A Multi-step Approach to Modeling the 24-hour Daily Profilesof Electricity Load using Seasonally-Varying Splines

Jornaz, Abdelmonaem [1], V.A. Samaranayake[1]
[1]Mathematics & Statistics, Missouri S&T, 400 West 12th Street, Rolla, MO 65409

Abstract

A multi-step approach to modeling hourly electricity load and estimating the 24-hour daily profile of electricity use is proposed. The methodology is motivated by the time-varying spline model introduced by Harvey and Koopman in 1993, but includes several important modifications. One is the multi-step approach to modeling daily load and then using the residuals to model the hourly data. The proposed method also use variables based on temperature to modify the model parameters, and in addition express these parameters as a time series, thus producing a model that is more dynamic and attuned to temperature related changes. Data from the Atlantic Electric Zone of the PJM market is used to illustrate the proposed method.

KeyWords: Electricity Load Forecasting time-varying splines, Daily Profile Load Curve

1. Introduction

Electricity demand is driven by many factors, including economic conditions and weather. Furthermore, the electricity demand varies with time, with different hours of the day and different days of the week having an effect on the load. Thus, modeling electricity demand can be a challenging task, but reliable models that can be used to forecast demand is of practical importance to those in the energy and public and private utility sectors. Therefore this filed has generated the interest of many researchers.

There is a long history of research work aimed at developing hourly electricity load models. For classical approaches the reader is referred to Bunn and Farmer (1985) which presents approaches that were used to forecast the short-term load. Another important reference is Alfares and Nazeeruddin (2002),which classified the various approaches into nine classes. The latter authors also commented that while the time series approach is widely used, hybrid approaches, which combine several techniques, have become more common. Taylor and McSharry (2007) conducted an empirical comparisonof some short-term forecasting methods using teninterday electricity demand time seriesfrom ten European countries; for more information see Weron (2007). Two key publications in this area are by Harvey and Koopman (1993) who proposed time-varying splines to model intra-weekly load and Cho et al. (2013) who proposed a hybrid approach using generalized additive model and curve linear regression to model weekly and daily electricity load. In this paper we adopt Koopman's approach to model an empirical data set, but incorporate several important modifications.

2. Data Description and the Analysis

The data used in this study consist of load data was obtained from the Pennsylvania-New Jersey-Maryland (PJM); the data cover Atlantic Electric zone (AE) in southern New Jersey. This data includes hourly observations over 20 years (1993-2012). The economic data was obtained from Federal Reserve Bank of ST. Louis. Moreover, the weather data was obtained from National Oceanic Atmospheric Administration (NOAA). Figure 1 shows the actual hourly load over 20 years.

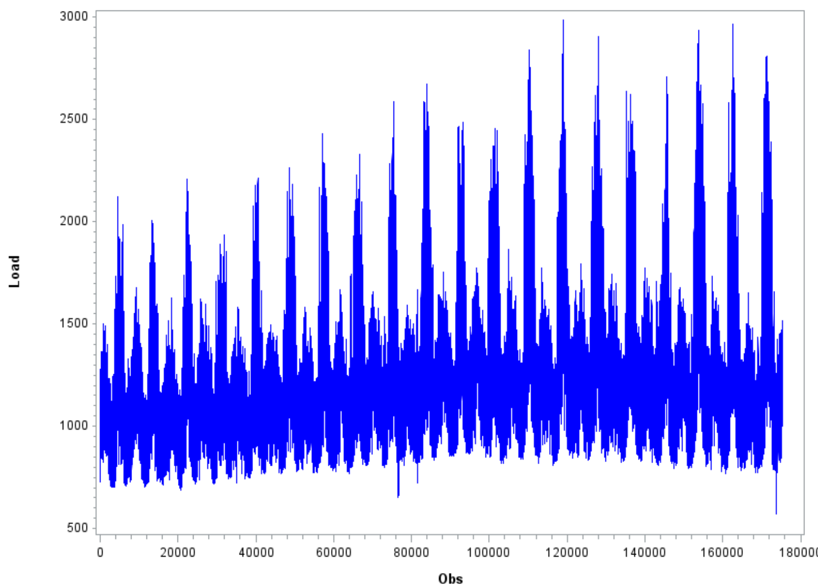

Figure 1. The actual hourly load over 20 years.

The time series of electricity demand Y_t was considered a composite of structural components consisting of a long term trend τ_t, a seasonal component S_t, a weekly cycle W_t, a 24-hour demand function $f(t)$, and an irregular stochastic component u_t as given in the model below:

$$Y_t = \tau_t + S_t + W_t + f(t) + u_t$$

We predict each component separately as described below.

