated as a regression with its own coefficients and error terms:

$$y_{ht} = x_{ht}\beta_h + u_{ht} \tag{3}$$

where h refers to the hour, t refers to the day and there are k variables in this regression. Some of these variables are deterministic, some are weather variables and some are lagged load variables. Many are interacted with binary variables. The error terms are assumed to follow an autoregression in days so that

$$u_{ht} = \rho_{h1} u_{ht-1} + \rho_{h2} u_{ht-2} + \cdots + \rho_{h5} u_{ht-5} + v_{ht} \quad (4)$$

The same model is more conveniently written as a panel model where each hour has its own equation. This is exactly the form of the hour by hour model. By grouping the variables we can write this as:

$$Y = X\beta + U \text{ and } E(uu') = \Omega \quad (5)$$

Eq. (5) is the well-known system of seemingly unrelated regressions or a SUR system. It is typically assumed that each row of U, which in our case corresponds to one day of hourly residuals, is contemporaneously correlated but not correlated over time. This assumption is reasonably accurate for this example except for two cases. The first has already been mentioned. There is correlation across days which is being modeled by Eq. (4). Thus, if the daily autocorrelation is taken into account in the x variables by including lagged errors, then the longer lagged correlations will not be present. Secondly, although the correlation between hours which are far apart can reasonably be taken as small, the correlation between hour 1 and hour 24 on the previous day may be large. Thus, the assumption of only contemporaneous correlation will not be strictly true for the first and last few hours. The beauty of the SUR model is that estimation of the system by Maximum Likelihood is accomplished by estimating each equation by ordinary least squares as long as the same variables enter each equation. This is discussed in any econometrics textbook; see for example Theil (1971) for an elegant proof. In our application, the same variables are considered for each equation, so this theorem applies exactly at the beginning of the specification search. Since insignificant variables are dropped from each equation, the final variables are different for different equations. This means that a slight improvement could be achieved by system estimation but it is presumably smaller than the gain from parsimoniously estimating each equation. This result explains why estimating each equation separately is a sensible procedure. It does not however say that forecasting from each separately is sensible, so we must now consider this question. The forecast from system (5) or system (3) has two components; the first is the systematic part and the second is the error forecast. This can be expressed as:

$$E_t(y_{ht+1}) = E_t(x_{ht+1})\beta_h + E_t(u_{ht+1}) \quad \text{for } h = 1, 2, \ldots, 24 \quad (6)$$

where the estimated value of β_h would be used with the weather service forecast of x_{ht+1} for the systematic part. The error forecast is generated from Eq. (4) as

$$E_t(u_{ht=1}) = \rho_{h1} u_{ht} + \cdots + \rho_{h5} u_{ht-4} + E_t(v_{ht+1}) \text{ for } h = 1, \ldots, 24 \quad (7)$$

assuming that u_{ht} is observed by the time the forecast is made. Since for some hours it is not, then the second lag is used both for estimation and forecasting. Because the ν_{ht+1} are correlated across hours, there is a possibility of forecasting the last term in Eq. (7). Suppose

$$E(v_{ht+1} | v_{h-1,t+1}) = \theta_{h-1} v_{h-1} \; h = 1, \ldots, 24 \quad (8)$$

