[1]Prediction of daily peak electricity demand in South Africa using volatility forecasting models

C. Sigauke[1*], D. Chikobvu[2]

[1*]Department of Statistics and Operations Research, School of Mathematical and Computer Sciences, University of Limpopo, Turfloop Campus, P. Bag X1106, Sovenga, 0727. South Africa sigaukec@ul.ac.za

[2]Department of Mathematical Statistics and Actuarial Science, University of the Free State, P. O. Box 339, Bloemfontein 9300, South Africa chikobvu@ufs.ac.za

Abstract

Daily peak electricity demand forecasting in South Africa using a seasonal autoregressive integrated moving average (SARIMA) model, a SARIMA model with generalized autoregressive conditional heteroskedastic errors (SARIMA-GARCH) and a regression-SARIMA-GARCH (Reg-SARIMA-GARCH) model are presented in this paper. The GARCH modelling methodology is introduced to accommodate the possibility of serial correlation in volatility since the daily peak demand data exhibits non-constant mean and variance, and multiple seasonality corresponding to weekly and monthly periodicity. The proposed Reg-SARIMA- GARCH model is designed in such a way that the predictor variables are initially selected using a multivariate adaptive regression splines algorithm. The developed models are used for out of sample prediction of daily peak demand. A comparative analysis is done with a piecewise linear regression model. Results from the study show that the Reg-SARIMA-GARCH model produces better forecast accuracy with a mean absolute percent error (MAPE) of 1.42%.

Key words: Volatility, Daily peak demand, SARIMA, GARCH, Piecewise linear regression.

1. Introduction

Prediction of daily peak load demand is very important for decision making processes in the electricity sector. Decision making in this sector involves planning under uncertainty. This involves for example finding the optimal day to day operation of a power plant and even strategic planning for capacity expansion. The demand of electricity forms the basis for power system planning, power security and supply reliability (Ismail *et al.*, 2009). It is important therefore to produce very accurate forecasts as the consequences of underestimation or overestimation can be costly. As noted by Taylor (2008), accurate short-term forecasts are needed by both generators and consumers of electricity particularly during periods of abnormal peak load demand.

In this paper seasonal autoregressive integrated moving average (SARIMA) model, a SARIMA with generalized autoregressive conditional heteroskedastic errors (SARIMA-GARCH) model and a regression-SARIMA-GARCH (Reg-SARIMA-GARCH) model are developed and used for out of sample predictions of daily peak demand (DPD) using South African data. The models are designed for short term forecasting, up to seven days ahead. The Reg-SARIMA-GARCH model captures factors such as day of the week, holiday and temperature effects including multiple seasonality.

[1]*Corresponding author: a. Email: sigaukec@ul.ac.za Tel. +27 15 268 2188, Fax +27 15 268 3075

The major challenge in most conventional Reg-SARIMA models is that of selecting a minimum number of predictor variables and ranking them in order of their importance. The proposed Reg-SARIMA-GARCH model developed in this paper is designed in such a way that the predictor variables are initially selected using a multivariate adaptive regression splines algorithm (Friedman, 1991).

Accurate prediction of daily peak load demand helps in the determination of consistent and reliable supply schedules during peak periods. Accurate short term daily peak load forecasts will enable effective load shifting between transmission substations, scheduling of startup times of peak stations, load flow analysis and power system security studies.

The rest of the paper is organized as follows, in Section 2 the data used is described and a preliminary analysis carried out. The models are presented in Section 3 and a detailed discussion of the results is covered in Section 4. A comparative analysis of the developed models is done with a piecewise linear regression model in Section 5. The summary and conclusion of the paper are covered in Section 6.

2. Data and Definitions

The data used in this paper is on net energy sent out (NESO) from distribution in response to some demand of electrical power. NESO (measured in megawatts) is defined as the rate at which electrical energy is delivered to customers. In this paper NESO is used as a proxy of electrical demand after adjusting for energy losses. This definition of electrical demand has its weakness. Electrical demand is bounded by the power plants available to provide supply at any time of the day including the need for reserve capacity. Demand cannot exceed supply and there are no market forces acting to influence electricity prices and hence reducing demand in the short run. Prices are generally fixed in the short run. If demand were to exceed supply, intervention takes place in the form of for example, load shedding. Load shedding is the last resort used to prevent a system-wide blackout. This NESO definition excludes the demand from households, companies etc, who are willing and able to pay for electricity but currently do not have access to electrical power. Despite the weakness in the NESO definition of electrical demand, it is still a good and measurable proxy of electricity demand.

The data is on daily peak demand (DPD) from 1 January 1996 to 14 December 2009 (n=5097 daily peak demand observations)[2]. Since demand is normally recorded on an hourly basis, DPD is the maximum hourly demand in a 24-hour period. Daily peak demand modeling is important as it provides short term forecasts which will assist in economic planning and dispatching of electric energy. Aggregated DPD data is collected for the industrial, commercial and domestic sectors of South Africa.

[2] The data used in this paper can be provided upon request

2.1 Preliminary Analysis

The time series plot of DPD in Figure 1 shows a positive linear trend and a strong seasonal fluctuation with DPD high in winter and low in summer. The trend is mainly due to economic development of the country. The winter peaks in the Southern Hemisphere are around June/July of each year. A casual inspection of the graph also shows that the variance is not constant.

