Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 24 (2008) 603 – 615
www.elsevier.com/locate/ijforecast

A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting

Luiz Felipe Amaral [a,*], Reinaldo Castro Souza [a], Maxwell Stevenson [b]

[a] *Electrical Engineering Department, DEE, PUC-Rio de Janeiro, Brazil*
[b] *Faculty of Economics and Business, The University of Sydney, Australia*

Abstract

This paper compares the short-term load performance of several forecasting models, including a new class of nonlinear models known as smooth transition periodic autoregressive (STPAR) models. A model building procedure is developed for the STPAR model, along with a linearity test against smooth transition periodic autoregressive behaviour. The predictive ability of the STPAR model is evaluated against alternative load forecasting models using load data from the Australian electricity market.

Keywords: Time series; Periodic and autoregressive models; STAR model; Load forecast

1. Introduction

Much of the early literature on electricity load forecasting models has been summarized by Bunn and Farmer (1985) and the references therein. More recent developments have typically focused on highly structured, larger scale models. The associated computationally intensive fitting, calibration and forecasting procedures have been made possible by the exponential growth in computing facilities over the last two decades. A key consideration in all these models is how best to capture the range of time scales, from half-hourly through to years, that are present in this type of data. Two broad classes of conceptual models have emerged which address these time scale issues in different ways.

One such class includes single-equation models, while multi-equation or vector models form the other. The single-equation models are specified with components that include trend, daily periodicity and exogenous variables, among others. These components are all defined on a half-hourly or hourly time scale, depending on the series being modeled. Examples of single-equation models are given by Smith (2000); Nowicka-Zagrajek and Weron (2002); Huang (2003); Taylor (2003); Taylor and Buizza (2003); Huang, Huang, and Wang (2005); and Taylor, de Menezes, and McSharry (2006).

Multi-equation models organize the data into daily (or weekly) vectors of half-hourly or hourly load, with the model specified over a daily or weekly time scale,

* Corresponding author.
E-mail address: luiz.amaral@vale.com (L.F. Amaral).

0169-2070/$ - see front matter © 2008 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2008.08.006

or both. For example, daily profiles of 48 half-hourly loads can be regarded as a vector time series with a daily time scale. For this class of model, intra-day half-hourly variation is accounted for using a cross-sectional model for each vector, and inter-day variation is modelled on the daily time scale. Articles using this conceptual approach include Fiebig, Bartels, and Aigner (1991); Peirson and Henley (1994); Ramanathan, Engle, Granger, Vahid-Arahi, and Brace (1997); Cottet and Smith (2003); Soares and Medeiros (2005); and Soares and Souza (2006).

In principle, all vector models can be reformulated as single-equation models and vice versa. However, the different conceptual frameworks discussed above inevitably lead to different stochastic models being adopted for each of the various model formulations.

Artificial Neural Network (ANN) models are an alternative approach for load forecast modelling, and, in terms of accuracy, compete directly with the single-equation and vector models cited above. However, the ANN models are complex and difficult to understand, and are often over-fitted. Indeed, their structure is sufficiently opaque that it is not clear why they should forecast as well as they do. As a result, the literature is still undecided as to their practical utility for electricity load forecasting. Articles that deal with ANN models include Khotanzad, Afkhami-Rohami, and Maratukulam (1998); Alves da Silva and Moulin (2000); Darbellay and Slama (2000); Metaxiotis, Kagiannas, Askounis, and Psarras (2003); Reis and Alves da Silva (2005); and Hippert, Bunn, and Souza (2005).

The main objective of this paper is to compare the short-term (up to seven days ahead or 336 half hours) load forecasting performance of several forecasting models, including the new single-equation STPAR class of nonlinear models. In order to evaluate the forecasting performance, an electricity load series from the Australian state of New South Wales (NSW) will be used.

The rest of the article is organized as follows. Section 2 presents the load and temperature data used in our analysis. Section 3 describes the nonlinear STPAR model, along with a modeling strategy for implementing the specification, estimation and forecasting analysis using this model. Alternative models that will serve as benchmarks in the evaluation process are also discussed in this section. After all models are estimated and forecasts produced from them, Section 4 reports the results from the forecasting evaluation. Concluding remarks are contained in Section 5.

2. NSW electricity load and temperature series

In this paper we use load and temperature observations from the Australian state of NSW for load model estimation and forecasting.[1] They are composed of half-hourly measurements of load and temperature from July 1, 2001 to June 30, 2005, a total of 4 years of data or 70128 points. The load data is aggregated across all retailers in NSW to become the NSW State Load. The observations from July 1, 2001 to June 30, 2004 are used as the estimation sample (in-sample), while the last year from July 1, 2004 to June 30, 2005 is used for the out-of-sample analysis. Fig. 1 graphically depicts twelve months of load observations from July 1, 2003 to June 30, 2004, while Fig. 2 illustrates two weeks of half-hourly loads from July 1, 2003 to July 15 2003.

An important component of the non-linear model that is the focus of this study is the temperature data set. It was collected at Bankstown Airport in the suburbs of Sydney, the largest city in NSW.

It has been suggested in the literature that electricity load time series may contain a trend component. This could take the form of either a deterministic or a stochastic trend. Over the time period considered for our analysis, the NSW State Load was clearly a mean-reverting series without a discernible deterministic trend. Further, after testing the null hypothesis of a stochastic trend (unit root) in the estimation sample using the Phillips-Perron Test[2] (PP), the hypothesis that the NSW State Load series contained a unit root was rejected at the 99% confidence level.[3] It followed that trend was not considered in the modelling process.

A discussion of the load series would be incomplete if no mention was made of how to consider the

[1] Integral Energy, an electricity retailer operating in the state of NSW, generously supplied the load and temperature data.

[2] Phillips and Perron (1988).

[3] However, it is well known that the power of the PP test is reduced under some circumstances like heterogeneity, a strong seasonal cycle and other common factors of the underlying series.