then

$$E_t(v_{ht+1}) = (\theta_{h-1}\theta_{h-2} \cdots \theta_8) v_{8t} \quad (9)$$

Since v_{8t}, the 8 A.M. error, is the most recent error observed when forecasting, and all the forecasts are at least 16 hours in the future, the products are always products of at least 16 terms. Even if the θ's are rather close to one, the product of so many terms is likely to be very small so that there is little gain from forecasting this component. Thus forecasting each equation separately is also very close to optimal in this comparison. Improvements would be possible primarily if shorter horizon forecasts were required. Thus, with the parameterization in Eq. (3) there is little cost to the very simple procedure of estimating and forecasting each equation separately. The remaining question, is whether the equations in (3) are correctly specified. In particular, do they have a similarity which may be exploited in joint estimation? For example, suppose it was believed that some or all of the β's were the same for all hours. In this case, system estimation, possibly using a relation such as Eq. (5) would allow this constraint. Practically all of the other models estimated by both EGRV and the other participants in this experiment constrained many of the coefficients to be the same across hours. Various hypothesis tests of Eq. (5) continually revealed that the coefficients were not constant across periods. Consequently new x variables were introduced which interacted the old x's with dummy variables representing particular hours. Thus the number of β's in Eq. (3) became very large. It is our general impression that in this application, the real benefit comes from allowing the β's to differ across equations. Nevertheless, it remains possible that only a subset of the β's truly change over the hours or at least that the changes should be more smooth. Thus a compromise between completely free coefficients and some cross equation constraints could prove even better.

5. Adaptive models

To allow for learning, we used a second, adaptive version of the models described above. The model is designed to adjust the forecasts for a persistent error (this is in addition to the presence of lagged errors in the model). It is certainly possible that the economic environment changes in ways not known by the model builders during the year. A judgemental forecaster will slowly change his methods to compensate for systematic errors. The adaptive version of the model is designed to do just this, but using statistical analysis. This benefit must be offset by a loss of accuracy in the situation where the model structure remains constant. The adjustment used in this model is simply an exponential smoothing of the forecast errors of each hour using the following rule [See Ramanathan (1995), pp. 607–611 for a discussion of exponential smoothing] which is readily seen as a linear combination of forecast and adjustment errors.

$$\tilde{y}_{t+1} - \hat{y}_{t+1} = \phi(y_t - \hat{y}_t) + (1 - \phi)(\tilde{y}_t - \hat{y}_t)$$

where $\hat{y}_t$ is the hour by hour model forecast and $\tilde{y}_t$ is the adaptive version. Expanding this and rearranging terms, we obtain the following equation:

$$\tilde{y}_{t+1} - \hat{y}_{t+1} = \tilde{y}_t - \hat{y}_t + \phi(y_t - \tilde{y}_t) \tag{10}$$

Thus, a preliminary forecast $\hat{y}_{t+1}$ is first made. It is then adjusted, using Eq. (10), to generate the exponentially smoothed forecast $\tilde{y}_{t+1}$. A slow adjustment will alter forecasts only after errors become very persistent, while a faster adjustment will pick up errors more quickly but will also respond to pure noise some of the time. Because the model was estimated and specified over a fixed sample period, the specification used is effectively constant over this period. Hence the optimal smoothing appeared to be zero from our first experiments. We chose a parameter 0.02 which corresponds to a half-life of about 50 days. As can be seen in the forecast comparisons discussed below, this constant lead the two models to be very similar. Consequently, in the 1991–92 comparison, the constant was increased to 0.05 which corresponds roughly to a 20 day half-life. The benefits of this process were particularly apparent in the weekend model where the adaptive version performed much better than the original model.

6. Results

As mentioned earlier, the official forecast period was November 1 through to March 31 of the following year. This section presents a comparative evaluation of the various models used in the two competitions. The comparative statistics were compiled by Mr. Casey Brace of Puget Power and we summarize them here.

6.1. The 1990–91 forecast results

In the 1990–91 competition, 11 models were entered and these are as labeled below. However, to preserve the confidentiality of the other teams, who are well respected in the area of forecasting, only codes were supplied to us with brief descriptions of the approach (provided below). Our models are called A and B.

A. ANN2
B. STAT
C. FS
D. L
E. NAIVE
F. OH2
G. OH3
H. A (our basic hour by hour model)
I. B (the adaptive version of A)
I. REG

J. SHELF

ANN2 used recurrent neural networks, L refers to the judgemental forecast produced by Puget Power's own personnel, OH models are multilayered perceptron artificial neural networks, and SHELF used time-varying splines [see Harvey and Koopman (1993)]. Details about other models are not available to us but might be obtainable from Puget Power Company.