2.1 Predicting long-term trend

The first step was modeling the hourly average electricity demand per year by using classical regression analysis with selected economic used as independent variables. The economic variables were selected based on stepwise regression. The final model explains 98.5% of variation in the annual load with mean RMSE 9.7. Moreover, we did not see any serious multicollinearity among the independent variables.

The regression model for the annual data is:

$$\overline{Y}_y = 270.95 – 8.15 \text{ RV} – 0.0001 \text{ ED} + 1.96 \text{ GOV} – 47.37 \text{ HV} + 5.82 \text{ IPRO}$$

where, $\overline{Y}_y$ is the hourly average load for year *y*, RV is the Rental Vacancy Rate in New Jersey (NJ), ED is the Durable Manufacturing Earnings in NJ, GOV is the Government Employment in NJ, HV is the Home Vacancy Rate in NJ, and IPRO is the Industrial Production Index of the US; all independent variables measured for the year *y*. The observed and values predicted using the above model are given in Figure 2.

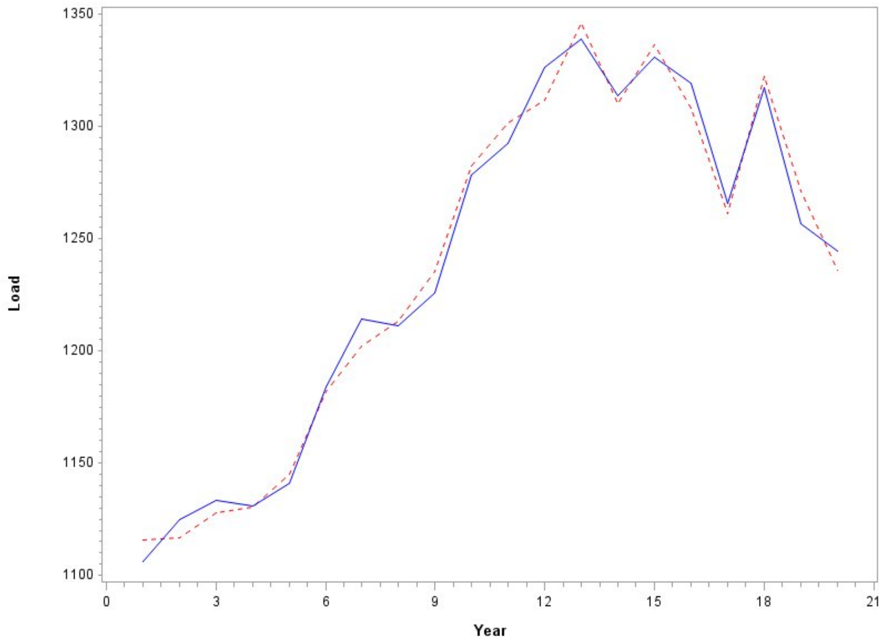

Figure 2. The annually average of hourly load (bluesolid) and the predicted (red dashed).

2.2 Predicting seasonal variation

The trend estimates for each year were transposed onto a weekly series, and a 52-week moving average was applied to this series. This smoothed trend was removed from weekly electricity demand data (averaged over the hours to reflect average per hour demand) by subtracting this resulting smoothed weekly series.The de-trended weekly time series (we shall denote it by $\tilde{W}_w$, where the subscript *w* denote the week) was modeled by using a subset ARMA model with added independent variables. The independent variables used are a holiday dummy h_w (equal to one if the week *w* had a holiday), minimum weekly temperature mT_w, $T_w(<65) = I_{[T<65]}(T_w)$ where T_w is the average weekly temperature, $T_w(>70) = I_{[T_w>70^0 F]}(T_w)$, $T_w(>80) = I_{[T_w>80^0 F]}(T_w)$, and $H_w(<50)$ the total number of hours below $50^0 F$.

The estimated ARMA model for the de-trended weekly data is:

$\tilde{W}_w$ = -113.64 + 0.71 L$\tilde{W}_{w-1}$ – 0.16 $\tilde{W}_{w-13}$ + 0.07 $\tilde{W}_{w-51}$ + 0.8 $\tilde{W}_{w-52}$ – 0.58 $\tilde{W}_{w-53}$ – 0.38 Z_{w-1} – 0.76 Z_{w-52} + 0.35 Z_{w-53} – 12.98 h_w + 1.34 mT_w – 1.57 $T_w(<65)$ +2.31 $T_w(>70)$ +3.17 $T_w(>80)$ + 0.82 $H_w(<50)$,

where the Z_w terms define the moving average portion of the ARMA process. The observed and predicted values for two select weeks are displayed in Figure 3.