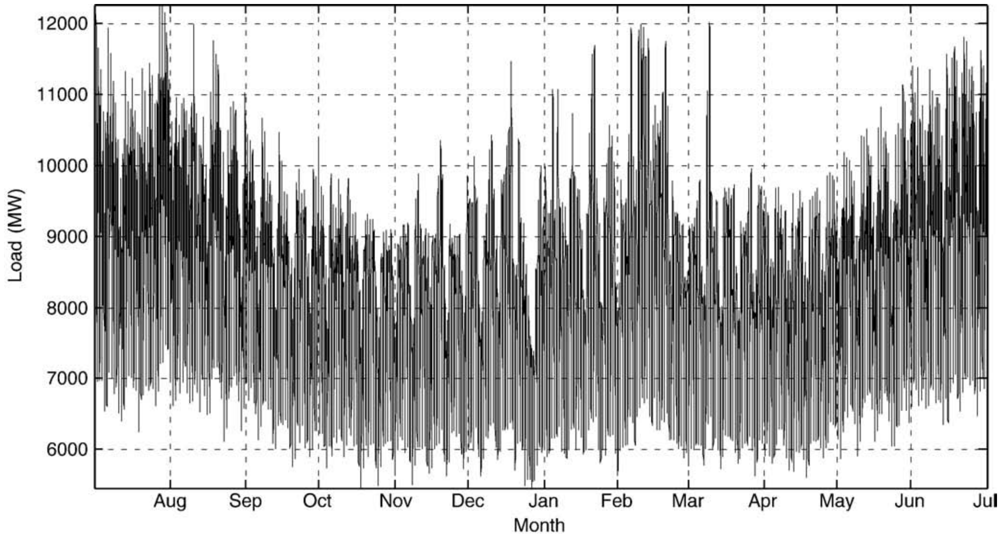

Fig. 1. Half-hourly loads in the Australian state of NSW for twelve months from Tuesday, July 1 2003 to Wednesday, June 30, 2004.

treatment of special days such as holidays. As previously stated, the main objective of this paper is to compare modeling approaches for forecasting load series. Taylor et al. (2006) argue that when a comparison of methodologies is the objective, the treatment of special days is likely to be meaningless because of the markedly different pattern of load on these days, with the inevitability that univariate methods will generate poor forecasts for these days. Therefore, when comparing the forecasting ability of different models, of which some are univariate, a logical approach is to not consider data from these days, and to treat them separately for forecasting purposes. From another point of view, if the

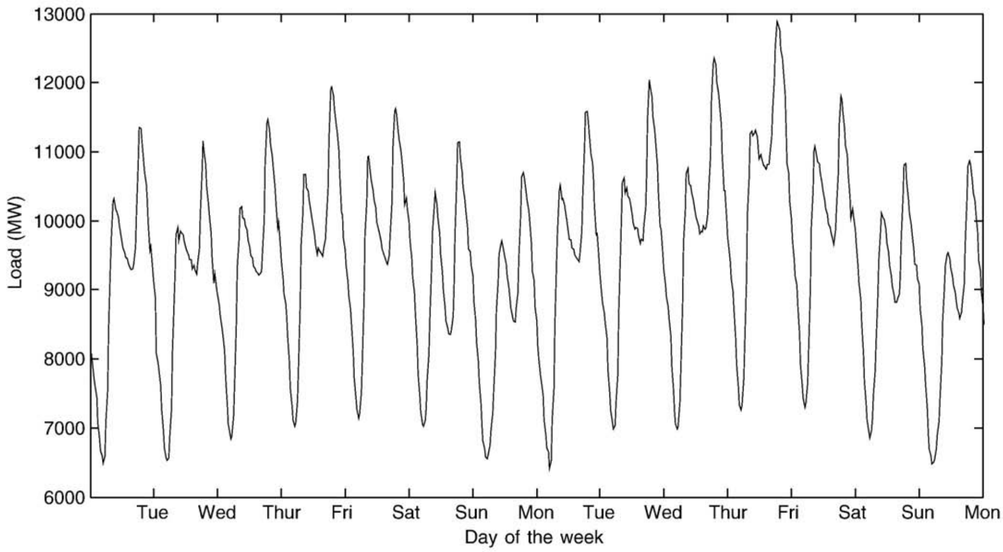

Fig. 2. Half-hourly loads in the Australian state of NSW for two weeks from Tuesday, July 1, 2003 to Tuesday, July 15, 2003.

purpose for which the model was developed is to assist in the hedging of electricity prices, special days are not likely to be a material issue. Loads are invariably smaller on special days than on normal days, and as such, special days do not drive hedging policy or practice. Based on the above considerations, the special day treatment adopted in this paper was to substitute for them the mean of the data from the same day of the week and the same half-hour, one week before and one week after.

3. Models for forecasting the electricity load

3.1. Linear periodic models

One class of models that has been used to model seasonal time series is the periodic autoregressive model (PAR). Franses and Paap (2004) point out that prior to 1996, not many academics and practitioners considered periodic models for modelling seasonal time series. However, in recent times the use of this type of model has increased notably. Studies using this modelling approach include Lund and Basawa (2000), Basawa and Lund (2001) and Ghysels and Osborn (2001). Periodic models have shown their usefulness in forecasting the intraday net imbalance volume (Taylor, 2006), as well as daily electricity spot prices (Koopman, Ooms, & Carnero, 2007).

An appealing characteristic of these models is that the autocorrelation at a particular lag of one half-hour varies across the week. This behaviour is not captured by a seasonal ARMA model. Taylor (2006) suggested the following periodic model:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_i y_{t-p} + \varepsilon_t \\ \beta_i &= \alpha_i + \sum_{j=1}^{h} \left(\lambda_{ij} \sin\left(2j\pi \frac{D(t)}{48}\right) + \tau_{ij} \cos\left(2j\pi \frac{D(t)}{48}\right) \right. \\ &\quad \left. + \varsigma_{ij} \sin\left(2j\pi \frac{W(t)}{336}\right) + \eta_{ij} \cos\left(2j\pi \frac{W(t)}{336}\right) \right), \end{aligned} \quad (1)$$

where $i = 0, 1, \ldots, p$, $j = 1, \ldots, h$, and $D(t)$ and $W(t)$ are repeating step functions that number the periodicity. For half-hourly load data, $D(t) = 1, 2, \ldots, 48$ and $W(t) = 1, 2, \ldots, 336$, with h being the number of harmonics. It follows that a Fourier form would represent the daily and weekly periodicity. This notation is used in order to overcome the problem of a highly parameterized model where the common notation of a periodic model suggests the need for the estimation of a parameter for each period of the daily and weekly periodicities.

Weather variables such as temperature, humidity, cloudiness, and other climatic variables are important in the modelling of demand series. As it stands, model (1) does not incorporate the information contained in these variables that is relevant to the modelling process. Temperature is a variable that is highly positively correlated with load. One way to include it is as an explanatory variable in the first linear autoregressive equation of the periodic model given by Eq. (1). However, the relationship between temperature and load is asymmetric. For example, if the variation in temperature from one half-hour to another is + 6°C, the effect on demand might be an increment of say 1000MW. If there is a variation in temperature of + 3°C, the effect on demand will not be 500MW as might be expected. This highlights the important non-linear relationship between load demand and temperature. Any time series modelling of such asymmetric behaviour is best represented by a non-linear model, not by typical linear time series models, with or without temperature incorporated. In this paper we propose a nonlinear model that can explicitly capture the periodicity of the load and the non-linear relationship between it and temperature.

