INDIAN INSTITUTE OF MANAGEMENT CALCUTTA

WORKING PAPER SERIES

WPS No. 656/ May 2010

Modeling Regional Electricity Load in India

by

Ashok Banerjee
Professor, IIM Calcutta, Diamond Harbour Road, Joka P.O., Kolkata 700 104 India

&

Sahadeb Sarkar
Professor, IIM Calcutta, Diamond Harbour Road, Joka P.O., Kolkata 700 104 India

Title: Modeling Regional Electricity Load in India

Authors:

Ashok Banerjee
Professor (Finance & Control)
Indian Institute of Management Calcutta
Diamond Harbour Road
Joka
Kolkata 700 104
INDIA
Email: ashok@iimcal.ac.in
Phone Number: 91-33-2467-8300
Fax Number: 91-33-2467-8307

Sahadeb Sarkar*
Professor (Operations Management)
Indian Institute of Management Calcutta
Diamond Harbour Road
Joka
Kolkata 700 104
INDIA
Email: sahadeb@iimcal.ac.in

*** Corresponding author**

<u>**Modeling Regional Electricity Load in India**</u>

Abstract

Electricity as a product cannot generally be stored. Hence, it is required to match demand and supply on a real time basis in order to avoid disturbance in grid frequency and consequently ensure quality of power supply. The agencies involved in scheduling of power in India divide a day into ninety-six time buckets - each bucket of fifteen minutes duration. The load is matched for each time bucket. This matching is done the night before the start of actual dispatch. In fact, the demand supply schedule for a day is finalized at 11 PM on *the previous day*. The regulated bulk supply tariff in India has an unscheduled interchange component, which is linked to grid frequency. Thus, any smart generator of electricity would make abnormal profits provided she is able to forecast the load properly.

The present paper attempts to model the hourly load in the northern grid in India. The hourly load is estimated by aggregating load over five-minute intervals. Data corresponding to each hour is treated as a single time series and each series is modeled independently. The paper has used an estimation window of eleven months and forecast window of one month. Autoregressive models with dummy variables to capture (a) day-of-the-week effect, (b) national non-Sunday holiday effect, and (c) seasonality effect turn out to be quite effective in explaining the hourly load behavior.

The results show evidences of clustering of load behavior. Five clusters of time period within a day were observed where the load behavior can be captured with a single model. These clusters take care of sixteen hours of a day. Each of the remaining eight hours behaves differently. Interestingly, the behavior of load around mid night and during morning hours of the day does not depend on day of the week. Furthermore, Thursdays and Saturdays had the least impact on the hourly load. Another interesting finding is that there is no effect of national non-Sunday holidays on the load. The mean absolute percentage error of the best-fit model in calculating one-day-ahead out-of-sample log-load forecast is quite small (ranging from 0.14% to 0.20%).

The findings of the paper may have profound implications for the regulator - (a) the regulator can use the information to introduce time-of-the-day pricing; (b) a fifteen-minute time bucket may be desirable for load scheduling, but for load forecasting one may use the clusters. Thus, the electricity generator may schedule its generation based on cluster-wise load forecast, rather than hourly forecast.

1. Introduction

Electricity as a product cannot generally be stored. Hence, it is required to match demand and supply on a real time basis in order to avoid disturbance in grid frequency and consequently ensure quality of power supply. The agencies involved in scheduling of power in India divide a day into ninety-six time buckets- each bucket of fifteen minutes duration. The load is matched for each time bucket. However, there must be some reserve margin in the system so that local disturbances do not lead to collapse of the grid. The issue of developing an appropriate mechanism for load forecasting is important in the following ways:

(a) In order to ensure minimum disturbance in the grid frequency 'it is relevant for electricity systems optimization to develop a scheduling algorithm for the hourly generation and transmission of electricity'([1]). Hourly load forecasts are one of the main inputs to this algorithm.

(b) The regulators in India are seriously discussing the possibility of introducing time-of-the-day pricing for bulk supply tariff. A proper understanding of intra-day load behaviour is a prerequisite for introducing such pricing system.

(c) The regulated bulk supply tariff in India has an unscheduled interchange component, which is linked to grid frequency. Thus, any smart generator of electricity would make abnormal profits provided she is able to forecast the load properly.

Forecasting electricity load demand has been a favorite subject of many researchers ([2]). Recently [3] followed a two-level seasonal autoregressive model and observed that this model reported better forecast results during the peak hours (i.e., hours 19-21). The paper

[4] fitted an autoregressive moving average model (ARMA) to estimate electricity loads in California power market and obtained an acceptable out of sample forecast. The obtained residuals seemed to be independent but with tails heavier than the Gaussian tails. Another paper [5] applied a multiple regression model for each hour and found the model performing extremely well.

The present paper attempts to model the hourly load in the northern grid in India. The hourly load is estimated by aggregating load over five-minute intervals. Data of each hour is treated as a single time series and each series is modeled independently. The results show evidences of clustering of load behavior.

The plan of the paper is as follows. The next section briefly explains the Indian electricity scenario. Section 3 describes the data and the model proposed to fit the load demand, while Section 4 analyses the results. Section 5 offers some concluding remarks.