The criterion used by Puget Power to compare forecasts is the mean absolute percentage error (MAPE) for each hour, plus the A.M. peak period (7 to 9 A.M.) and the P.M. peak period (4 to 7 P.M.). Table 2 shows which forecasting method won each cell, with weekdays summarized separately from weekends. It is seen that one or the other of the EGRV models (labeled A and B) wins 82 of 120 weekday hours (that is, 68%), 13 of the 48 weekend hours (that is, 27%), 5 out of 7 A.M. peaks (71%), 4 of the 7 P.M. peak periods (57%), and 4 of the 7 totals. Table 3 shows the mean absolute percentage error (MAPE) for the A forecasts. It finds MAPE values for the daily average over 5 only for Saturday, Sunday, and Monday.

Puget Power received its forecasts of hourly temperatures from a local weather service and we were provided with these forecasts for the periods we were forecasting. Table 4 shows the MAPEs for the temperature forecasts and it indicates that Saturday, Sunday, and Monday's temperatures are forecast less well than other days. The Sunday and Monday forecast of temperatures are both issued on Friday and are for longer horizons than the other days, and so forecast errors are expected to be higher, with Monday worse than Sunday, as observed. However, why Saturday's temperature forecast should be worse than the other weekdays is a mystery, unless less

Table 2
Best forecast (lowest MAPE) for each cell

Time	Mon	Tue	Wed	Thur	Fri	Sat	Sun
1 A.M.	B	A	A	A	FS	SHELF	SHELF
2 A.M.	STAT	STAT	A	A	FS	L	STAT
3 A.M.	FS	STAT	B	SHELF	FS	A	FS
4 A.M.	FS	B	B	B	FS	A	SHELF
5 A.M.	STAT	B	B	B	SHELF	SHELF	SHELF
6 A.M.	B	FS	B	B	FS	B	B
7 A.M.	A	FS	B	B	FS	FS	B
8 A.M.	A	FS	A	B	FS	FS	B
9 A.M.	STAT	FS	A	A	FS	B	B
10 A.M.	STAT	FS	A	B	B	A	STAT
11 A.M.	FS	FS	A	A	A	B	L
NOON	FS	B	A	A	A	L	L
1 P.M.	L	B	A	A	A	L	L
2 P.M.	B	FS	B	B	B	L	B
3 P.M.	B	B	B	A	A	L	L
4 P.M.	A	A	B	B	B	L	A
5 P.M.	A	B	B	B	B	L	L
6 P.M.	L	A	B	B	B	L	FS
7 P.M.	FS	A	A	A	A	FS	L
8 P.M.	L	B	A	A	A	L	FS
9 P.M.	FS	B	A	B	A	L	L
10 P.M.	FS	B	B	B	A	L	L
11 P.M.	FS	A	A	B	L	SHELF	ANN2
MIDNIGHT	FS	L	B	A	A	SHELF	B
A.M. PEAK	A	FS	B	B	FS	B	B
P.M. PEAK	L	A	B	B	B	L	FS

"A" is our basic model, "B" is our adaptive model, and "SHELF" used time-varying splines. Details about other models are unavailable to us, but might be obtainable from Puget Sound.