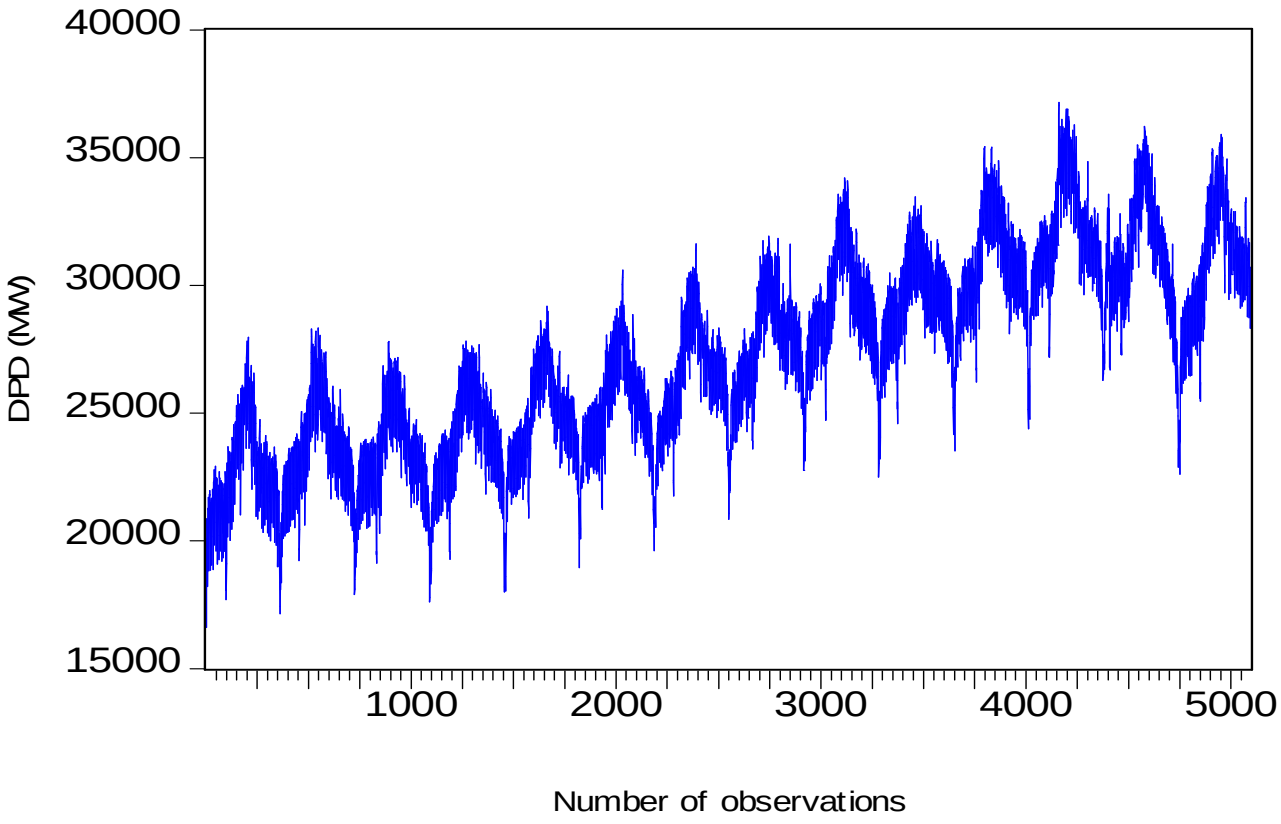

Figure 1: Time series plot for DPD (in Megawatts) for the period 1/1/1996-14/12/2009

A test for a stochastic trend in the DPD series using the Augmented Dickey Fuller test shows that the series is not stationary. The null hypothesis of a stochastic trend was accepted showing that there is a unit root. Stationarity was achieved by a first difference of the data.

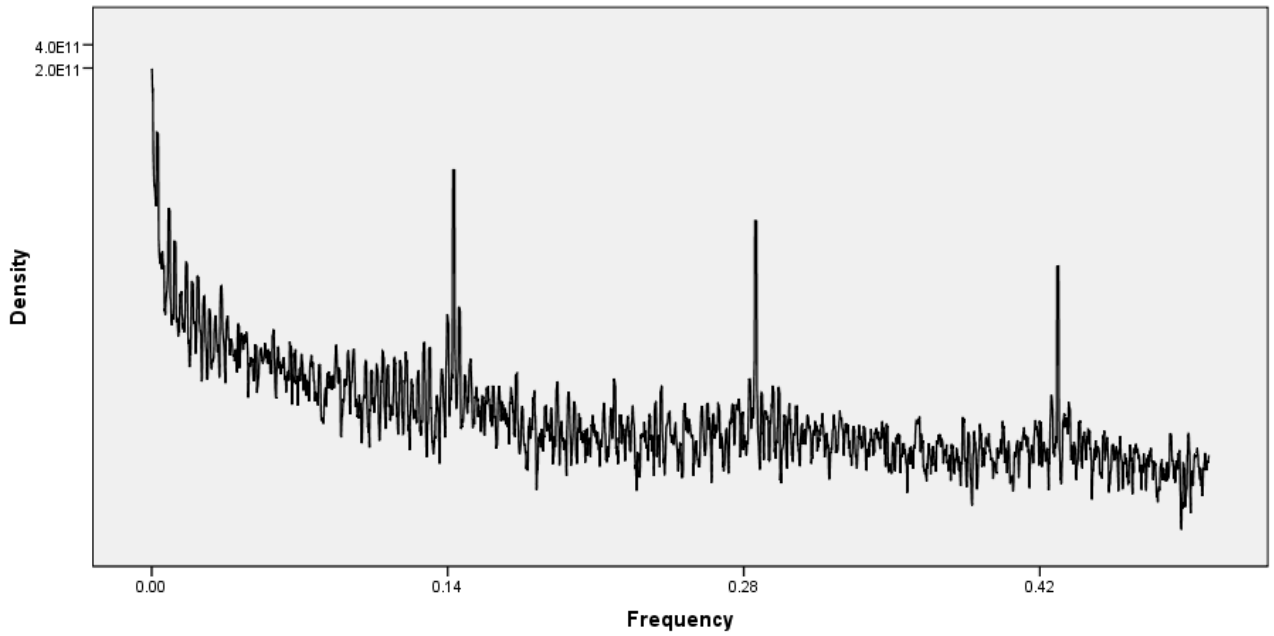

Figure 2: Spectral density of DPD

A spectral analysis revealed periodicity in the data. Figure 2 shows the spectral density of DPD. The first major peak in the spectral density is around 0.14 indicating the presence of a periodic movement of seven days.

3. The Models

SARIMA, SARIMA-GARCH and Reg-SARIMA-GARCH models are presented in this section. The developed models are then used for out of sample predictions of DPD. In all models DPD is taken as the dependent variable. The data is transformed by taking natural logarithms to reduce the impact of heteroskedasticity that may be present because of the large data set (Hekkenberg *et al.*, 2009).