2. Indian Scenario

The electricity industry in India is mired in a complex network of problems. They range from inadequate capacities in generation, transmission and distribution, outdated technologies especially in transmission and distribution, poor maintenance, cross-subsidization and consequent financial non-viability of a large part of the sector, over staffing, lack of a commercial culture, poor management and accounting practices, etc. The grid situation in India witnessed low frequency during peak load hours; rapid and wide changes in frequency – 1 Hz change in 5 to 10 minutes, for many hours every day; and very frequent grid disturbances, causing tripping of generating stations, interruption of supply to large blocks of consumers, and disintegration of the regional grids. The regulator has introduced a tariff mechanism (called, the availability based tariff) that would address the grid behaviour and at the same time would give incentive to participants for restoring grid discipline.

The heart of the availability based tariff (ABT) system is scheduling wherein both the generator and the subsequent beneficiary has to pre-commit themselves to day ahead schedules via declaration of plant availability status on ex-bus basis and drawal schedule at nodal intake point. The methodology is as framed below:

(i) Each day starting from 00.00 hrs will be divided into 96 time blocks of 15 minutes intervals.

(ii) The generator will make an advance declaration of capability of its generating station by 10:00 hrs. The declaration will be for that capability which can be actually made available.

(iii) The regional load dispatch centres (RLDC) shall communicate the respective shares to the beneficiaries based on the declaration of the generator by 11:00 hours.

(iv) Based on the requisitions given by the beneficiaries by 15:00 hours and taking into account transmission system constraints, if any, RLDC prepares the economically optimal generation schedules and drawal schedules and communicate the same to generator and beneficiaries by 17:00 hours. RLDC also formulates the procedure for meeting contingencies both in the long run and in the short run (daily scheduling). All the scheduled generation and actual generation shall be at the generator's ex-bus. For beneficiaries, the scheduled and actual net drawals are at their respective receiving points. For calculating the net drawal schedules of beneficiaries, the transmission losses are apportioned to their drawals.

(v) The generators can however revise their schedule, which they intimate to the RLDC's by 22:00 hours.

(vi) Based on the final revised documents received by the RLDC, it draws up the final schedule by 23:00 hours and issues the schedule to both generators and beneficiaries alike. The new schedule comes into application at 00:00 hours.

The commercial mechanism of the ABT contemplates the disciplining of all three entities in the grid viz., the generator, transmitter and the beneficiaries. It accords a uniform treatment to all participants in the grid. The basic advantage in ABT is that the total tariff payable by the beneficiary to the generating station is divided into 3 components viz. 1) Capacity charge 2) Energy charge and 3) the unscheduled Interchange (UI) charges. Variation in actual generation/drawal and scheduled generation/drawal is accounted for through Unscheduled Interchange (UI) charges.

Though the UI charges are primarily intended to act as a penalty charge thus preventing the tendency of generators to over-generate in times of high frequency and the states to involve in under-drawal via the penalty mechanism, yet its very nature also provides a market pricing mechanism for the sale of power according to the intensity of demand.

All the generators irrespective of ownership would be dispatched with frequency based dispatch guidelines where at each frequency level, output of the generators are regulated by comparing their own variable cost with the frequency linked UI price. Thus, merit order scheduling as well as merit order dispatch of generators are ensured. Any generator can take advantage of a windfall in the form of UI, provided it can perform the hourly load forecast well.

3. Data and Methodology

The data consist of electricity load demands at five-minute intervals observed from January 1, 2005 through December 31, 2005 (1,05,120 data points). For each day, the hourly load demand is then estimated by averaging the load demands observed at twelve five-minute intervals. Data corresponding to each hour over the days is treated as a single time series ("sectional data") and each of these 24 daily series is modeled independently of others (Figure 1a-1x). Modeling these "sectional" data is quite appropriate since the objective is to obtain 24-hour ahead forecasts. This helps avoid complex intra-day patterns in the hourly data and allows each daily series corresponding to an hour to capture a distinctive weekly behavior. The day of the week is expected to affect the load demand during the middle hours, when business and industrial activities are it their peak, more than the first and last hours of the day. This hour-by-hour approach has been adopted by [1] and [5] among others. However, we have used the lags of the load data and binary dummy variables representing the week-day effect, national non-Sunday holiday effect and seasonal effects as explanatory variables in linear regression models. No external variables such as the hourly average temperature of the day have been used in our study as such data are not available to us. Such variables are expected to improve prediction of the load demand ([6]). However, for our data the linear regression models considered by us performed quite well.

For each hour h of the day we have computed the average electricity load and then computed its logarithm r_t^h on day t, where h = 1, 2, …, 24. We chose to model with the logarithm of load demands since it allows one to model weekly seasonality and national non-Sunday holiday effect through simple linear models. For fixed h, we have considered various linear regression models ([7]), to explain the mean-behavior of the series r_t^h, and tried GARCH models ([8], [9]) to capture possible heteroscedasticity of the series.