Table 3
Mean percentage absolute forecast error for Model A

Time	Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Hourly Avg
1 A.M.	4.53	3.97	3.24	3.64	3.81	4.77	4.65	4.08
2 A.M.	4.67	4.51	3.14	4.47	4.54	4.26	5.50	4.44
3 A.M.	4.62	5.04	3.74	5.74	4.70	4.25	5.81	4.84
4 A.M.	5.10	5.08	3.75	5.74	5.01	4.54	5.96	5.03
5 A.M.	5.90	5.26	4.29	5.87	4.97	5.38	6.61	5.47
6 A.M.	5.59	5.54	4.39	4.94	5.35	6.11	7.70	5.66
7 A.M.	5.47	5.12	4.13	4.64	5.33	6.62	6.66	5.43
8 A.M.	4.89	4.47	3.21	3.89	4.45	6.49	5.90	4.76
9 A.M.	4.69	4.18	2.64	3.36	4.07	5.20	5.44	4.22
7–9 A.M.	5.02	4.59	3.32	3.96	4.61	6.10	6.00	4.80
10 A.M.	4.53	3.75	2.93	2.81	3.93	4.02	4.96	3.84
11 A.M.	5.12	3.42	3.29	3.28	3.65	4.31	5.21	4.04
12 P.M.	4.81	3.08	3.20	4.09	3.50	4.64	5.51	4.12
1 P.M.	5.31	3.13	3.55	4.43	3.80	5.47	6.16	4.55
2 P.M.	5.60	3.87	3.67	4.34	4.44	6.09	7.10	5.02
3 P.M.	5.17	4.08	4.21	4.36	6.37	6.22	6.73	5.31
4 P.M.	4.63	4.01	4.59	5.56	6.15	6.95	6.00	5.43
5 P.M.	5.11	4.10	4.48	5.21	5.92	6.77	5.88	5.36
6 P.M.	5.43	4.21	4.53	4.97	4.94	5.27	5.22	4.94
7 P.M.	5.23	3.53	3.51	3.90	3.84	4.34	4.99	4.19
4–7 P.M.	5.10	3.96	4.28	4.91	5.21	5.83	5.52	4.98
8 P.M.	5.73	5.53	2.99	3.99	3.65	4.43	5.79	4.30
9 P.M.	5.83	3.84	3.05	3.98	3.97	4.54	5.75	4.42
10 P.M.	5.33	4.06	2.75	4.16	4.38	4.43	5.77	4.40
11 P.M.	5.06	4.76	3.27	4.61	4.34	4.81	5.17	4.57
12 A.M.	4.43	5.44	5.06	4.68	4.59	5.31	5.73	5.04
Day Avg	5.12	4.25	3.65	4.44	4.57	5.22	5.84	4.73

effort is made to produce this forecast. As all the models relied heavily on temperature forecasts, the quality of our forecasts will decrease as the temperature forecasts get poorer, which possibly explains why the EGRV models do less well on Saturday, Sunday, and Monday.

There were two distinct cold snaps during the 1990-91 forecasting competition. In each the MAPE of the EGRV models increased and our overall rankings dropped — to second in the first cold snap (19–23 December 1990) and to fifth in a brief second snap (28–30 December 1990). A detailed analysis of the MAPEs for the A model around and including the first cold snap shows that some large biases occur and MAPE values are generally higher. However, the temperature forecasts for this period are also worse than usual. If the local service is unable to forecast cold spells, this implies that models such as EGRV's that rely on these forecasts will produce less satisfactory forecasts of load during such spells. It would be useful, however, to investigate why some of the other models (which also were faced with poor weather forecasts) did considerably better than the EGRV models.

6.2. The 1991–92 forecast results

Based on the above experience, we made a number of improvements in the model specifications. In the 1991–92 winter competition, all the models were respecified to some extent and several new models were added. Altogether now 14 models were compared. These are as labeled below:

A. ANN2
B. GJP
C. L
D. OH

Table 4
Mean absolute percentage error in forecast temperatures

Time	Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Hourly Av
1 A.M.	9.23	6.25	4.77	7.18	6.14	8.18	7.58	7.04
2 A.M.	10.03	6.39	4.64	7.93	7.23	7.31	8.31	7.39
3 A.M.	10.26	6.72	6.10	8.72	5.00	8.96	8.46	7.73
4 A.M.	11.09	6.42	5.54	7.51	5.63	8.78	7.78	7.52
5 A.M.	11.38	6.81	6.74	7.76	5.32	9.08	8.58	7.94
6 A.M.	12.12	7.50	6.81	7.79	5.46	10.64	8.87	8.44
7 A.M.	10.53	6.88	6.53	6.68	5.54	9.70	7.31	7.58
8 A.M.	10.14	6.04	6.01	7.70	5.41	10.89	8.01	.7.74
9 A.M.	9.77	5.68	6.41	7.63	5.29	10.38	7.68	7.54
7–9 A.M.	10.15	6.20	6.32	7.33	5.41	10.32	7.67	7.62
10 A.M.	8.98	4.86	6.34	5.44	5.60	9.50	8.86	6.94
11 A.M.	8.55	5.78	6.83	6.53	7.05	9.52	7.61	7.41
12 P.M.	8.46	5.30	7.46	7.12	7.86	9.30	6.78	7.33
1 P.M.	8.62	6.01	6.20	7.09	8.03	7.03	6.04	7.00
2 P.M.	10.51	6.67	5.71	7.36	9.49	7.10	8.25	7.86
3 P.M.	9.85	7.24	6.33	7.27	9.46	8.05	7.55	7.96
4 P.M.	9.86	6.14	6.08	7.32	8.78	8.56	7.76	7.78
5 P.M.	9.37	7.07	6.28	6.82	7.81	8.08	9.54	7.85
6 P.M.	8.50	5.78	5.62	6.10	7.26	8.86	9.89	7.43
7 P.M.	9.11	6.70	6.77	6.61	7.95	7.63	10.67	7.92
4–7 P.M.	9.21	6.42	6.19	6.71	7.95	8.28	9.47	7.57
8 P.M.	9.26	8.63	7.04	8.20	8.34	8.48	10.01	8.56
9 P.M.	9.20	8.71	8.68	8.18	7.48	9.67	9.81	8.82
10 P.M.	9.85	10.18	8.46	8.37	8.03	9.63	9.52	9.14
11 P.M.	10.47	11.04	9.51	9.15	9.15	8.29	8.68	9.46
12 A.M.	10.11	10.94	9.88	8.85	7.88	8.56	8.64	9.25
Day Av	9.80	7.07	6.70	7.47	7.13	8.80	8.38	7.90

E. A (EGRV's basic hour by hour model)
F. B (the adaptive version of A)
G. RAHMAN
H. SHELF
I. SMU1
J. SMU2
K. SMU3
L. SMU4
M. SMUA
N. SMUB

GJP used the Box–Jenkins transfer function intervention-noise methodology. SMU models are based on adaptive multi-layered feed-forward neural networks trained with back-propagation [see Brace, et al. (1993) for more details on these and on RAHMAN). The forecasts are tabulated separately for Tues–Fri, Monday and weekends. In each case the morning peak, afternoon peak and daily mean absolute percentage errors (MAPE) are computed and tabulated by Puget Power in Table 5. Of the six weekday tabulations, F (our adaptive Model B) was the best model for the four peak periods and was beaten only by E (our basic Model A) for the Tues–Fri daily average. Several models did slightly better for the Monday average. Our basic Model A (E in Table 5) was the second best model for each of the peak periods.

On weekends the results are also good but not overwhelming. The adaptive model B (F in Table 5) is best for weekend afternoons and the basic Model A (E in Table 5) falls slightly into third behind L. For weekend mornings SMU4 (labelled L in Table 5) was the best, and for overall average, model N, SMUB is the best. However, it is also the worst for weekend afternoons, so there is some instability there.

In order to understand further the strengths and

Table 5
Mean absolute percentage error (MAPE) for 1991–92

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Tue–Fri														
A.M.	3.70	4.22	3.40	4.40	3.32	**3.18**	4.05	3.86	4.83	4.65	4.18	4.09	3.86	4.09
P.M.	4.00	4.25	3.57	5.13	3.29	**3.28**	3.67	4.04	4.77	4.55	4.35	3.89	5.53	5.53
Daily	2.77	3.28	2.73	3.59	**2.45**	2.55	2.68	2.87	2.91	3.06	2.57	2.65	2.88	2.85
Weekend														
A.M.	NA	6.05	5.68	NA	7.86	6.38	5.69	5.59	5.71	5.71	5.71	**4.66**	5.71	
P.M.	NA	4.73	4.10	NA	4.13	**3.82**	4.67	4.65	4.21	4.21	4.21	4.21	6.35	6.35
Daily	NA	3.47	3.56	NA	4.15	3.53	3.66	3.80	3.64	3.64	3.64	3.64	3.42	**3.30**
Monday														
A.M.	6.71	6.36	6.05	6.92	5.99	**5.75**	7.83	6.34	6.32	6.45	6.38	6.19	6.21	6.19
P.M.	5.11	5.62	4.94	6.54	4.74	**4.64**	6.53	5.70	5.60	5.86	5.39	5.59	7.70	7.70
Daily	5.81	4.60	5.26	6.40	4.97	4.63	5.75	5.24	4.61	4.85	**4.49**	4.51	5.19	4.95