3.1 Seasonal ARIMA Model (SARIMA)

Load demand forecasting has been studied extensively over the years using time series, regression based methods and artificial and computational intelligence (Ramanathan *et al.*, 1997; Mirasgedis *et al.*, 2006; Amaral *et al.*, 2008; Amin-Naseri and Soroush, 2008; Gosh, 2008; Soares and Medeiros, 2008; Taylor, 2008; Truong *et al.*, 2008; Goia *et al.*, 2010, among others). Updated review of different methods can be found in (Feinberg and Genethliou, 2005; Hahn *et al.*, 2009). The general multiplicative SARIMA $(p,d,q)\times(P,D,Q)_s$ model used in this paper is given in equation (1) and the derivation is done in appendix A1.

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D\, y_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \quad \varepsilon_t \sim N(0,\sigma_t^2) \tag{1}$$

where y_t represents daily peak demand (in megawatts) observed on day $t\,(t=1,2,...,n)$ and ε_t represents the error term at time t with variance σ_t^2 and potentially subject to conditional heteroskedasticity, s is the seasonal length and B is a backshift operator defined as $By_t = y_{t-1}$, $\phi_p(B), \Phi_P(B^s), \theta_q(B), \Theta_Q(B^s)$, are backshift operator polynomials of orders p, P, q, Q respectively, modeling the regular and weekly autoregressive and mean average effects respectively. ∇^d and ∇_s^D are difference operators defined as $\nabla^d y_t = (1-\mathrm{B})^{\mathrm{d}} y_t$ and $\nabla_s^D y_t = (1-\mathrm{B}^{\mathrm{s}})^{\mathrm{D}} y_t$, and c is a constant term.

3.2 Volatility Forecasting Models

In conventional SARIMA models, the variance of the disturbance term is assumed to be constant. Causal inspection of the electricity demand time series plot shown in Figure 1 suggests that the series does not have a constant variance. The series exhibits phases of high demand (in winter) followed by periods of low demand (in summer). The assumption of homoskedasticity (constant variance) seems inappropriate, since the data exhibits non-constant mean and variance, and multiple seasonality corresponding to weekly and monthly periodicity. The GARCH modelling methodology is introduced to accommodate the possibility of serial correlation in volatility. Models for volatility forecasting were first developed by Engle (1982). These models known as the autoregressive conditional heteroskedasticity (ARCH) models were developed to capture the non constant variance. ARCH models were later extended to generalized ARCH (GARCH) models by Bollerslev (1986) and Nelson (1991). Modelling volatility in time series data using GARCH – type models has been studied extensively over the past three decades, (Engle, 1982; Bollerslev, 1986; Nelson, 1991; Taylor, 2006; Aknouche and Bentarzi, 2008; Mulera and Yohaib, 2008; Doornik and Ooms, 2008; Ghahramani and Thavaneswaran, 2008; Horv´ath *et al*., 2008; Ismail *et al.*, 2009; He and Maheu, 2010; Kim *et al*., 2010, among others). Work closely related to ours is that of Taylor (2006) who investigated methods for Net Imbalance Volume density forecasting. The author decomposed the problem into point forecasting and volatility forecasting. A seasonal ARMA model and a periodic AR model with simplistic volatility forecasting gave good results.

3.2.1 SARIMA-GARCH Model

The SARIMA-GARCH model is one in which the variance of the error term of the SARIMA model follows a GARCH process. The model used for the DPD series can be written as:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$
$$\varepsilon_t = z_t\sigma_t,$$

$$z_t \sim i.i.d \text{ with } E(z_t) = 0, \text{Var}(z_t) = 1$$
$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2 \quad \textbf{(2)}$$

where y_t represents the time series as defined in (1), p is the order of GARCH process; q is the order of ARCH process; a_0, a_i and b_j are constants; ε_t is the error term; σ_t^2 is the conditional variance of ε_t; ε_{t-i}^2 is the news about the volatility from the i^{th} lag period and σ_{t-j}^2 is the j^{th} lag period forecast error variance, z_t is a standardized error term.

3.2.2 Reg-SARIMA-GARCH Model

There are various factors that influence electricity load demand. Some of these factors are temperature, day of the week, holidays, daily and monthly seasonality. In this section, a Reg-SARIMA-GARCH model is developed which will capture the day of the week, holiday and monthly seasonality effects. Temperature is not included in the Reg-SARIMA-GARCH model. The authors are aware that the inclusion of this factor could have a significant improvement on prediction particularly in winter when heating systems are used and also in summer when air conditioning appliances are used. It should be noted that it is easy to include weather variables such as temperature in the model. The influence of temperature on energy demand will studied elsewhere. Several papers in literature have adopted the same strategy of not including temperature (Carpinteiro *et al.*, 2004; Taylor *et al.*, 2006; Sores and Souza, 2006; Soares and Medeiros, 2008).

The Reg-SARIMA-GARCH model is a regression seasonal ARIMA model with error terms following a GARCH process. The model can be written as:

$$\begin{aligned}
&\phi_p(B)\Phi_P(B^s)\psi_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \\
&\varepsilon_t = z_t\sigma_t, \\
&z_t \sim i.i.d \text{ with } E(z_t) = 0, \text{Var}(z_t) = 1 \\
&\sigma_t^2 = \gamma w_t^{|} \\
&\text{where} \\
&w_t = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, \sigma_{t-1}^2, ..., \sigma_{t-p}^2), \\
&\gamma = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p) \\
&\text{and } \psi_t = (1-B)^d(1-B^s)^D y_t - \sum_{g=1}^{G} \beta_g (1-B)^d (1-B^s)^D x_{gt}
\end{aligned} \quad \textbf{(3)}$$

where x_{gt} is the g^{th} regression variable at time t, β_g is the g^{th} regression parameter and the other variables and parameters are as defined in sections 3.1 and 3.2.1 respectively.
The derivation of the Reg-SARIMA-GARCH model in equation (3) is given in appendix A2.

4. Results and Discussion

This section presents the results of estimating the parameters of the SARIMA, SARIMA-GARCH and Reg-SARIMA-GARCH models represented by equations (1)-(3) along with their diagnostic tests. These results are compared with a piecewise linear regression model. The SARIMA model presented in Table 1 can be written as:

$$(1-\phi_1 B-\phi_2 B^2-\phi_4 B^4-\phi_6 B^6)\omega_{tt}=(1-\theta_2 B^2)(1-\Theta_1 B^7-\Theta_2 B^{14}-\Theta_6 B^{42})\varepsilon_t+c \qquad \textbf{(4)}$$

where $\omega_t=(1-B)^d(1-B^s)^D \ln y_t$, with d = 0, s = 7 and D = 1. The coefficients of the auto regressive parameters ϕ_1, ϕ_2 and ϕ_6 are all positive implying that there will be an increase in DPD when there is an increase in the corresponding lagged DPD and the rest are negative meaning a decrease in DPD. The seasonal moving average parameters at lags 7, 14 and 42 are an indication of strong seasonality effects.