Let

$$\mu_t^h = E(r_t^h \mid F_{t-1}^h), \qquad {\sigma_t^h}^2 = V(r_t^h \mid F_{t-1}^h) = E((r_t^h - \mu_t^h)^2 \mid F_{t-1}^h) \qquad (1)$$

where F_{t-1}^h is the information made available upto time (t−1). The series $\{r_t^h\}$ may be either serially uncorrelated or may have minor lower order serial correlations, but it may yet be dependent. To capture such possible dependence in series volatility models may be used. Generally, the conditional mean μ_t^h of such series $\{r_t^h\}$ can be modeled using a simple model such as a stationary AR(p) model with some additional variables like dummy variables to account for effect of a day of the week, and national non-Sunday holidays, and also the seasonal effect, i.e.,

$$r_t^h = \mu_t^h + a_t^h, \qquad \mu_t^h = \phi_0^h + \sum_{i=1}^{p} \phi_i^h r_{t-i}^h + \sum_{j=1}^{6} \beta_j^h D_{j,t}^h + \beta_7^h H_t^h + \sum_{k=1}^{2} \gamma_j^h S_{k,t}^h , \qquad (2)$$

where the shock a_t^h represents a white noise series with mean zero and variance ${\sigma_a^h}^2$, and p a non-negative integer. We have used dummy variables $D_{j,t}^h$ and $S_{k,t}^h$. For j = 1, 2,

…, 6, the variables are defined as follows: $D^h_{j,t} = 1$ if t = day j of the week and $D^h_{j,t} = 0$, otherwise. The binary variable H^h_t is defined as: $H^h_t = 1$ if t = a national non-Sunday holiday and $H^h_t = 0$, otherwise. Exploratory data analysis indicated that the regional electricity load is affected by roughly 3 seasons, namely, winter, summer and fall (Figure 2). For three seasons only two dummy variables $S^h_{k,t}$ are required to be included in the model. For k = 1, 2, the variable $S^h_{k,t}$ is defined as $S^h_{k,t} = 1$ if t belongs to season k and $S^h_{k,t} = 0$ otherwise.

From equations (2) and (3) we have

$$\sigma_t^{h\,2} = V(r_t^h \mid F_{t-1}^h) = V(a_t^h \mid F_{t-1}^h) \qquad (3)$$

In model (2), the unconditional variance $V(a_t^h)$ may be constant, yet the conditional variance $\sigma_t^{h\,2} = V(a_t^h \mid F_{t-1}^h)$ may depend on t. Volatility models attempt to express the evolution of $\sigma_t^{h\,2}$, or its positive square root σ_t^h using an exact function or a stochastic equation. The equation for μ_t^h is called the *mean equation* for the r_t^h, and that for $\sigma_t^{h\,2}$ its volatility (or conditional variance) equation. A GARCH model describes the volatility evolution through a simple parametric function. A detailed discussion on GARCH models can be found in [9] and [10].

4. Results

We have used an estimation window of eleven months (January- November 2005) and forecast window of one month (December 2005). We ran various regression models for each hour separately with the data specified in the previous section to find the best fit. The results are reproduced in Table 1. It is observed that AR(2) with dummies is the best fit model for all the hourly time series. The coefficient of any higher order lags proved to be insignificant. The usual diagnostics of residuals along with results from GARCH models (not reported here for brevity) indicate absence of heteroskedasticity in the data. This result is in conformity with that in [4].

Table 1 reports only the models where the coefficients were significantly different from zero (at 10% level). An interesting feature of the result is that there is no effect of national (non-Sunday) holidays (β_7) on the load (in the northern grid of India). The paper considered the holidays as per Government of India's Gazette of holidays. This implies that factories and offices in the private sector do not follow holidays as per Government of India's Gazette. Seasonal dummies were used for winter (1 January- 15 April) and summer (16 April- 31 August). The seasonal dummy variable corresponding to the period Sep 1 – Dec 31 was not included in the model in the presence of the intercept variable. The negative coefficients (γ_1) of winter dummy indicate usual inverse relationship between load and the winter season. The northern region of India experiences severe summer and winter. This relationship also implies that electricity load is more influenced by running of air conditioners rather than room heaters. The results also show that hourly load behavior does not always depend on the day of the week. In fact, Thursdays (β_4) and Saturdays (β_6) had the least impact on the hourly load.

The models reported in Table 1 also show quite high adjusted R^2 values, which indicate the strength of goodness-of-fit. The Durbin-Watson (DW) statistic values and Ljung-Box (LB) test P-values indicate validity of the independence assumption of the errors.

Results show similar load behavior in certain clusters. For example, clusters were observed during 2300-0300 hours, 0500-0900 hours, 1400-1700 hours, and 2000-2300 hours. This load behaviour makes sense. For example, during early hours of the day, most of the load arises from continuous process industry and residences. This load remains same throughout the week.

These clustering of load behavior indicate that the regulators may use these clusters for time-of-the-day pricing. In other words, although hourly models have been used, time-of-the-day pricing need not be different in every hour.

After obtaining the best-fit model- AR(2) with dummies, the out-of-sample forecast is made for all the days of December 2005. We compute the mean absolute percentage error (MAPE) for each hourly forecast. The results are reported in Table 2. It is observed that the transmission lines in the Northern Grid suddenly tripped during 22-23 December 2005 due to heavy fog coupled with pollution ([11]). This disturbance is reflected in sudden fall in electricity load during 22-25 December 2005. Such a break in the data set is difficult to forecast well with the constructed best-fit model unless modified appropriately. Hence, Table 2 shows two sets of MAPE numbers - one for the entire month of December 2005 and the other for the normal days (excluding the days of grid disturbance) of December 2005. Our forecast results are quite encouraging. The MAPE ranges between 0.14% to 0.20% for normal days of December 2005. This is much better than that reported in Table 2 of [1] .