3.2. Non-linear periodic models

In recent years, applied research using nonlinear time series techniques has been comprehensive. The several new models and methodologies that have proved popular include some that are based on multiple regimes.[4] The smooth transition autoregressive model (STAR) is one such model; it was first introduced by Chan and Tong (1986), and further developed by Teräsvirta (1994). It incorporates a smooth transition between two or more regimes. Each regime is modelled as an autoregressive process of order p with a transition function that determines the data composition of each regime. In the new formulation proposed here, we will incorporate periodic behaviour into a STAR model

[4] See van Dijk, Teräsvirta, & Franses (2002) and Stevenson and Peat (2001), to mention only two applications.

framework. The model created is referred to as a Smooth Transition Periodic Autoregressive (STPAR) model. It is given by

$$
\begin{aligned}
y_t &= \beta_{1,0} + \beta_{1,1} y_{t-1} + \ldots + \beta_{1,i} y_{t-i} \\
&\quad + \left\{ \left(\beta_{2,0} + \beta_{2,1} y_{t-1} + \ldots + \beta_{2,i} y_{t-i} \right) \right. \\
&\quad \left. \times F(\gamma(s_{t-d} - c)) \right\} + \varepsilon_t \\
\beta_{R,i} &= \alpha_{k,i} + \sum_{j=1}^{h} \lambda_{R,ij} \sin\left(2j\pi \left(\frac{D(t)}{48} \right) \right) + \tau_{R,ij} \cos\left(2j\pi \left(\frac{D(t)}{48} \right) \right) \\
&\quad + \zeta_{R,ij} \sin\left(2j\pi \left(\frac{W(t)}{336} \right) \right) + \varsigma_{R,ij} \cos\left(2j\pi \left(\frac{W(t)}{336} \right) \right), \qquad (2)
\end{aligned}
$$

where $i = 0, 1, \ldots, p$ and $j = 1, \ldots, h$. The number of regimes is R, where $R = 1, 2$, and the random error term is $\varepsilon_t \sim NID(0, \sigma^2)$. $F(\cdot)$ is the logistic function defined by:

$$
F(\gamma(s_t - c)) = (1 + \exp\{-\gamma(s_{t-d} - c)\})^{-1}, \quad \gamma > 0.
$$

The parameter γ determines the smoothness of the change in the value of the logistic function, and is therefore a measure of the smoothness of the transition from one regime to the other. c is a location parameter embedded in the transition function, and only indicates the location of the transition. In addition, s_{t-d} is a transition variable that can be represented by either a lag value of the dependent variable y_{t-d}, or another variable that is highly correlated with the electricity load. Temperature is such a variable, and will take the role of the transition variable in our empirical analysis. The delay parameter d of the transition variable can take values in the range $1 \leq d \leq p$.

When the value of γ is very large, the logistic function $F(\cdot)$ becomes an indicator function. In this case, model (2) becomes a periodic version of the Self-Exciting Threshold Autoregressive (SETAR) model with two regimes.[5]

3.2.1. A STPAR modelling strategy

A modelling procedure for the logistic STPAR model follows the steps advocated by Teräsvirta (1994) for his LSTAR model. Appropriate modifications are required to include the periodic component in the specification. The steps of the proposed STPAR modelling strategy are as follows:

(i) Specify a linear autoregressive model of order p using a model selection criterion such as the AIC (Akaike, 1974) or SBIC (Schwarz, 1978);
(ii) Select the Fourier forms for each lag selected in (i) by analysing the auto correlation function for every half-hour across the week;
(iii) Test the null hypothesis of linearity against STPAR nonlinearity. If linearity is rejected, select the appropriate transition variable s_t and the number h;
(iv) Estimate the parameters in the selected STPAR model using nonlinear least squares, and select the correct value of h using a selection criterion such as AIC or SBIC;
(v) Evaluate the model;
(vi) Finally, obtain the forecasts via Monte Carlo, or use the model for descriptive purposes.

3.2.2. A linearity test

After steps (i) and (ii) are completed, the null hypothesis of linearity against the alternative of non-linearity embedded in the STPAR model is tested. The LM-type linearity test used here is similar to that described by Teräsvirta (1994). It is based on expanding the transition function in (2) into a third order Taylor series, merging terms, and further parameterization. The final auxiliary regression forms the basis for the tests. Because the Taylor expansion is of order zero under the null hypothesis of linearity, the remainder does not affect the asymptotic distribution theory.

Testing for linearity is important for two reasons. Firstly, the non-linear model should only be used if linearity is rejected under the null hypothesis, otherwise a linear model would be suitable for modeling the underlying time series. Secondly, the linearity test allows for the selection of the transition variable, s_t, its delay, d, and the number of harmonics, h, in the Fourier form. The lagged transition variable, s_{t-d}, and the number of harmonics, h, are chosen by running the linearity test and minimizing the p value of the test.[6]

[5] See Tong (1983, 1990) for more details about the SETAR model.

[6] See Teräsvirta (1994) for the motivation behind this selection rule.

3.2.3. STPAR model estimation

Once the variables, the transition variable and h have been selected, the next stage in the modeling cycle involves the estimation of the parameters of a STPAR model.

The parameters in Eq. (2) can be estimated by conditional maximum likelihood or non-linear least squares. When $\varepsilon_t \sim \mathrm{NID}(0,\sigma^2)$, the two methods are coincident. In this case, the vector of parameters ψ can be estimated as:

$$\begin{aligned}\hat{\psi} &= \underset{\psi}{\operatorname{argmin}}\, Q_T(\psi)\\ &= \underset{\psi}{\operatorname{argmin}} \sum_{t=1}^{T} (y_t - G(w_t, s_t; \psi))^2,\end{aligned} \tag{3}$$

where $w_t = [1, y_{t-1}, \ldots y_{t-\mathrm{p}}]'$, and $G(\cdot)$ is the function on the right-hand side of Eq. (2).

One important issue in this type of model is the selection of starting values for the optimization routine. Estimation has proved to be very sensitive to the starting values. In this article we used a two-dimensional grid search over values of γ and c, following van Dijk et al. (2002).[7]

After estimating the parameters, the model needs to be evaluated to verify that it was correctly specified. A Lagrange Multiplier (LM) test for the hypothesis of no error autocorrelation, as well as a LM-type test for the hypothesis of no remaining nonlinearity, was implemented. The construction of these tests was based on the work of Eitrheim and Teräsvirta (1996).