The model parameters were estimated using the maximum likelihood method. The best model has a root mean square error (RMSE) of 565 and a mean absolute percentage error (MAPE) of 1.44%. The parameter estimates of the best SARIMA model developed for the DPD series are presented in Table 1 with the p-values shown in parentheses.

Table 1: Parameter estimates of the SARIMA Model

Par	c	ϕ_1	ϕ_2	ϕ_4	ϕ_6	θ_2	Θ_1	Θ_2	Θ_6
Coef	0.000364 (0.0366)	0.839 (0.000)	0.110 (0.000)	-0.066 (0.000)	0.068 (0.000)	-0.161 (0.000)	-0.826 (0.000)	-0.087 (0.000)	-0.056 (0.000)

Residual analysis shows hetreskedasticity with a mean around zero. A graphical plot of the residuals is shown in Figure 3.

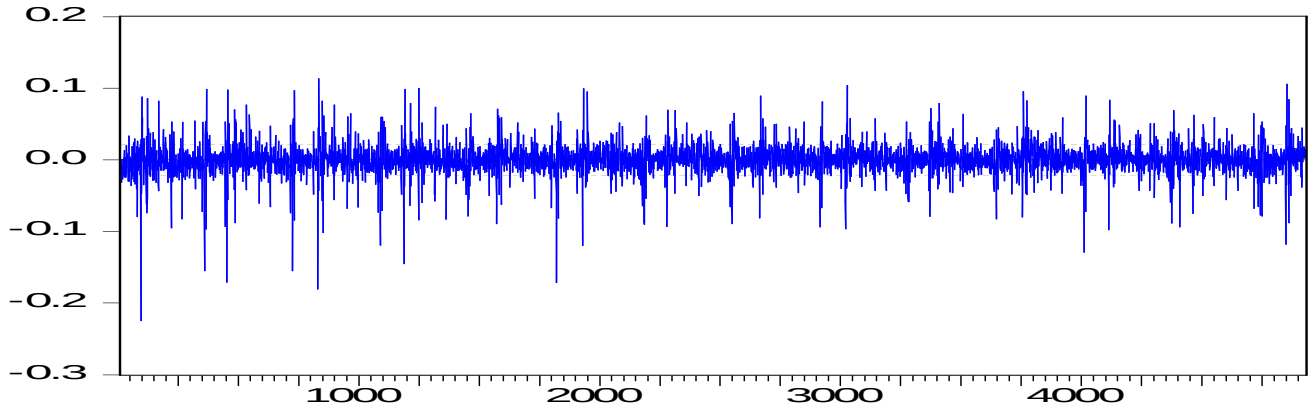

Figure 3: Graphical plot of standardized residuals from the SARIMA model.

The residuals were further investigated for heteroskedasticity. The graphical plot of the squared residuals is shown in figure 4. Since the residual variance, σ_t^2, is unobservable, the squared residuals serve as a proxy. Figure 4 suggests heteroskedasticity.

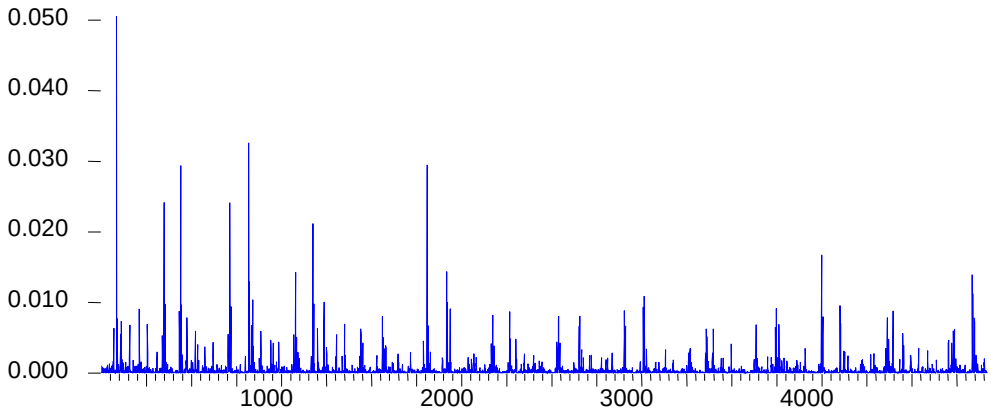

Figure 4: Graphical plot of squared residuals from the SARIMA model

The Ljung-Box Q-statistics on standardardized residuals with various lag values up to lag 40 with their p-values shown in parentheses are Q(10) = 3.7 (0.158), Q(20) = 10.2 (0.601), Q(30) = 14.9 (0.867) and Q(40) = 35.7 (0.297). All the Q statistics were insignificant up to lag 40 at the 5% level indicating that there is no excessive autocorrelation left in the residuals. The Ljung-Box Q-Statistics on squared standardardized residuals were $Q^2(10)$ = 425.4 (0.000), $Q^2(20)$ = 433.6 (0.000), $Q^2(30)$ = 443.2 (0.000) and $Q^2(40)$ = 453.9 (0.000). All the Ljung-Box Q statistics were significant up to lag 40 at the 5% level indicating that there is heteroskedasticity. Engle's LM test was carried out and the null hypothesis that there are no GARCH effects in the residuals is rejected at the 5% level. This suggests that there can be some improvement on the current model through volatility modelling.

Several SARIMA–GARCH models were considered and the best model is selected based on the Akaike information criterion (AIC). The model parameters are estimated using the maximum likelihood method. The estimates are obtained by the Berndt *et al*. (1974) algorithm using numerical derivatives. The parameter estimates of the best model along with their p-values in parentheses are presented in table 2.