5. Conclusions

The findings of the paper may have profound implications for the regulator - (a) the regulator can use the information to introduce time-of-the-day pricing; (b) a fifteen-minute time bucket may be desirable for load scheduling, but for load forecasting one may use the clusters or individual hours. Thus, the electricity generator may, in some cases, schedule its generation based on cluster-wise load forecast, rather than hourly forecast. The paper observes five clusters covering sixteen hours of a day. The forecast for the remaining eight hours of a day can be made using respective hourly models. The paper did not use any external variables such as the hourly average temperature of the day. Inclusion of such variable may further improve the prediction of load demand. The paper used data from only one region of the country. We intend to cover the load of other regions in our future research.

References

[1] Soares, L.J., and Souza, L.R. (2006), 'Forecasting electricity demand using generalized long memory', *International Journal of Forecasting* 22, 17-28.

[2] Bunn, D. and Farmer, E., *Comparative Models for Electrical Load Forecasting*, New York: Wiley, 1985.

[3] Soares, L.J. and Medeiros, M.C.,'Modeling and forecasting short-term electric load demand: a two-step methodology', *Discussion Paper* No. 495, Departmento De Economia, Brazil.

[4] Nowicka-Zagrajek, J. and Weron, R. (2002), 'Modeling electricity loads in California: ARMA models with hyperbolic noise' *Signal Processing* 82, Issue 12, 1903-1915.

[5] Ramanathan, R., Engle, R., Granger, C. W. J. , Vahid-Araghi, F. & Brace, C. (1997), 'Short-term forecasts of electricity loads and peaks', *International Journal of Forecasting*, 13, 161-174.

[6] Hippert, H. S., Pedreira, C. F. & Souza, R. C. (2001), 'Neural Networks for short-term load forecasting: A review and evaluation', *IEEE Transactions and Power Systems*, 16, 44-55.

[7] Draper, N., and Smith, H., *Applied Regression Analysis*, 2nd Edition, Wiley, 1981.

[8] Alexander, C., Market *Models: A Guide to Financial Data Analysis*, John Wiley & Sons Ltd, 2001.

[9] Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics,* 31, 307-327.

[10] Bollerslev, T., Engle, R.F., and Nelson, D.B., ARCH Models, in *Handbook of Econometrics*, R.F. Engle and D.L. McFadden (eds.) 4, Elsevier Science B.V. 1994.

[11] Agenda for Joint Meeting of 372nd OCC and 121st Commercial committee of Northern Regional Electricity Board, 13 January 2006.