3.2.4. Forecasting with the STPAR model

Forecasting with a periodic autoregressive model essentially involves the same procedure as a non-periodic autoregressive model. The difference lies in the incorporation of the Fourier-based coefficients into the forecast function. It follows that the procedure for forecasting the autoregressive STPAR model will closely follow the procedure followed to produce forecasts from the autoregressive models and the better-known STAR models.

Further, a Monte Carlo simulation was applied by using the error distribution from the estimated model to sample the innovations. The advantage of a simulation approach, rather than a point forecast, is that a number of point forecasts are generated at each period across the forecast horizon. This provides a density forecast at each tick of time in the prediction period, and facilitates the calculation of an expected value at each tick, along with higher moment measures. Following these procedures does not necessarily account for sampling uncertainty, which therefore results in somewhat narrow confidence intervals (see Lundbergh & Teräsvirta, 2000).

The k-step ahead forecast function is given by Eq. (4) below:

$$y^f_{t+k,i|t} = 1/N \sum_{i=1}^{N} G\left(w^f_{t+k,i}, s^f_{t-d+k,i}; \psi\right) \tag{4}$$

Each of the N values of $\varepsilon_{t+(k-1)}$ in $w^f_{t+k,i}$ is drawn independently from an error distribution of the estimated model, with constant mean and variance. By the weak law of large numbers, the expected forecast is asymptotically unbiased for very large values of N.

The N forecasts of temperature (the transition variable, $s^f_{t-d+k,i}$) used as input in Eq. (4) are point forecasts from the Svec and Stevenson (2007) temperature-forecasting model, the specification of which is given in Eq. (5) below:

$$\begin{aligned} T_t &= \beta_0 + \beta_1 t + S_t + D_t + C_t + \sigma\varepsilon_t\\ S_t &= \sum_{i=1}^{8}\left(A_i\cos\left(2\pi i\frac{d(t)}{52560}\right) + B_i\sin\left(2\pi i\frac{d(t)}{52560}\right)\right)\\ D_t &= \sum_{j=1}^{2}\left(\Gamma_j\cos\left(2\pi j\frac{\varpi(t)}{48}\right) + \Delta_j\sin\left(2\pi j\frac{\varpi(t)}{48}\right)\right)\\ C_t &= \sum_{l}^{L}\rho_{t-l}T_{t-l}.\end{aligned} \tag{5}$$

Eq. (5) is the classical time series decomposition, where ε_t is distributed as an independently and identically distributed (i.i.d.) standard normal variable with $\varepsilon_t \sim$ i.i.d.$N(0,1)$. $d(t)$ is the step function and $\sigma = 1$. A cycle component, C_t, is added to account for short-term autoregressive behavior. S_t repeats every 52560 steps, corresponding to three years of in-sample data, while D_t cycles through the 48 half hours.

[7] Replacing the transition function in Eq. (2) by $F(\gamma(s_t - c)) = (1 + \exp\{-\gamma/\hat{\sigma}_{S_t}(s_t - c)\})^{-1}$ where $\hat{\sigma}_{s_t}$ is the sample standard deviation of s_t, makes γ approximately normalized. Then a set of grid values for c is defined as sample percentiles of the transition variable.

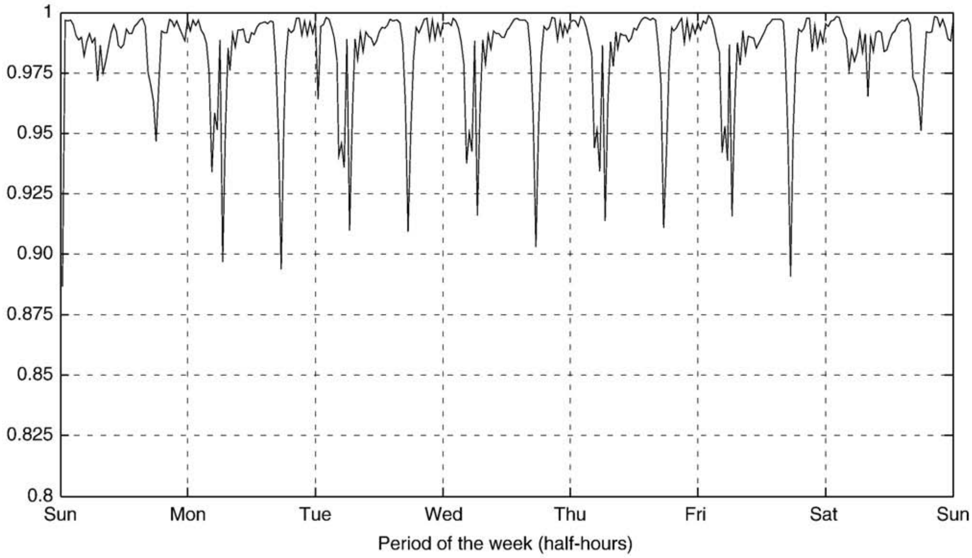

Fig. 3. Lag 1 autocorrelation estimated separately for each period of the week.

We tested the temperature series for the presence of a unit root (or stochastic trend), and, as a consequence, evidence of non-stationary behavior. For this purpose we used the Augmented-Dickey Fuller and Phillips-Perron unit root tests. Under the null hypothesis of a stochastic trend (unit root) in the estimation sample, we rejected the null at the 1% level of significance.[8]

To ensure that non-stationarity in temperatures has not been caused by a combination of large-scale changes in the climate, such as possible global warming, and local changes due to urbanization around measurement stations, we included a trend term in the temperature model given by Eq. (5). The results indicated that the coefficient of the trend term was not statistically significant.[9] This result is consistent with that reported by Svec and Stevenson (2007).

3.3. Other forecasting methods

Load forecasts were also generated from other models to serve as benchmarks in the evaluation of the forecasting performance of the STPAR model. First, a naïve benchmark was considered. In this method the forecast is the last observation available for the same period in the last week. This is the longest cycle considered in the load series. Hence, the forecast function is written as $\hat{y}_t(k)=y_{t+k-336}$.

The second model estimated and used for forecasting the load is a simple feed forward (back propagation) artificial neural network (ANN). In addition, a simple autoregressive process is also specified, estimated and used for generating load forecasts. Both methods included temperature as an explanatory variable.

4. Results

A comparison was made between each of the forecasting models after choosing the best STPAR model. The steps taken in the selection of the preferred model are detailed below, while the results of the forecasting evaluation are discussed in the following sub-section.