Table 2: Parameter estimates of the SARIMA-GARCH Model

Parameter	ϕ_1	ϕ_6	Φ_1	θ_2	Θ_1
Coefficient	0.8668 (0.0000)	0.1048 (0.0001)	0.1511 (0.0000)	-0.0755 (0.0000)	-0.9708 (0.0009)

Variance Equation

Parameter	c	α_1	α_2	β_1
Coefficient	0.0000264 (0.0003)	0.5550 (0.0000)	-0.4069 (0.0000)	0.8164 (0.0000)

The best model has an RMSE of 553 and a MAPE of 1.41%. The positivity conditions imposed on the GARCH model parameters are relaxed in line with Nelson and Cao (1992) where the sum of the parameters in the model should be less than one. Volatility shocks are persistent in the time series data since the sum of the ARCH and GARCH terms (which is 0.9645) in the variance equation are close to one. The Ljung-Box Q-statistics on squared standardardized residuals with various lag values up to lag 40 with their p-values shown in parentheses are $Q^2(6)$ = 1.2947 (0.255), $Q^2(10)$ = 5.2 (0.392), $Q^2(20)$ = 13.9 (0.531), $Q^2(30)$ = 18.7 (0.812) and $Q^2(40)$ = 20.3 (0.977). All the Ljung-Box Q statistics are insignificant up to lag 40 at the 5% level indicating that there is no excessive serial autocorrelation left in the residuals. Engle's LM test was carried out and the null hypothesis that there are no GARCH effects in the residuals is accepted at the 5% level.

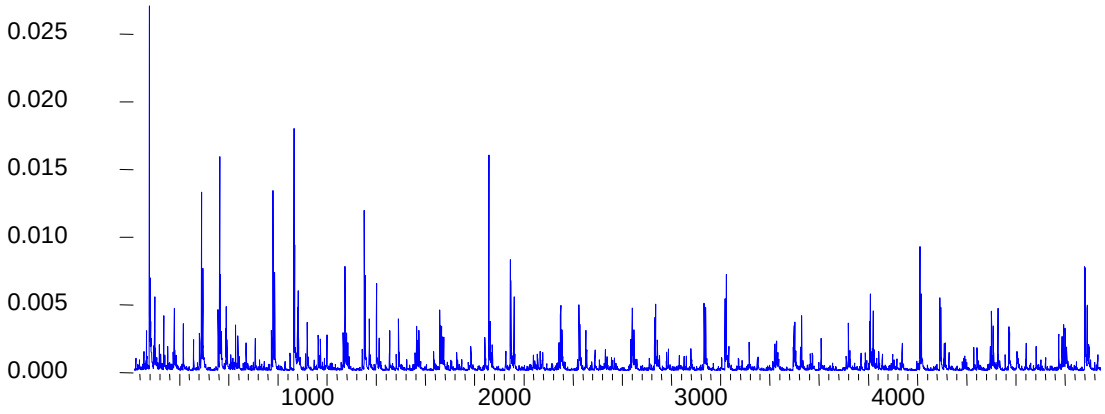

Figure 5: Graphical plot of squared standardized residuals from the SARIMA-GARCH model

Figure 5 shows the graphical plot of the squared standardized residuals (residuals squared divided by the estimated values for σ_t^2, that is $\sigma_t^2 = \frac{\varepsilon_t^2}{z_t^2}$). A comparison of Figures 5 and 4

shows that the SARIMA-GARCH model has accommodated much of the heteroskedasticity in the residuals as we now have smaller spikes. Other SARIMA-GARCH models were considered but the authors were not able to improve results on the squared standardized residuals. This led to the development of Reg-SARIMA-GARCH models in an effort to further improve the SARIMA-GARCH model. Table 3 shows a summary of the estimates of the parameters of the Reg-SARIMA-GARCH model developed along with their p-values shown in parentheses. The predictor variables shown in Table 3 are initially selected from a total of 23 predictor variables using a multivariate adaptive regression splines algorithm (Friedman, 1991). The model parameters are then estimated using the maximum likelihood method. The estimates are obtained by the Berndt *et al*. (1974) algorithm using numerical derivatives.

The Ljung-Box Q-statistics on squared standardardized residuals with various lag values up to lag 40 with their p-values shown in parentheses are $Q^2(10) = 8.9$ (0.113), $Q^2(20) = 11.8$ (0.692), $Q^2(30) = 13.9$ (0.963) and $Q^2(40) = 17.3$(0.995). All the Ljung-Box Q statistics were insignificant up to lag 40 at the 5% level indicating that there is no serial autocorrelation left in the residuals. The Engle's LM test is carried out and the null hypothesis that there are no GARCH effects in the residuals is accepted at the 5% level.

The best model has an RMSE of 549 and a MAPE of 1.40%. Volatility shocks are persistent in the time series data since the sum of the GARCH terms (which is 0.928914) in the variance equation are close to one. The developed model shown in table 3 can be written as:

$$(1-\phi_1 B-\phi_2 B^2)(1-B^7)(\ln y_t+\gamma_5\text{Friday}+\gamma_6\text{Saturday}+\gamma_7\text{Sunday}-g_2\text{February}+g_7\text{July} + \lambda \text{H}_{h-1}+\mu \text{H}_h+\delta \text{H}_{h+1}) = (1+\Theta_1 B^7+\Theta_2 B^{14}+\Theta_6 B^{42})\varepsilon_t + c \quad \textbf{(5)}$$

with the variance equation written as $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. Table 3 shows the estimated parameters.

Table 3: Parameter estimates of the Reg – SARIMA – GARCH Model

Par	c	H_{h-1}	H_h	H_{h+1}	Friday	Saturday	Sunday
Coef	0.00085 (0.0041)	-0.00715 (0.0003)	-0.00451 (0.0041)	-0.00148 (0.0372)	-0.00022 (0.0014)	-0.00024 (0.0033)	-0.00025 (0.0258)
Par	-February	July	ϕ_1	ϕ_2	Θ_1	Θ_2	Θ_6
Coef	0.00240 (0.5632)	0.00138 (0.3013)	0.85805 (0.0000)	0.08301 (0.0025)	-0.80683 (0.0000)	-0.09244 (0.0048)	-0.03281 (0.0524)

Variance Equation

Par	c	α	β
Coef	0.00007 (0.0000)	0.47206 (0.0000)	0.45578 (0.0000)

From Table 3 all the coefficients of the dummy variables representing holiday, day before a holiday, day after a holiday, Friday, Saturday and Sunday are negative meaning that DPD decreases during these days. This is due to the fact that most companies will be closed during these days. The dummy variable for month of July is positive meaning that DPD increases. This

month is in winter when heating appliances are used. The dummy variable for February is negative, meaning that there will be a decrease in DPD. This is possibly due to the fact that February is in summer. In South Africa DPD is more sensitive to winter periods than summer periods (Sigauke and Chikobvu, 2010).