Entries in bold type refer to the best model for that period.

weaknesses of the EGRV models, we did our own decomposition of MAPE by hour for each model and mean percentage error (not presented here). The latter measure does not take absolute values and therefore can be positive or negative. If the forecasts are biased then the MPE will exhibit the bias. Both models have particular biases on Monday from hours 1 through to 7. These show up in high MAPE and in the much higher Monday averages. By the time of the peak hours, the biases are over so that the weaker performance on Mondays than other days can basically be attributed to these hours. Since the problem shows up in a bias it should be possible to correct it through respecification of these equations. A similar set of biases occur, particularly in A on Sunday and to some extent on Saturday mornings. These however, extend over into the peak period resulting in inferior peak performance.

Thus, the overall analysis of the forecast errors indicates that the weaknesses of these models are in the very early mornings of Saturday, Sunday, and Monday. If a model is to be inaccurate somewhere, this seems as good a place as anywhere. The overall impression especially in comparison with the other models is of extraordinarily good performance which could possibly be improved even a little more. Although a detailed analysis of cold snaps was not carried out with the 1991–92 data, preliminary study indicates that the days when the model does not do well also coincide with days when the weather service forecasts were not very accurate. This was the pattern noticed earlier in the 1990–91 data.

7. Conclusions

We have developed a simple and flexible set of models for hourly load forecasting and probabilistic forecasting. These have performed extremely well in tightly controlled experiments against a wide range of alternative models. When the participants were able to revise their models along the lines of our successes of 1990–91, they still were unable to equal the performance of the hour by hour model presented here. Although the model has many parameters, the estimation method is very simple; just ordinary least squares with general order serial correlation. It should be possible for other utilities to adapt this approach for their use. An econometrician familiar with the use of least squares software should be capable of handling general order autoregressive errors. Data handling and forecasting can be done using spreadsheet software, although more user friendly code can be written directly.

A useful extension to the present project would be to use optimum combination techniques that are likely to be superior to individual forecasts. Another useful extension would be to use "switching-regime" techniques, especially during cold-snaps. Thus, when forecasts are going badly during abnormally cold spells a different model will "kick-in" that, hopefully, will yield better forecasts. This approach has been investigated by Deutsch et al. (1994).

Acknowledgments

We are grateful for the financial support of the Electric Power Research Institute (Project No. RP2919-04) and to its project manager Phillip Hummel. Thanks are also due to Peter Cleary, former manager of the project for EPRI, for his encouragement and enthusiastic support of the project, and to Allan Mitchem of the Virginia Electric Power Company for sharing the details of their hour by hour model. We also acknowledge the invaluable help of our research assistants Chor Yiu Sin, Megan Werner, and Diana Weinhold. Finally, thanks are also due to the referees of this paper.

References

Ackerman, G., 1985, Short-term load prediction for electric-utility control of generating units, in: D.W. Bunn and E.D. Farmer, eds., Comparative Models for Electrical Load Forecasting (Wiley, New York).

Brace, M.C., V. Bui–Nguyen and J. Schmidt, 1993, Another look at forecast accuracy of neural networks, Paper presented to the Second International Forum on the Applications of Neural Networks to Power Systems, April 19–22, 1993, at the Meji University, Yokohama, Japan.

Bunn, D.W. and E.D. Farmer, 1985, Comparative Models for Electrical Load Forecsting (Wiley, New York).