4.1. STPAR model

When the order of the autoregressive model was determined, the autocorrelation function, together with the AIC and BIC, suggested the use of lags 1, 48 and

[8] ADF Test: Test statistic = −7.06; 1% critical value = −3.96. Phillips-Perron Test: Test statistic = −17.44; 1% critical value = −3.96.

[9] The coefficient of the trend was very small (that is, 1.74E−08 degrees), with a probability value of 0.8834, indicating that it was not statistically different from zero.

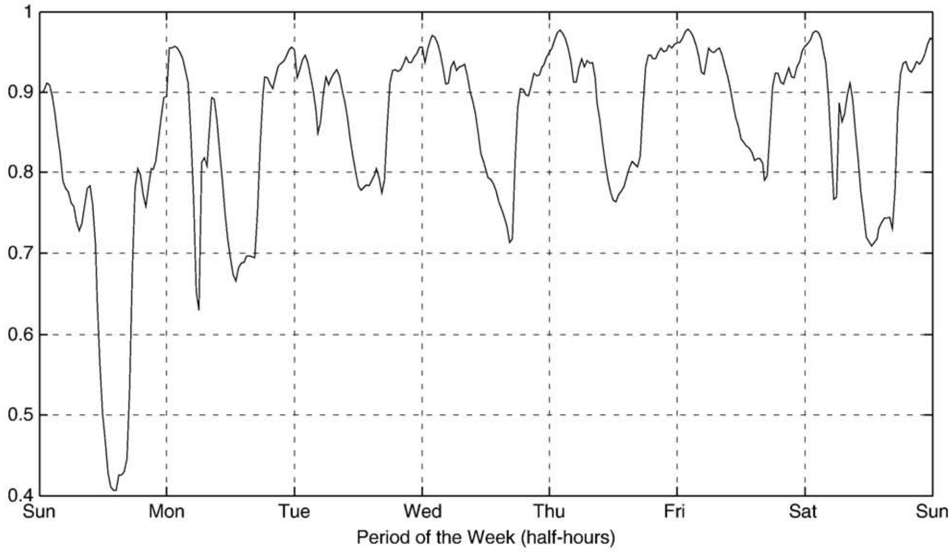

Fig. 4. Lag 48 autocorrelation estimated separately for each period of the week.

336. These lags formed the basis of the linearity test according to the Ljung and Box (1978) statistic.

Figs. 3, 4 and 5 diagrammatically depict the estimated autocorrelations at lags 1, 48 and 336, respectively. It is obvious that the autocorrelations at lags 1, 48 and 336 are time-varying across the 336 half-hours of the week. From Figs. 3 and 5 we can conclude that the variation of the autocorrelation does not differ much across the week for lags 1 and 336. On the other hand, there is clearly daily variation that is much stronger across the week than in the case of the autocorrelations for lag 48. This suggests that weekly periodicity may be relevant for this lag. Further, as it is not clear whether the weekly variation in the autocorrelations at lags 1 and 336 is significant or not,

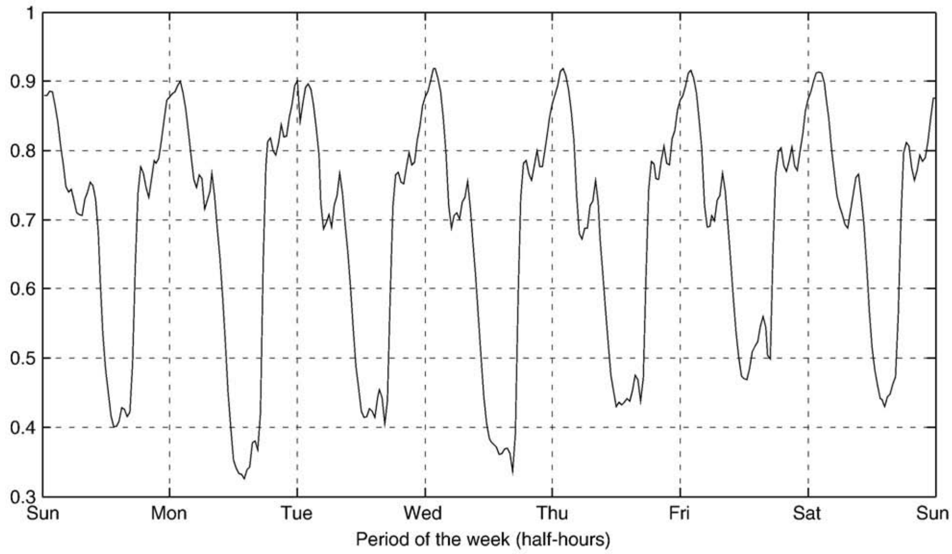

Fig. 5. Lag 336 autocorrelation estimated separately for each period of the week.

Table 1
Value of the test statistic for different values of h and d (p value in parentheses)

h	1	2	3	4	5	6
2	103.7 (0.00)	92.8 (0.00)	85.1 (0.00)	81.5 (0.00)	80.5 (0.00)	81.3 (0.00)
3	94.5 (0.00)	83.4 (0.00)	74.7 (0.00)	70.0 (0.00)	67.2 (0.00)	66.3 (0.00)
4	72.1 (0.00)	65.4 (0.00)	60.3 (0.00)	57.5 (0.00)	56.0 (0.00)	55.0 (0.00)
5	65.7 (0.00)	59.1 (0.00)	54.6 (0.00)	51.8 (0.00)	50.6 (0.00)	50.2 (0.00)

both specifications were tested. We also considered the possibility of both daily and weekly periodicity in the constant term.

After selecting the periodicity, we conducted the linearity test. Table 1 reports the results for the linearity test when the transition variable is temperature. It was also tested using values of h from 2 to 5 and lags of temperature d from 1 to 6.

It is clear that the null hypothesis of linearity is rejected at the 99% level of confidence for all specifications. However, the results are inconclusive concerning the choice of the lag for the transition variable d and the value of h. The p values are of similar magnitude, so that a number of models were candidates for the status of preferred model. The final choice of the values of d and h that specify the most accurate forecasting model was left to the stage where the forecasting performance was evaluated.