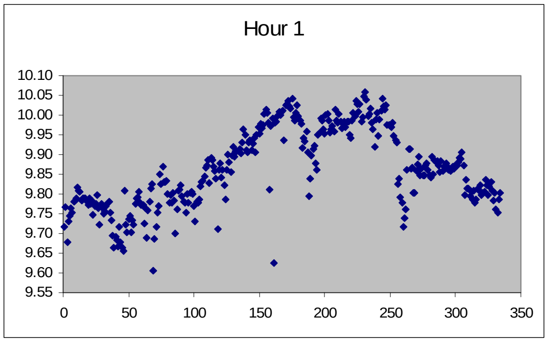

Fig. 1a

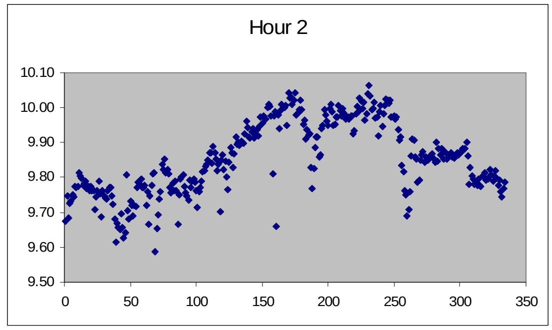

Fig. 1b

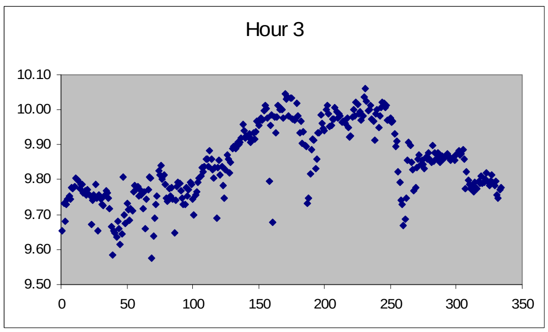

Fig. 1c

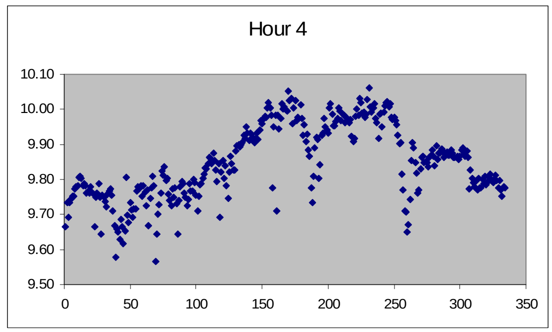

Fig. 1d

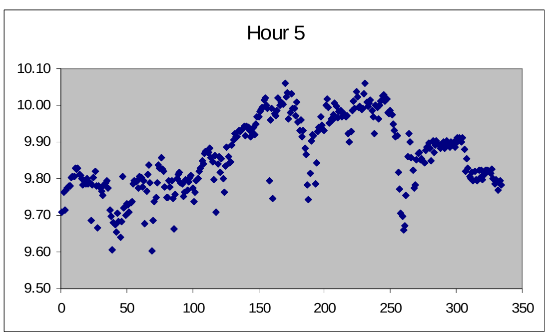

Fig. 1e

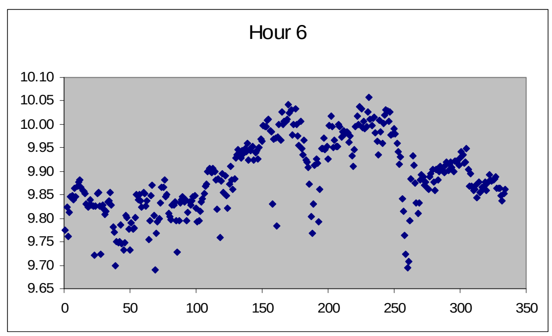

Fig. 1f

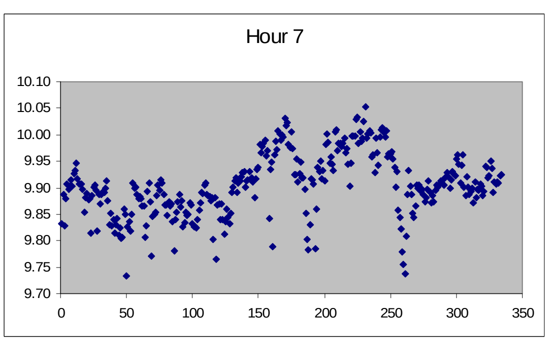

Fig. 1g

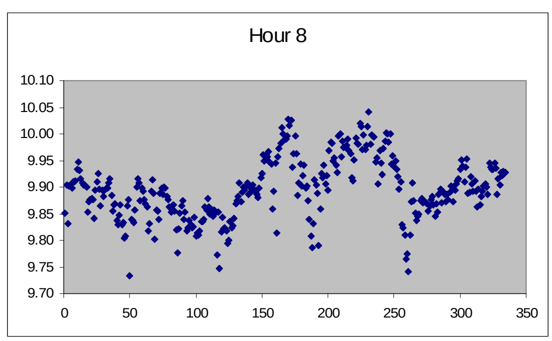

Fig. 1h

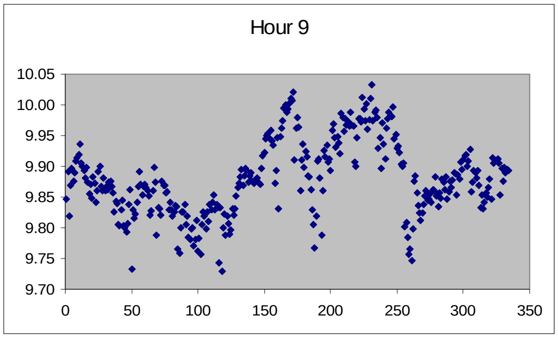

Fig. 1i

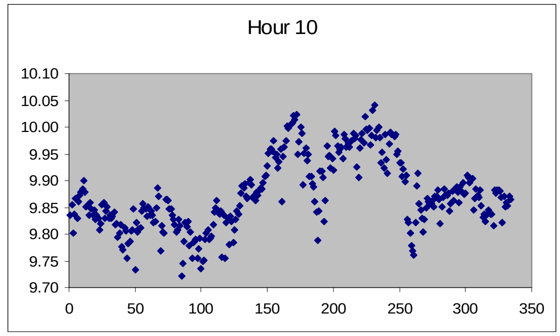