Deutsch, M., Granger, C.W.J., Teräsvita, T., 1994, The combination of forecasts using changing weights. International J. of Forecasting 10, 47–57.

Electric Power Research Institute, 1977, Forecasting and Modeling Time-of-day and Seasonal Electricity Demands, (EPRI EA-578-SR, Palo Alto, California).

Electric Power Research Institute, 1993, Probabilistic Methods in Forecasting Hourly Loads (EPRI TR-101902, Palo Alto, California).

Gupta, P.C., 1985, Adaptive short-term forecasting of hourly loads using weather information, in: D.W. Bunn and E.D. Farmer, eds., Comparative Models for Electrical Load Forecasting (Wiley, New York).

Harvey, A., Koopman, S.J., 1993, Forecasting hourly electricity demand using time-varying splines. J. Amer. Stat. Assoc. 88, 1228–1237.

Ramanathan, R., C.W.J. Granger and R. Engle, 1985, two-step modelling for short-term Forecasting, in: D.W. Bunn and E.D. Farmer, eds., Comparative Models for Electrical Load Forecasting (Wiley, New York).

Ramanathan, R., 1995, Introductory Econometrics with Applications, Third Edition (The Dryden Press, Forth Worth, Texas).

Schneider, A.M., T. Takenawa and D.A. Schiffman, 1985, 24-hour electric utility load forecasting, in: D.W. Bunn and E.D. Farmer, eds., Comparative Models for Electrical Load Forecasting (Wiley, New York).

Theil, H. (1971), Principles of Econometrics, New York: John Wiley.

Biographies: Robert ENGLE, Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, Phone (619) 534-4453, FAX (619) 534-7040. Professor Engle received his Ph.D. from Cornell University in 1969. He is presently a Senior Professor of Economics at the University of California, San Diego, where he was a chairman from 1990 to 1994. He is also a Fellow of both the American Academy of Arts and Sciences and the Econometric Society. He is an expert in time series analysis with a long time interest in the analysis of financial markets. He has published over a hundred academic papers and three books and has pioneered such innovative statistical methods as ARCH, cointegration, band spectrum regression, and most recently, common features.

Clive W.J. GRANGER, Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, Phone (619) 534-3856, FAX (619) 534-7040. Professor Granger received a Ph.D. from Nottingham in 1959. He is presently a Senior Professor at the University of California, San Diego, where he was a chairman from 1982 to 1984. He has been a Visiting Fellow at All Souls College, Oxford, and a Visiting Fellow Commoner at Trinity College, Cambridge. He has been a Guggenheim Fellow and is also a Fellow of the American Academy of Arts and Sciences, the Econometric Society, and the International Institute of Forecasters. His primary interests are in time series analysis, forecasting, applied econometrics, and finance, and has written over 200 academic papers and ten books.

Ramu RAMANATHAN, Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, Phone (619) 534-6787, FAX (619) 534-7040. Professor Ramanathan received his Ph.D. in 1967 from the University of Minnesota. He is presently a Senior Professor of Economics at the University of California, San Diego, and the Director of Undergraduate Studies in Economics. His areas of interest are applied econometrics, economic growth, and international trade. His articles in these areas have appeared in leading journals in the United States, United Kingdom, Australia, and India. He is also the author of three books and a co-author of two other books.

Farshid VAHID–ARAGHI, Department of Economics, Texas A and M University, College Station, TX 77483-4228, Phone (409) 845-0676, FAX (409) 847-8757. Farshid Vahid received his Ph.D. from the University of California, San Diego in 1994. He is presently an Assistant Professor of Economics at Texas A and M University. His research interests include applied econometrics and multivariate time series analysis.

Casey BRACE, Puget Sound Power and Light Company, P.O. Box 97034, Bellvue, WA 98009-9734, Phone (206) 462-3125, FAX (206) 462-3175. Milan Casey Brace is a Senior Consulting Engineer with Puget Sound Power and Light Company. He has been developing short and long range forecasting models for the utility industry for over twenty-five years. He is currently applying decision analysis to a wide range of utility problems.