Next we estimated the candidate models using nonlinear least squares. Two general specifications were estimated. They were given by:

$$\begin{aligned} y_t &= \beta_{1,0} + \beta_{1,1}y_{t-1} + \beta_{1,48}y_{t-48} + \beta_{1,336}y_{t-336} \\ &\quad + \{(\beta_{2,0} + \beta_{2,1}y_{t-1} + \beta_{2,48}y_{t-48} + \beta_{2,336}y_{t-336}) \\ &\quad \times F(\gamma(s_{t-d} - c))\} + \varepsilon_t \\ \beta_{R,i} &= \alpha_{R,i} + \sum_{j=1}^{h} \left\{ \lambda_{R,ij}\sin\left(2j\pi\left(\frac{D(t)}{48}\right)\right) + \tau_{R,ij}\cos\left(2j\pi\left(\frac{D(t)}{48}\right)\right)\right\}, \end{aligned} \tag{6}$$

and

$$\begin{aligned} y_t &= \beta_{1,0} + \beta_{1,1}y_{t-1} + \beta_{1,48}y_{t-48} + \beta_{1,336}y_{t-336} \\ &\quad + \{(\beta_{2,0} + \beta_{2,1}y_{t-1} + \beta_{2,48}y_{t-48} + \beta_{2,336}y_{t-336}) \\ &\quad \times F(\gamma(s_{t-d} - c))\} + \varepsilon_t \\ \beta_{R,i} &= \alpha_{R,i} + \sum_{j=1}^{h} \left\{ \lambda_{R,ij}\sin\left(2j\pi\left(\frac{D(t)}{48}\right)\right) + \tau_{R,ij}\cos\left(2j\pi\left(\frac{D(t)}{48}\right)\right) \right. \\ &\quad \left. + \zeta_{R,ij}\sin\left(2j\pi\left(\frac{W(t)}{336}\right)\right) + \varsigma_{R,ij}\cos\left(2j\pi\left(\frac{W(t)}{336}\right)\right)\right\}, \end{aligned} \tag{7}$$

where R = 1, 2 and i = 0, 1, 48 and 336.

The two specifications were estimated across a range of values of h and d. This resulted in 17185 × 336 matrices containing all of the 336 half-hours forecast ahead for each half-hour in the out-of-sample data set, and for each of the models that were estimated. In each case we ran 250 Monte Carlo interactions to obtain the forecasts.

The Mean Absolute Percentage Error (MAPE) was used as a measure of forecasting performance in the post sample analysis. It usually expresses accuracy as a percentage, and it is therefore a preferable measure for analysing the performances of different models. Tables 2 and 3 report the MAPEs for the estimated models according to Eqs. (6) and (7), respectively. At each half-hour across the forecast horizon, as determined by the holdout sample, we calculated a forecast for the next week of 336 half-hours. We then calculated the MAPEs for each of these series of forecasts, one for each half-hour over the forecast horizon. This way, a total of 336 values of this statistic were obtained, one for each half-hour horizon from 1 up to 336 steps ahead. Obviously, as one would expect, these estimated MAPEs tend to increase with the horizon. However, in order to obtain a single measure of forecasting performance for each pair (h,d) tried in the model formulation, we calculated the mean of these MAPEs, which resulted in the overall MAPE corresponding to a particular model with specific values of h and d. These overall MAPEs are the values

Table 2
Mean of the MAPEs calculated from Model (6) for each half-hour over the forecast horizon

	d					
h	1	2	3	4	5	6
2	6.61	6.39	6.25	5.73	5.51	5.38
3	**3.42**	**3.47**	**3.55**	**3.55**	**3.66**	**3.88**
4	11.17	11.06	10.93	10.87	10.83	10.61
5	16.48	16.44	16.91	16.46	16.43	16.85

Table 3
Mean of the MAPEs calculated from Model (7) for each half-hour over the forecast horizon

	d					
h	1	2	3	4	5	6
2	8.63	8.07	7.70	7.28	6.98	6.74
3	**3.56**	**3.57**	**3.57**	**3.70**	**3.91**	**4.07**
4	8.82	9.17	9.27	9.45	9.42	9.75
5	12.91	13.36	14.02	14.11	14.27	14.58

presented in these tables, and tabulated according to the various combinations of h and d.

It is clear from the results that the use of h equal to 3 is the best in both specifications. However, the choice of the value of d is not so clear. A comparison of the MAPEs indicates that when d is equal to 1, the MAPE is slightly better when considering the daily (Model 6) or weekly (Model 7) periodicity.

In order to select between the two models for each specification of h and d, in Fig. 6 we graphically depict the forecast of the mean of the MAPEs for both models at each half-hour over the 336 half-hours, making up the forecast horizon of a week. The points in the graph are the MAPEs of the forecasts from each of the preferred models corresponding to Models 6 and 7 for half-hours over the forecast horizon of a year.

What is interesting about the results shown in Fig. 6 is that the forecast skill of each model is quite different across the half-hours of a week, even though the means of the MAPEs are similar. Except for Thursday, Friday and half of Saturday (the first 120 half-hours of a 336 half-hour week), Model 6 with daily periodicity appears to exhibit greater forecast skill than Model 7, which has both daily and weekly periodicity embedded in its specification. Further, the forecasting performance varies across the day. The specification of Model 6 was selected as our preferred model. This was due to the lower level of variation in its performance, as measured by the variation in the mean MAPE values at each half-hour across the prediction interval of a week.

In Fig. 7 we graph the observations in the transition function for the range of temperatures. For high temperatures, the value of the transition function approaches unity. In this case, the transition is more of an immediate jump process that is consistent with

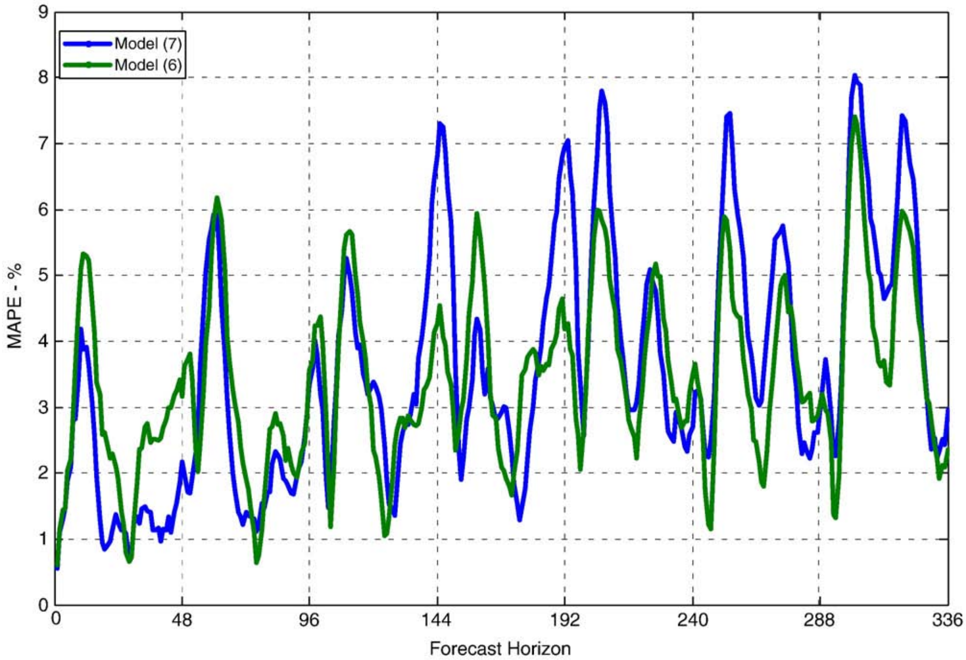

Fig. 6. A comparison of forecasts between specifications (6) and (7) over the forecast horizon.