Fig. 1j

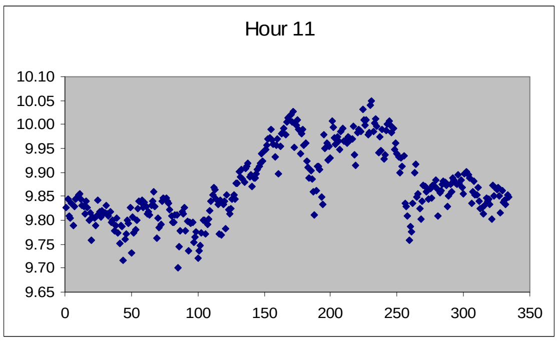

Fig. 1k

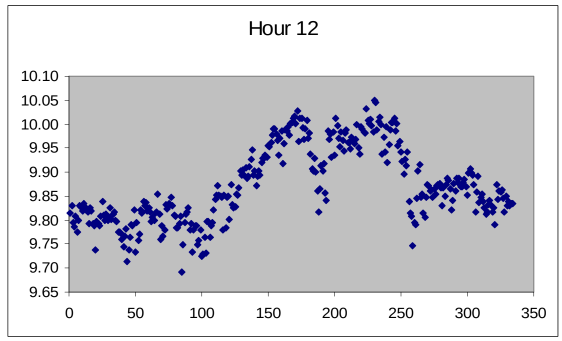

Fig. 1l

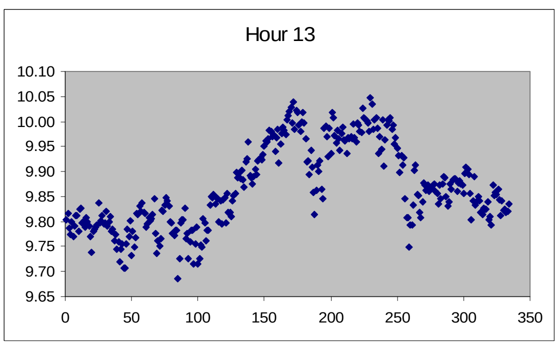

Fig. 1m

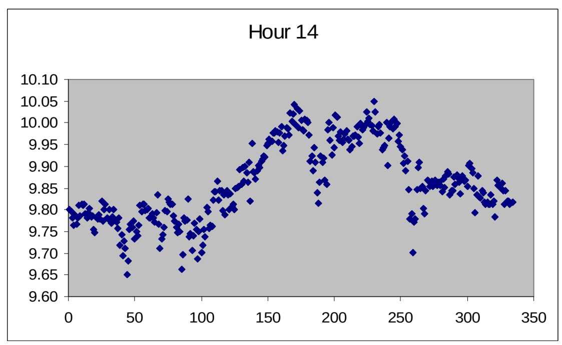

Fig. 1n

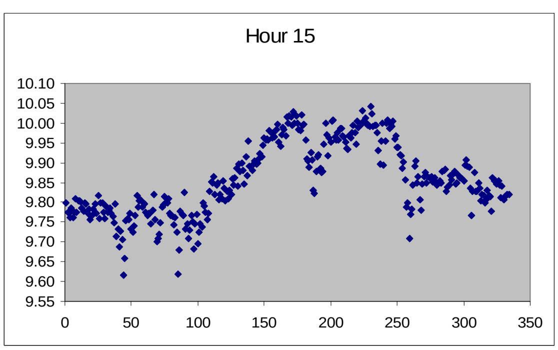

Fig. 1o

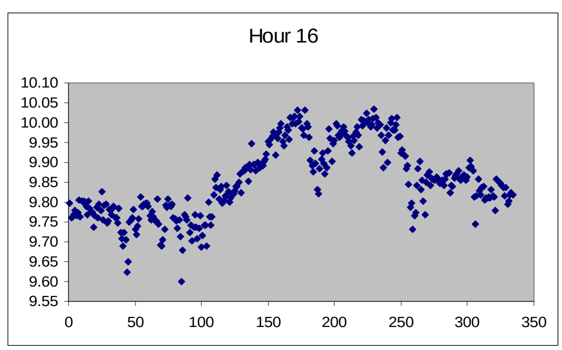

Fig. 1p

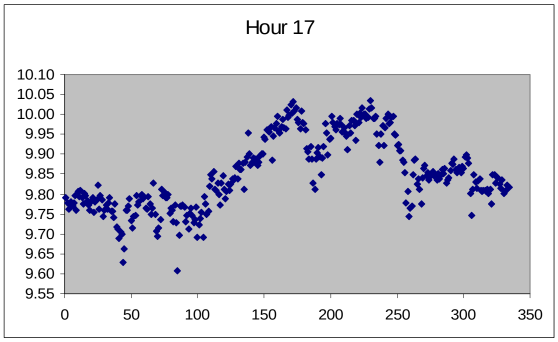

Fig. 1q

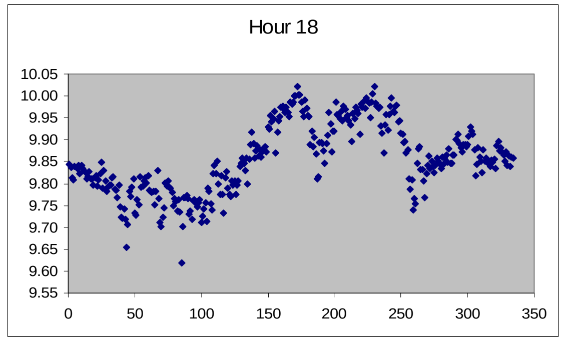

Fig. 1r

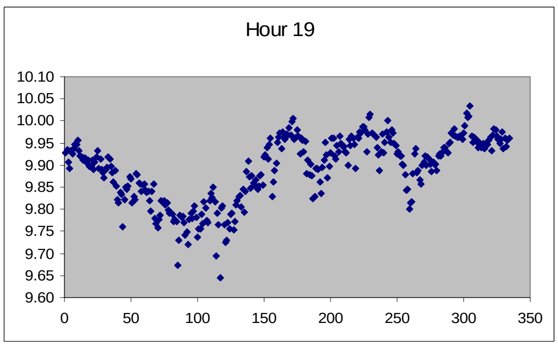

Fig. 1s