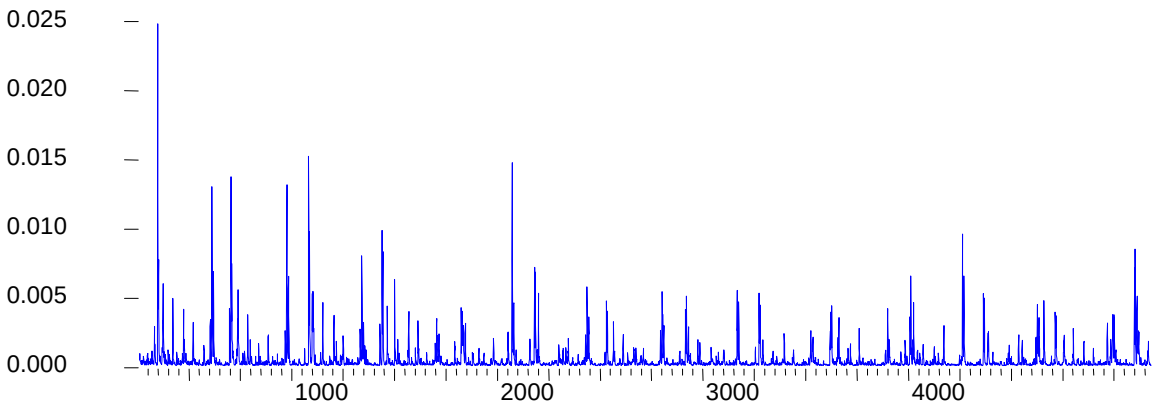

Figure 6: Graphical plot of squared standardized residuals from the Reg-SARIMA–GARCH model

Figure 6 shows the graphical plot of the squared standardized residuals. A comparison of Figure 6 with Figure 5 shows that the Reg-SARIMA-GARCH model has accommodated much of the heteroskedasticity in the residuals. Other Reg-SARIMA-GARCH models were considered but the authours were not able to improve further on the heteroskedasticity problem revealed by the squared standardized residuals.

5. Evaluating the predictive abilities of the models

In short term load forecasting RMSE and MAPE are generally used to present load forecasting error (Munoz *et al.*, 2010). These accuracy measures are used for the evaluation of the developed models for peak load demand forecasting in the out of sample predictions for the period 1 July to 14 December 2009. The training period was 1 January 1996 to 30 June 2009. The paper concentrated on daily peak demand modeling which is important for providing short term forecasts which will assist in optimal dispatching of electrical energy.

The performance of the developed models is evaluated by comparing them with a naïve simple piecewise linear regression model. The piecewise linear regression model is found in appendix A3. A detailed discussion of the piecewise linear regression model is found in (Sigauke and Chikobvu, 2010). Table 4 presents the comparative analysis of the models. The developed models outperformed the piecewise linear regression model.

Table 4: Out-of-sample forecast evaluation for the period 01/07/2009-14/12/2009

Forecasting Models	Performance Criteria (Validation Period)

	MAPE	RMSE
Piecewise Linear Regression	2.77	941
SARIMA	1.47	571
SARIMA-GARCH	1.43	556
RegSARIMA-GARCH	1.42	554

SARIMA models work well when the data exhibits a linear trend and are only good for short term forecasting. The Reg-SARIMA-GARCH model has the least MAPE, showing that it is the best fitting model. The Reg-SARIMA-GARCH model is simple to implement, reliable and provides information about the importance of each predictor variable. The results from using the Reg-SARIMA-GARCH model are relatively robust.

6. Conclusion

This paper has investigated some hybrid models for daily peak load demand forecasting. The problem is decomposed into point and volatility forecasting. Results show that the regression-seasonal autoregressive integrated moving average (Reg-SARIMA-GARCH) model with heteroskedastic error terms produces better forecast accuracy with a mean absolute percent error of 1.42%. Accurate prediction of daily peak load demand is very important for decision makers in the energy sector. This helps in the determination of consistent and reliable supply schedules during peak periods. Accurate short term load forecasts will enable effective load shifting between transmission substations, scheduling of startup times of peak stations, load flow analysis and power system security studies.

Areas for further study would include combining forecasts produced by the different methods. This can be done through use of techniques which will minimize variance of the forecast and also the mean absolute percentage error. Another interesting area for further study would be to model annual winter peaks using extreme value theory. The development of a stochastic integer recourse model to optimize electricity distribution would be another interesting area to study including a formal test to see whether an improvement in forecast accuracy between two models is statistically significant. These areas will be studied elsewhere.

Acknowledgments

The authors are grateful to the referees for their useful comments and suggestions. We would like to thank Eskom for providing the data, Department of Statistics and Operations Research, University of Limpopo and Department of Mathematical Statistics and Actuarial Science, University of the Free State for using their resources and to the numerous people who assisted in making comments on this paper.

Appendix

A1. Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

Let y_t represent daily peak electricity demand. Then y_t follows a multiplicative seasonal

ARIMA model for daily peak demand time series given by:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D y_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,\sigma^2) \tag{6}$$

where

$$\phi_p(B) = 1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p,\ \Phi_P(\mathrm{B}^s) = 1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_P B^{Ps},$$
$$\theta_q(\mathrm{B}) = 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q,\ \Theta_Q(\mathrm{B}^s) = 1-\Theta_1 B^s-\Theta_2 B^{2s}-\ldots-\Theta_Q B^{Qs},$$

c is a constant term, ε_t is an independent and identically distributed random noise. B is a backshift operator and $\phi_p(B), \Phi_P(B^s), \theta_q(B), \Theta_Q(B^s)$, are backshift operator polynomials of orders p, P, q, Q respectively, modeling the regular, daily and weekly autoregressive and mean average effects respectively. ∇^d and ∇_s^D are difference operators defined as $\nabla^d y_t = (1-\mathrm{B})^{\mathrm{d}} y_t$ and $\nabla_s^D y_t = (1-\mathrm{B}^{\mathrm{s}})^{\mathrm{D}} y_t$, and s is the seasonal length.
Let $\omega_t = (1-\mathrm{B})^{\mathrm{d}}(1-\mathrm{B}^{\mathrm{s}})^{\mathrm{D}} y_t$, where $(1-\mathrm{B})^d$ represents nonseasonal differencing operator of order d and $(1-\mathrm{B}^{\mathrm{s}})^D$ is the seasonal differencing operator of order D and both d and D are positive integers. Then ω_t is a stationary SARMA $(p,q)\times(P,Q)_s$ model given by:

$$\phi_p(B)\Phi_P(B^s)\omega_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \tag{7}$$