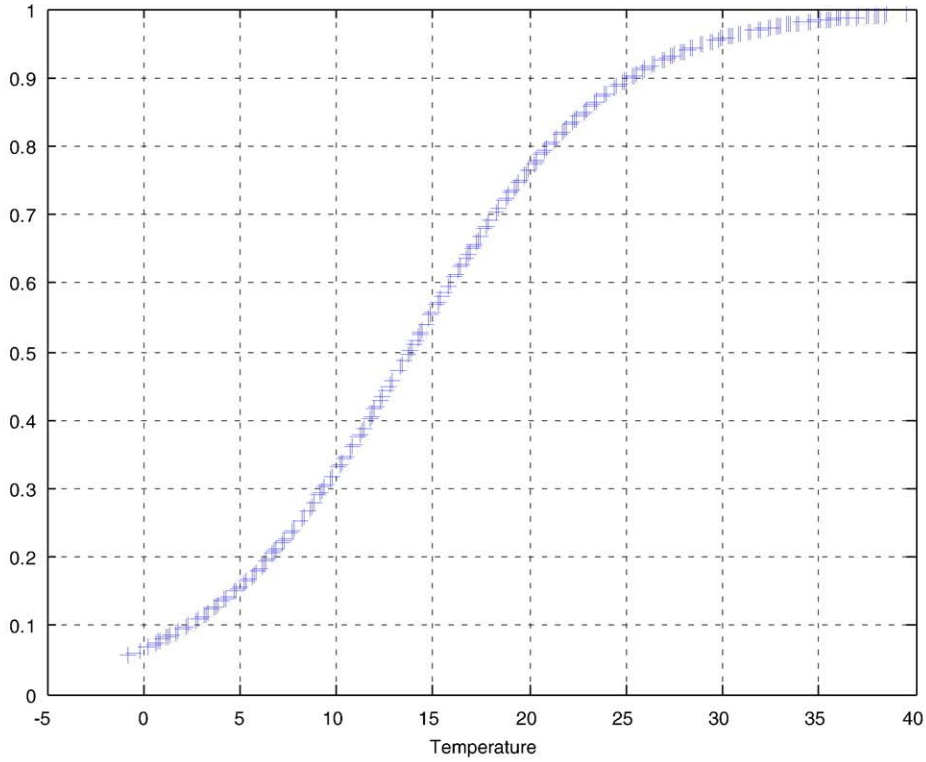

Fig. 7. Transition function for the best STPAR (daily periodicity – Model 6).

the STPAR load-forecasting model, resembling a SETAR-type model.[10] Table 4 shows the results of Lagrange Multiplier (LM) tests at the model estimation stage. The LM tests test the null hypothesis that no statistically significant autocorrelation remains in the error term, thereby obviating the need to model further non-linearity in the data. The results indicate no further model misspecification.

4.2. Comparison of forecasts

Finally, we compare the forecasting performance of the preferred STPAR model (Model (6) with $h=3$ and $d=1$) with that of the other models. The models chosen for the comparison included a simple autoregressive process (AR), a naïve benchmark (NAÏVE) and a feed forward, back propagation artificial neural network (ANN).

The estimated AR model contained lags of 1, 48 and 336, and temperature as an explanatory variable with lag 1. While other lags were tested, there was no statistically significant increase in performance. Several structures of the ANN were tested by changing the

Table 4
Results of the misspecification tests of the estimated STPAR model

	Test for qth order serial correlation								No remaining nonlinearity
q	1	2	4	12	24	48	96	336	
p-value	0.36	0.58	0.63	0.82	0.24	0.16	0.7	0.21	0.83

Table 5
Mean of the MAPEs for all models for each half hour of the prediction interval

	STPAR	AR	NAÏVE	ANN
MAPE	**3.42**	**6.20**	**5.32**	**6.41**

[10] When the immediate jump from one regime to another is determined by a threshold value of a variable of interest (load), the process is self-exciting.

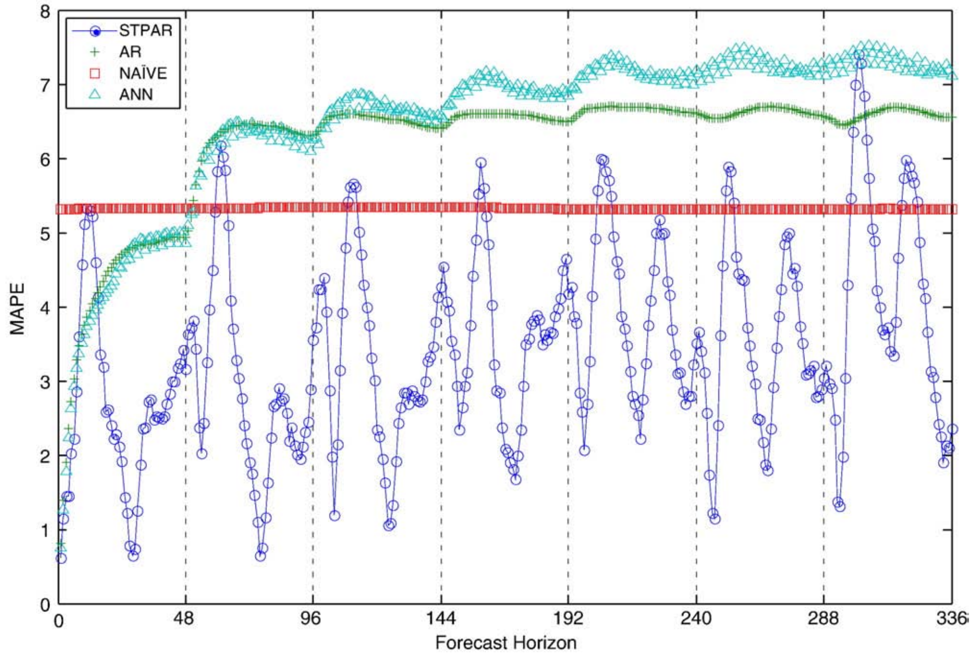

Fig. 8. A comparison of forecasts from the ANN, STPAR, AR and NAÏVE models over the prediction interval.

inputs, number of layers, neurons and activation functions. In the end, the best network had 4 neurons with autoregressive terms of lags 1, 2, 3, 48, 49 and 336, and temperature with lag 1 as inputs. Table 5 gives the means of the MAPEs for all 336 forecasts. Clearly, the STPAR model emerges as the best of the four alternatives.