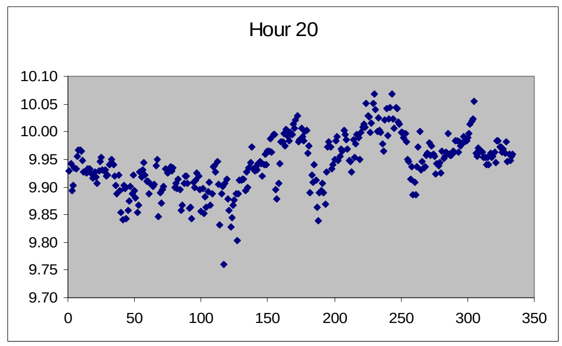

Fig. 1t

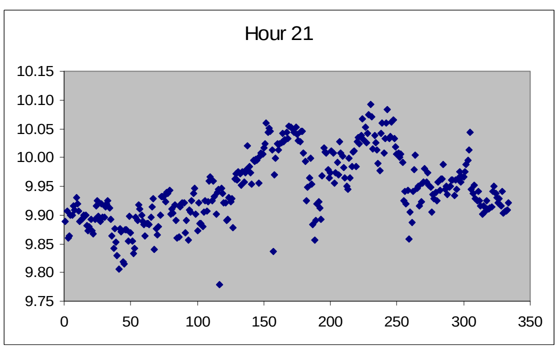

Fig. 1u

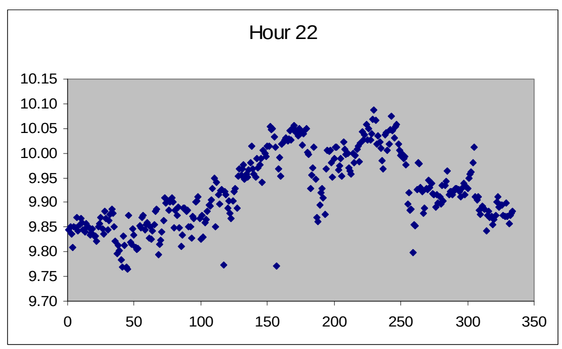

Fig. 1v

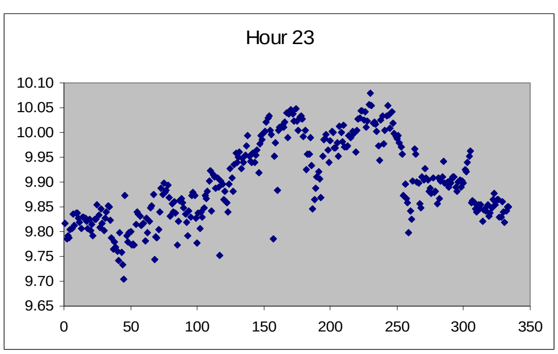

Fig. 1w

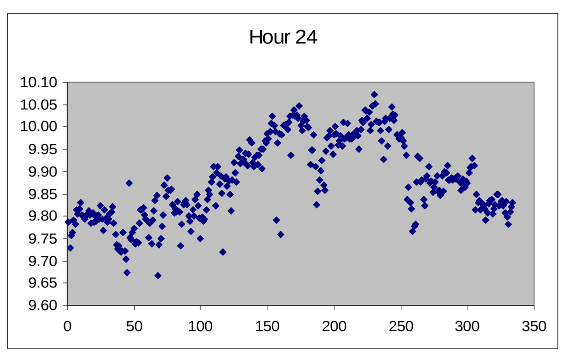

Fig. 1x

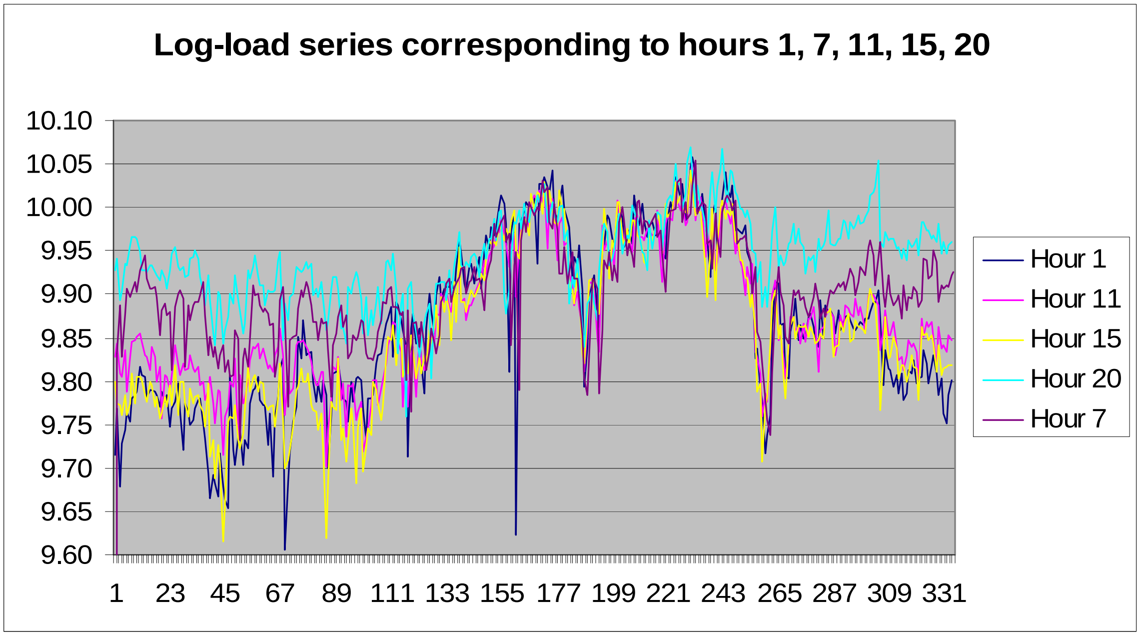
Log-load series corresponding to hours 1, 7, 11, 15, 20
10.10
10.05
10.00
9.95
9.90
9.85
9.80
9.75
9.70
9.65
9.60
1
23
45
67
89
111
133
155
177
199
221
243
265
287
309
331
Hour 1
Hour 11
Hour 15
Hour 20
Hour 7