A2. The Reg-SARIMA-GARCH model

A general multiplicative SARIMA model for a variable v_t can be written as:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D v_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t. \tag{8}$$

The model in (8) can be extended by use of a time varying mean function which can be modeled through linear regression effects. A linear regression equation for a time series y_t can be written as:

$$y_t = \sum_{g=1}^{G}\beta_g x_{gt} + v_t \tag{9}$$

where x_{gt} is the g^{th} regression variable at time t, β_g is the g^{th} regression parameter and $v_t = y_t - \sum_{g=1}^{G}\beta_g x_{gt}$ are the time series regression errors which are assumed to follow the SARIMA model in equation (8). Equations (8) and (9) taken together define the Reg-SARIMA model for the DPD series which can be written as a single equation as:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D(y_t-\sum_{g=1}^{G}\beta_g x_{gt})=c+\theta_q(B)\Theta_Q(B^s)\varepsilon_t, \quad \textbf{(10)}$$

The model in (9) implies that the regression effects are first subtracted from the time series y_t, which results in a series v_t with a zero mean. The series v_t is then differenced to get a stationary series. Let this series be denoted by ψ_t, then ψ_t follows a stationary Reg-SARMA model given by:

$$\phi_p(B)\Phi_P(B^s)\psi_t=c+\theta_q(B)\Theta_Q(B^s)\varepsilon_t. \quad \textbf{(11)}$$

The Reg-SARIMA model can also be written as:

$$(1-B)^d(1-B^s)^D y_t=\sum_{g=1}^{G}\beta_g(1-B)^d(1-B^s)^D x_{gt}+\psi_t, \quad \textbf{(12)}$$

where ψ_t follows the stationary Reg-SARMA model.

In order to capture the day of the week effect dummy variables are introduced and defined as follows:

$$\mathrm{D}_r=\begin{cases}1, & \text{if } r=\text{Tuesday},\ldots,\text{Sunday}\\ 0, & \text{otherwise}\end{cases} \quad \textbf{(13)}$$

The day of the week effect is represented by D_r which represents the significant daily variability of daily peak electricity demand. The index r takes values in the interval [2,7] representing days in the week except for Monday which represents the base period ($r=2$ for Tuesday, $r=3$ for Wednesday,…., $r=7$ for Sunday). The use of the base period is done to avoid the problem of multicollinearity which will affect the stability of the regression coefficients (Mirasgedis *et al.*, 2006). The daily peak demand decreases during holidays. The day before and after a holiday has an effect on demand. To take into account the effects of holidays the following dummy variables H_h, H_{h-1} and H_{h+1} are introduced. H_h, H_{h-1} and H_{h+1} are dummy variables representing holiday, day before and after a holiday respectively. South African holidays are: New years day (NY), Human rights day (HR), Good Friday (GF) (Easter holiday), Family day (F) (Easter holiday), Freedom (FD) day, Workers (W) day, Youth (Y) day, National women's (NW) day, Heritage (H) day, Day of reconciliation (DR), Christmas (C) day and Day of goodwill (DG). If a holiday falls on a weekend the following Monday is declared a public holiday. School holidays were not considered in this study. In this paper all holidays are equally weighted.

$$\mathrm{H}_h=\begin{cases}1, & \text{if day } h \text{ is a holiday}\\ 0, & \text{otherwise}\end{cases} \quad \textbf{(14)}$$

$$\mathrm{H}_{h-1} = \begin{cases} 1, & \text{if day } h-1 \text{ is a day before a holiday} \\ 0, & \text{otherwise} \end{cases} \quad \textbf{(15)}$$

$$\mathrm{H}_{h+1} = \begin{cases} 1, & \text{if day } h+1 \text{ is a day after a holiday} \\ 0, & \text{otherwise} \end{cases} \quad \textbf{(16)}$$

To take into account the monthly seasonality effect a dummy variable M_l is introduced.

$$\mathrm{M}_l = \begin{cases} 1, & \text{if } l = \text{February},\ldots, \text{December} \\ 0, & \text{otherwise} \end{cases} \quad \textbf{(17)}$$

The monthly seasonality effect is represented by M_l, where l represents the months February up to December with January as the base month. The index l takes values in the interval [2,12] representing months in a year except for January which represents the base period ($l = 2$ for February, $l = 3$ for March,…., $l = 12$ for December).

The Reg-SARIMA-GARCH model is a regression seasonal ARIMA model with error terms following a GARCH process. The model can be written as:

$$\begin{aligned} &\phi_p(B)\Phi_P(B^s)\psi_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \\ &\varepsilon_t = z_t\sigma_t, \\ &z_t \sim i.i.d \text{ with } E(z_t) = 0, \mathrm{Var}(z_t) = 1 \\ &\sigma_t^2 = \gamma w_t^{|} \\ &\text{where} \\ &w_t = (1, \varepsilon_{t-1}^2,\ldots,\varepsilon_{t-q}^2, \sigma_{t-1}^2,\ldots,\sigma_{t-p}^2), \\ &\gamma = (\alpha_0, \alpha_1,\ldots,\alpha_q, \beta_1,\ldots,\beta_p) \\ &\text{and } \psi_t = (1-B)^d(1-B^s)^D y_t - \sum_{g=1}^{G} \beta_g (1-B)^d (1-B^s)^D x_{gt} \end{aligned} \quad \textbf{(18)}$$