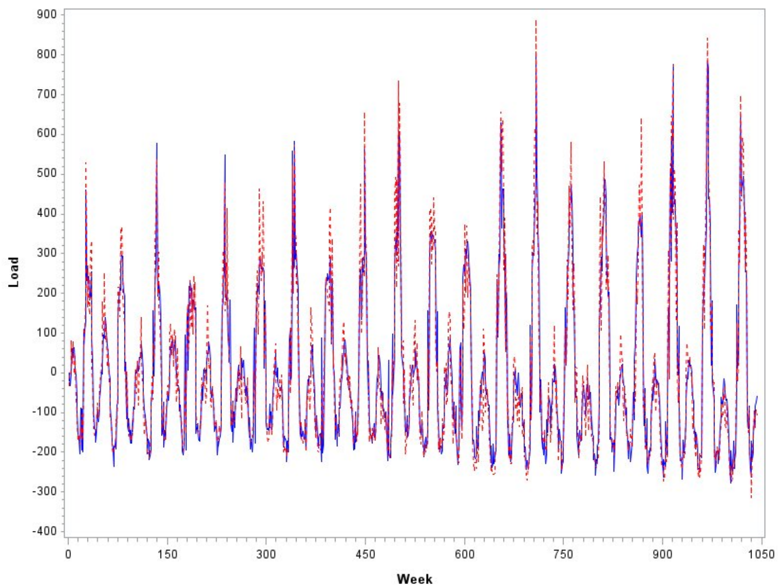

Figure 3. The weekly average of hourly load (blue solid) and the predicted (red dashed).

2.3 Predicting hourly load using cubic splines

After removing trend and seasonality from the hourly time series, a new de-trended and de-seasonalized time series was obtained. Note there in this notation *t* denotes hour and *h* = *h*(*t*) the hour of the day so that $h \in \{1, 2, \ldots, 24\}$. The daily load profile $f(h)$ was modeled by using cubic splines with different spline estimates obtained for each season, weekdays, and weekends. That is, we pooled all the hourly $\tilde{L}_t$ values for all weekdays for a given season with hour of the day *h* as the primary independent variable. The temperature (*T*) and its interactions with the spline coefficients were also fitted. Residuals of the spline model were then modeled using a factored ARMA model across the 20 year period. Given below are four of these models, namely winter weekend, winter weekday, summer weekend, and summer weekday. Results for other seasons are available from the first author upon request.

The spline model for weekends during the winter is:

$\hat{f}(h)$ = 151.52 – 74.34h + 10.5 h^2 – 0.24h^3 + 2.88$(h – 5)^2$ – 20.17$(h - 9)^2$ – 52.59$(h – 14)^2$ – 95.88$(h – 17)^2$ – 0.88$(h - 5)^3$ + 4.88$(h - 9)^3$ + 6.39$(h – 14)^3$ – 8.68$(h – 17)^3$ – 6.17 T– 0.08 $T*h$ + 0.7 $T*(h – 14)^2$ + 0.01 $T*(h - 5)^3$ – 0.05 $T*(h - 9)^3$+ 0.05 $T*(h – 17)^3$.

The spline model for weekdays during the winter is:

$\hat{f}(h)$ = 140.26 – 80.32h + 15.12h^2 – 0.74h^3 + 67.32$(h – 5)^2$ – 68.69$(h - 8)^2$ – 62.63$(h – 15)^2$ – 153.72$(h – 17)^2$ – 15.81$(h - 5)^3$ + 17.65$(h - 8)^3$ + 26.37$(h – 15)^3$ – 25.88$(h – 17)^3$ – 6.43T + 1.31$T*(h – 15)^2$ + 0.01$T*(h - 5)^3$ – 0.02$T*(h - 8)^3$ – 0.32 $T*(h – 15)^3$ + 0.34 $T*(h – 17)^3$.

The spline model for weekends during the summer is:

$\hat{f}(h)$ = –1164.01.26 + 4.7h + 14.18h^2 – 0.5h^3 – 12.05$(h – 4.5)^2$ + 5.52$(h - 9)^2$ + 7.66$(h – 16)^2$ – 33.82$(h – 19)^2$ – 2.87$(h – 4.5)^3$ + 8.26$(h - 9)^3$ – 8.76$(h – 16)^3$ – 12.05$(h – 19)^3$ + 14.79T– 1.7$T*h$ – 1.62$T*(h - 9)^2$+ 1.21$T*(h – 19)^2$ + 0.11$T*(h – 4.5)^3$ – 0.14 $T*(h - 9)^3$.

The spline model for weekdays during the summer is:

$\hat{f}(h)$ = –1304 + 131.24h + 12.35h^2 – 3.68h^3 + 98.06$(h – 4.5)^2$ + 82.6$(h - 9)^2$ + 2.94$(h – 16)^2$ – 31.51$(h – 19)^2$ – 9.01$(h – 4.5)^3$ + 16.45$(h - 9)^3$ – 5.86$(h – 16)^3$ – 1.94$(h – 19)^3$ + 16.13T – 3.29$T*h$ – 1.19$T*(h - 9)^2$ + 1.01$T*(h – 19)^2$ + 0.13$T*(h – 4.5)^3$ – 0.2 $T*(h - 9)^3$.