Fig. 2

Table 1: Model Parameter Estimates for each hour along with adjusted R^2, DW and LB-P values

Hour	ϕ_0	ϕ_1	ϕ_2	β_1	β_2	β_3	β_4	β_5	β_6	γ_1	γ_2	Adj R^2	DW	LB P-value
1	2.349	0.541	0.221	×	×	×	×	×	×	-0.019	0.025	0.810	2.106	0.436
2	2.164	0.582	0.198	×	×	×	×	×	×	-0.018	0.024	0.822	2.073	0.648
3	2.125	0.608	0.176	×	×	×	×	×	×	-0.018	0.023	0.820	2.060	0.764
4	2.067	0.634	0.156	×	×	0.01	×	×	×	-0.018	0.022	0.822	2.047	0.584
5	2.064	0.671	0.120	×	×	×	×	×	×	-0.016	0.020	0.814	2.036	0.535
6	1.839	0.714	0.099	×	×	×	×	×	×	×	0.020	0.784	2.037	0.201
7	1.917	0.661	0.145	×	×	×	×	×	×	×	0.011	0.684	2.049	0.371
8	1.522	0.666	0.180	×	×	×	×	×	×	×	0.008	0.717	2.057	0.508
9	1.348	0.638	0.225	×	×	×	×	×	×	×	0.009	0.760	2.071	0.557
10	1.251	0.680	0.193	0.010	×	×	×	×	×	×	0.012	0.824	2.068	0.270
11	1.205	0.629	0.248	0.013	×	×	×	0.008	×	×	0.014	0.857	2.059	0.267
12	1.489	0.591	0.258	0.015	×	×	×	0.010	×	-0.009	0.014	0.871	2.051	0.292
13	1.457	0.670	0.182	0.014	×	×	×	0.008	×	-0.010	0.014	0.881	2.053	0.502
14	1.437	0.636	0.218	0.014	×	×	×	×	×	-0.010	0.015	0.882	2.055	0.172
15	1.529	0.711	0.133	0.020	0.010	×	×	0.010	×	-0.012	0.016	0.889	2.038	0.261
16	1.474	0.736	0.113	0.023	0.011	×	×	0.011	×	–0.012	0.015	0.893	2.037	0.052
17	1.391	0.741	0.117	0.023	0.011	×	×	0.010	×	–0.011	0.014	0.895	2.038	0.052
18	1.221	0.739	0.137	0.017	0.014	×	×	×	×	–0.010	0.009	0.875	2.029	0.289
19	0.838	0.739	0.175	0.031	0.030	0.016	0.019	0.021	0.016	–0.008	×	0.853	2.055	0.616
20	1.813	0.659	0.160	0.009	0.011	×	×	0.008	×	–0.009	×	0.704	2.036	0.433
21	2.302	0.613	0.156	×	0.008	×	×	×	×	–0.012	0.011	0.757	2.025	0.615
22	2.210	0.592	0.185	×	0.012	×	×	×	×	–0.013	0.017	0.806	2.047	0.440
23	2.101	0.575	0.213	×	0.012	×	×	×	×	–0.013	0.020	0.832	2.069	0.798
24	2.449	0.564	0.188	×	×	×	×	×	×	–0.016	0.025	0.813	2.085	0.433

Table 2: MAPE for the month of December 2005 for our best fitted models, as summarized in Table 1, for each hourly log-load data

	MAPE calculated over all 31 days of December 2005											
Hour	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
MAPE	**0.14%**	**0.15%**	**0.16%**	**0.18%**	**0.18%**	**0.18%**	**0.21%**	0.41%	0.48%	0.35%	0.31%	0.28%
Hour	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
MAPE	0.23%	**0.20%**	**0.19%**	**0.18%**	**0.18%**	**0.17%**	**0.17%**	**0.16%**	**0.14%**	**0.13%**	**0.14%**	**0.16%**
	MAPE calculated over 27 days of December 2005, excluding 22-25 December											
Hour	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
MAPE	**0.14%**	**0.14%**	**0.14%**	**0.14%**	**0.13%**	**0.14%**	**0.17%**	0.19%	0.20%	0.18%	0.17%	0.15%
Hour	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
MAPE	0.14%	**0.16%**	**0.17%**	**0.16%**	**0.16%**	**0.17%**	**0.16%**	**0.14%**	**0.13%**	**0.13%**	**0.14%**	**0.16%**