A3 The Piecewise Linear regression Model

The piecewise linear regression model used for comparative analysis with the models developed in this paper is shown in equation (19)

$$y_t = \beta_0 + \beta_1 \mathrm{t} + \beta_2(T_{pt} - t_w)x_{1t} + \beta_3(T_{pt} - t_s)x_{2t} + \sum_{r=2}^{7} \alpha_r \mathrm{D}_r + \sum_{l=2}^{12} \tau_l \mathrm{M}_l + \mu \mathrm{H}_h + \delta \mathrm{H}_{h-1} + \lambda \mathrm{H}_{h+1} + R_t \quad \textbf{(19)}$$

where T_{pt} represents peak temperature (in degrees Celsius). The peak temperature is the

temperature recorded at the hour of peak demand on day t, y_t denotes daily peak demand (in megawatts) observed on day t, t_w temperature to identify where the winter sensitive portion of demand join the non-weather sensitive demand component, t_s temperature to identify where the summer sensitive portion of demand join the non-weather sensitive demand component, β_0 represents the mean daily peak demand observed in the non-weather sensitive period $(t_w \le T_{pt} \le t_s)$. It should be noted that daily peak demand during non-weather sensitive days does not depend on temperature (T_{pt}) .The variable t represents the trend component, H_h, H_{h-1} and H_{h+1} are dummy variables representing holiday, day before and after a holiday respectively. The day of the week effect is represented by D_r, where r represents the days Tuesday up to Sunday with Monday as the base period. The monthly effect is represented by M_l, where l represents the months February up to December with January as the base month. $R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + \phi_5 R_{t-5} + \phi_7 R_{t-7} + \varepsilon_t$, where R_t is a stochastic disturbance term and ε_t is the innovation in the disturbance term.

$$x_{1t} = \begin{cases} 1, & if \quad T_{pt} - t_w < 0 \\ 0, & otherwise \end{cases} \quad \text{and}$$

$$x_{2t} = \begin{cases} 1, & if \quad T_{pt} - t_s > 0 \\ 0, & otherwise \end{cases}$$

References

Aknouche A. and Bentarzi M. 2008. On the existence of higher-order moments of periodic GARCH models. Statistics and Probability Letters 78, 3262 – 3268.

Amaral L.F., Souza R.C. and Stevenson M. 2008. A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. International Journal of Forecasting 24, 603-615.

Amin-Naseri M.R., Soroush A.R. 2008. Combined use of unsupervised and supervised learning for daily peak load forecasting. Energy Conversion and Management 49, 1302-1308.

Berndt E. K., Hall B. H., Hall R. E. and Hausman J. A. 1974. Estimation and Inference in Nonlinear Structural Models. Annals of Economic and Social Measurement 4, 653-665.

Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307-327.

Carpinteiro O., Reis A. and Silva A. 2004. A hierarchical neural model in short-term load Forecasting. Applied Soft Computing 4, 405-412.

Doornik J.A. and Ooms M. 2008. Multimodality in GARCH regression models. International Journal of Forecasting 24, 432–448.

Engle R.F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom Inflation. Econometrica 50, 987-1008.

Feinberg E. and Genethliou D. 2005. Load Forecasting. In J. Chow, F. Wu and J. Momoh (Eds.). Applied Mathematics for Restructured Electric Power Systems: Optimization, Control and Computational Intelligence. (pp. 269-285). Springer.

Friedman J.H. 1991. Multivariate adaptive regression splines. The Annals of Statistics 19(1), 1-141.

Ghahramani M. and Thavaneswaran A. 2008. A note on GARCH model identification. Computers and Mathematics with Applications 55, 2469–2475.

Gosh S. 2008. Univariate time-series forecasting of monthly peak demand of electricity in northern India. International Journal of Indian Culture and Business Management. 1, 466-474.

Goia A., May C. and Fusai G. 2010. Functional clustering and linear regression for peak load forecasting. International Journal of Forecasting. Article in press.

Hahn H., Meyer-Nieberg S. and Pickl S. 2009. Electric load forecasting methods: Tools for decision making. European Journal of Operational Research 199, 902-907.

Horv´ath L., Kokoszka P. and Zitikis R. 2008. Distributional analysis of empirical volatility in GARCH processes. Journal of Statistical Planning and Inference 138, 3578 – 3589.

He Z. and Maheu J.M. 2010. Real time detection of structural breaks in GARCH models. Computational Statistics and Data Analysis, Article in Press.

Hekkenberg M., Benders R.M.J., Moll H.C. and Schoot Uiterkamp A.J.M. 2009. Indications for a changing electricity demand pattern: The temperature dependence of electricity demand in the Netherlands. Energy Policy 37, 1542-1551.

Ismail Z., Yahya A. and Mahpol K.A. 2009. Forecasting peak load electricity demand using statistics and rule based approach. American Journal of Applied Sciences. 6 (8) 1618-625.

Kim Y.S., Rachev S.T., Bianchi M.L. and Fabozzi F. J. 2010. Tempered stable and tempered infinitely divisible GARCH models. Journal of Banking & Finance, Article in Press.

Mirasgedis S., Sarafidis Y., Georgopoulou E. , Lalasa D.P., Moschovits M., Karagiannis F. and Papakonstantinou D. 2006. Models for mid-term electricity demand forecasting incorporating weather influences. Energy 31, 208–227.

Mulera N. and Yohaib V.J. 2008. Robust estimates for GARCH models. Journal of Statistical Planning and Inference 138, 2918 – 2940.

Munoz A., Sanchez-Ubeda E.F., Cruz A. and Marin J. 2010. Short-term forecasting in power sytems: a guided tour. Energy Systems 2, 129-160.

Nelson D.B. and Cao C. Q. 1992. Inequality Constraints in the Univariate GARCH Model. Journal of Business & Economic Statistics 10 (2), 229-235.

Nelson D.B. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 59, 347–370.

Ramanathan R., Engle R., Granger C.W.J, Vahid-Araghi F. and Brace C. 1997. Short-run forecasts of electricity loads and peaks. International Journal of Forecasting 13, 161-174.

Sigauke C. and Chikobvu D. 2010. Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach. ORiON, 26(2), 97-111.

Soares L.J. and Medeiros M.C. 2008. Modelling and Forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. International Journal of Forecasting 24, 630-644.

Soares L. and Souza L. 2006. Forecasting electricity demand using generalized long memory. International Journal of Forecasting 22, 17-28.

Taylor J.W. 2008 An evaluation of methods for very short-term load forecasting using minute-by-minute British data. International Journal of Forecasting 24, 645-658.

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

Truong N-V., Wang L. and Wong P.K.C. 2008. Modelling and short-term forecasting of daily peak power demand in Victoria using two-dimensional wavelet based SDP models. Electrical Power and Energy Systems. 30, 511-518.