Observe that the positions of the knots are different for each season and type of day. This is because the shape of the load profile is quite different for those season by day combinations. In addition, terms that are statistically insignificant were eliminated from the models. This approach is different from that proposed by Koopman (1993), who suggests fitting the same set of splines but allow the spline coefficients to vary stochastically as a set of individual time series. While that is a viable approach, we assumed that most of the stochasticity of spline coefficients is due to seasonal variations in electricity use patterns that are not attributable seasonal weather plus those that are due to weather variables (e.g. temperature). As such we propose our approach as a viable alternative. The hourly average of the observed $\tilde{L}_t$ values for a given season and type of day and the predicted values $\hat{f}(h)$ are displayed in Figure 4. As seen from the figure, the estimated splines do a very good job in approximating the observed data.

Once the splines are estimated, the predicted values can be subtracted from the $\tilde{L}_t$ values and the residuals estimated as a pure stationary time series. We have done this but these results are not reported here because the estimation of the splines were our main objective. If one is to predict the load use profile for a future week, however, the predicted values from the residual model as well as the spline predictions, together with the predicted values from the yearly and the weekly models can be combined to produce a composite prediction.

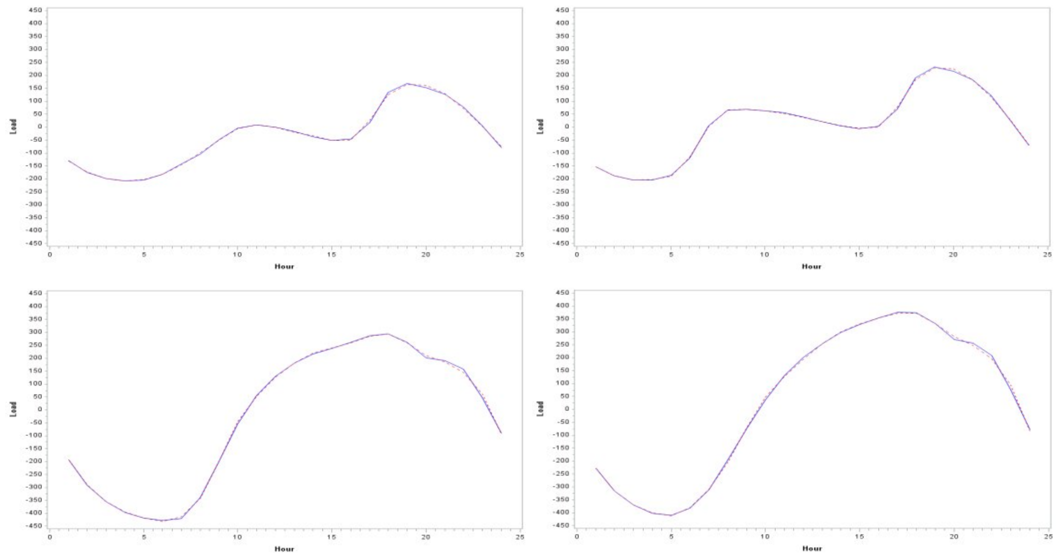

Figure 4. The average of hourly load (blue solid) and the spline predicted (red dashed),
top left: weekends in the winter, top right: weekdays in the winter.
bottom left: weekends in the summer, bottom right: weekdays in the summer.

References

Alfares, H., and Nazeeruddin, M. (2002), "Electric Load Forecasting: Literature Survey and Classification of Methods," *International Journal of Systems Science*, 33, 23-34.

Bunn, D., and Farmer, E. (1985), *Comparative Models for Electrical Load Forecasting*, New York: John Wiley & Sons.

Cho, H., Goude, Y., Brossat, X., and Yao, Q. (2013), "Modeling and Forecasting Daily Electricity Load Curves: A Hybrid Approach," *Journal of the American Statistical Association*, 108, 7-21.

Harvey, A., and Koopman, S. (1993), "Forecasting Hourly Electricity Demand Using Time-Varying Splines," *Journal of the American Statistical Association*, 88, 1228-1236.

Taylor, J. W., and McSharry, P. E. (2007), "Short-term Load Forecasting Methods: An Evaluation Based on European Data," *Power Systems, IEEE Transactions on*, 22, 2213-2219.

Vilar, J., Cao, R., and Aneiros, G. (2012), "Forecasting next-day electricity demand and price using nonparametric functional methods,"*International Journal of Electrical Power and Energy Systems*, 39, 48-55.

Weron, R. (2006), *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, UK: John Wiley & Sons.