Fig. 8 shows the mean of the MAPEs from the four alternative load forecasting models, at each half-hour and as an average for the year. Again, it shows that the mean MAPEs exhibit substantial variability. According to this criterion, the STPAR has a better performance than the other models. Of the remaining models, the NAÏVE method outperforms the others, except for the first day of forecasting when the ANN and the simple AR are better.

5. Concluding Remarks

In this paper, we have considered a generalization of the logistic Smooth Transition Autoregressive (STAR) model for short-term load forecasting. It is a combination of periodic models with a smooth transition between the regimes. In addition, we compared the forecasting skill of the STPAR model with that of other alternative forecasting model specifications.

We concluded that, based on its superior performance when compared to the alternative specifications, the STPAR is likely to prove a useful tool for practitioners forecasting the electricity load. Ongoing work will concentrate on a comparison between other popular and robust methods in the literature for forecasting the short-term electricity load series.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions Automatic Control*, *19*(6), 716−723.

Alves da Silva, A. P., & Moulin, L. S. (2000). Confidence intervals for neural network based short term load forecasting. *IEEE Transactions on Power Systems*, *15*(4), 1191−1196.

Basawa, I. V., & Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, *22*(6), 651−663.

Bunn, D. W. & Farmer, E. D. (Eds.). (1985). *Comparative Models for Electrical Load Forecasting.* John Wiley & Sons.

Chan, K. S., & Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, *7*, 178−190.

Cottet, R., & Smith, M. (2003). Bayesian modelling and forecasting of intraday electricity load. *Journal of the American Statistical Association*, *98*, 839−849.

Darbellay, G., & Slama, M. (2000). Forecasting the short term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting*, *16*, 71−83.

Eitrheim, O., & Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, *74*, 59−75.

Fiebig, D., Bartels, R., & Aigner, D. (1991). A random coefficient approach to the estimation of residual end-use load profiles. *Journal of Econometrics*, *50*, 297−327.

Franses, P. H., & Paap, R. (2004). *Periodic Time Series Models.* Oxford: Oxford University Press.

Ghysels, E., & Osborn, D. R. (2001). *The Econometric Analysis of Seasonal Time Series.* Cambridge: Cambridge University Press.

Hippert, H. S., Bunn, D. W., & Souza, R. C. (2005). Large neural networks for electricity load forecasting: Are they overfitted? *International Journal of Forecasting*, *21*, 425−434.

Huang, S. J. (2003). Short term load forecasting via ARMA model identification including non-Gaussian process consideration. *IEEE Transactions on Power Systems*, *18*, 673−679.

Huang, C.–M., Huang, C.–J., & Wang, M.–L. (2005). A particle swarm optimization to identify the ARMAX model for short-term load forecasting. *IEEE Transactions on Power Systems*, *20*, 1126−1133.

Khotanzad, A., Afkhami-Rohami, R., & Maratukulam, D. (1998). ANNSTLF – Artificial Neural Network Short Term Load Forecaster – Generation Three. *IEEE Transactions on Power Systems*, *13*(4), 1413−1422.

Koopman, S. J., Ooms, M., & Carnero, M. A. (2007). Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. *Journal of the American Statistical Association*, *102*, 16−27.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*, 297−303.

Lund, R., & Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, *21*(1), 75−93.

Lundbergh, S., & Teräsvirta, T. (2000). *Forecasting with smooth transition autoregressive model.* Working Paper Series in Economics and Finance, vol. 390. Stockholm School of Economics.

Metaxiotis, K., Kagiannas, A., Askounis, D., & Psarras, J. (2003). Artificial intelligence in short term electric load forecasting: A state of the art survey for the researcher. *Energy Conversion and Management*, *44*, 1525−1534.

Nowicka-Zagrajek, J., & Weron, R. (2002). Modeling electricity loads in California: ARMA models with hyperbolic noise. *Signal Processing*, *82*(12), 1903−1915.

Peirson, J., & Henley, A. (1994). Electricity load and temperature. *Energy Economics*, *16*, 235−243.

Phillips, P., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*, 335−346.

Ramanathan, R., Engle, R. F., Granger, C., Vahid-Arahi, F., & Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, *13*, 161−174.

Reis, A. J. R., & Alves da Silva, A. P. (2005). Feature Extraction Via Multi-Resolution Analysis for Short Term Load Forecasting. *IEEE Transactions on Power Systems*, *20*(1), 189−198.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461−464.

Smith, M. (2000). Modelling and short-term forecasting of New South Wales electricity system load. *Journal of Business and Economic Statistics*, *18*(4), 465−478.

Soares, L. J., & Medeiros, M. C. (2005). Modeling and forecasting short-term electric load demand: a two step methodology. *Working Paper, Department of Economics, PUC-Rio*.

Soares, L. J., & Souza, L. R. (2006). Forecasting electricity demand using generalized long memory. *International Journal of Forecasting*, *22*, 17−28.

Stevenson, M., & Peat, M. (2001). Forecasting Australian unemployment rates. *Australian Journal of Labour Economics*, *4*(1), 38−52.

Svec, J., & Stevenson, M. (2007). Modelling and forecasting temperature based weather derivatives. *Global Finance Journal*, *18*(2), 185−204.

Taylor, J. W. (2003). Short term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, *54*, 799−805.

Taylor, J. W. (2006). Density forecasting for the efficient balancing of the generation and consumption of electricity. *International Journal of Forecasting*, *22*, 707−724.

Taylor, J. W., & Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, *19*, 57−70.

Taylor, J. W., de Menezes, L. M., & McSharry, P. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, *22*, 1−16.

Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, *89*, 208−218.

Tong, H. (1983). *Threshold models in non-linear time series analysis. Lecture Notes in Statistics*, *vol. 21*. Heidelberg: Springer.

Tong, H. (1990). *Non Linear Time Series: A Dynamical System Approach. Oxford Statistical Science Series*, *vol. 6*. Oxford University Press.

van Dijk, D., Teräsvirta, T., & Franses, P. H. (2002). Smooth Transition Autoregressive Models – A survey of recent developments. *Econometric Reviews*, *21*(1), 1−47.