



Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach

Caston Sigauke*

Delson Chikobvu†

Received: 5 October 2010; Revised: 8 November 2010; Accepted: 11 November 2010

Abstract

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 enable effective load shifting between transmission substations, scheduling of startup times of peak stations, load flow analysis and power system security studies. A *multivariate adaptive regression splines* (MARS) modelling approach towards daily peak electricity load forecasting in South Africa is presented in this paper for the period 2000 to 2009. MARS is a non-parametric multivariate regression method which is used in high-dimensional problems with complex model structures, such as nonlinearities, interactions and missing data, in a straight forward manner and produces results which may easily be explained to management. The models developed in this paper consist of components that represent calendar and meteorological data. The performances of the models are evaluated by comparing them to a piecewise linear regression model. The results from the study show that the MARS models achieve better forecast accuracy.

Key words: MARS, temperature, peak electricity demand forecasting, piecewise linear regression.

1 Introduction

One of the most weather-sensitive sectors of any economy is the energy sector. In this sector accurate prediction of daily peak electricity demand is very important. It provides short term forecasts which are required for dispatching and economic grid management of electric energy [1, 2, 3, 8, 16, 19, 21, 22]. The most important weather factors which affect *daily peak demand* (DPD) is temperature. Changing weather conditions represent the major source of variation in peak demand forecasting and the inclusion of temperature has a significant effect due to the fact that in winter heating systems are used, whilst in summer

*Corresponding author: Department of Statistics and Operations Research, School of Mathematical and Computer Sciences, University of Limpopo, Turfloop Campus, Private Bag X1106, Sovenga 0727, South Africa, email: sigaukec@ul.ac.za

†Department of Mathematical Statistics and Actuarial Science, University of the Free State, P.O. Box 339, Bloemfontein, 9300, South Africa.

air conditioning appliances are used [6, 10, 11, 13, 14, 15, 17, 18]. Other weather factors include relative humidity, wind speed and cloud cover. Electricity demand forecasting has received extensive attention in the literature using various techniques ranging from classical time series methods and neural networks to regression methods. In this paper a *multivariate adaptive regression splines* (MARS) model is developed and used to predict daily peak electricity demand for South Africa. An updated review of different forecasting methods may be found in [7, 12].

The remainder of the paper is organised as follows. In Section 2 the data are described and a preliminary data analysis is carried out. The piecewise linear regression and the MARS models are presented in Section 3. A discussion of the results is presented in Section 4, and the paper closes in Section 5.

2 Definitions and data

The data considered in this paper are on *net energy sent out* (NESO) in response to some demand for 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. The data are for the period 2000 to 2009.

This definition of electrical demand has its weaknesses. Electrical demand is bounded by the power plants' capacity 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. 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 people, companies, *etc.* who are willing (or unwilling) and able (or unable) 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 for electrical demand.

The *daily peak demand* (DPD) is the maximum hourly demand in a 24-hour period. Aggregated DPD data were used for the industrial, commercial and domestic sectors of South Africa. Historical data on temperature were also collected from 22 meteorological stations from all the provinces of the country. The data were aggregated to obtain average daily, maximum and minimum temperatures for the entire country.

The time series plot of DPD in Figure 1 shows a positive linear trend and a strong seasonal fluctuation. The trend is mainly due to economic development of the country. Figures 2 and 3 shows daily and monthly index plots respectively. The basis for each index is 100. The seasonal peak is in July, which is a winter month. There is another small summer peak in October. The daily index plot shows that demand for electricity during week days is above the average consumption and decreases significantly on Saturdays and Sundays. A better representation of the relationship between DPD and temperature is shown in Figure 4. The peak temperature is the temperature recorded during the hour of peak demand on day t . The relationship is nonlinear. The demand for electricity is highly

sensitive to temperature fluctuations in winter and less sensitive in summer. DPD increases sharply as temperature decreases. The non-linear relationship between temperature and DPD calls for the derivation of two functions: one for cooling degree-days and the other for heating degree-days. Cooling degree-days (CDD_t) and heating degree-days (HDD_t) are estimated on the basis of the two linear functions

$$CDD_t = \max\{T_t - T_{\text{ref}}, 0\}$$

and

$$HDD_t = \max\{T_{\text{ref}} - T_t, 0\},$$

respectively, as defined in [12], where T_{ref} represents the temperature which separates the winter and summer periods and where T_t represents the peak temperature on day t . The reference temperature (T_{ref}) has been selected to be equal to 20.5°C ; this appears to be the temperature at which the minimum demand for electricity occurs. Above this temperature, electricity demand tends to rise slightly and below this temperature electricity demand increases significantly.

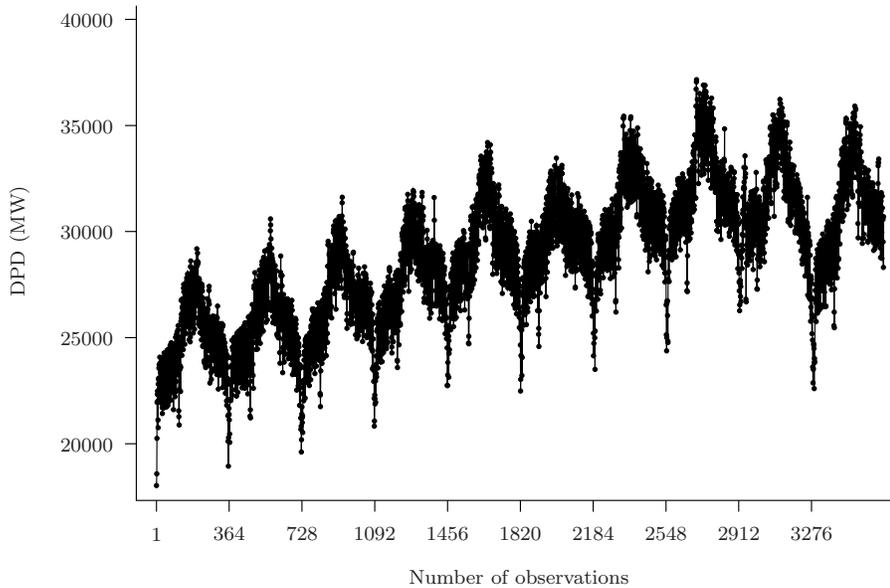


Figure 1: Time series plot of daily peak electricity demand for the period 1/1/2000 to 14/12/2009.

3 The models

A piecewise linear regression model and a MARS model are presented in this section. These models are used later for out-of-sample predictions of DPD. In both models DPD is taken as the dependent variable.

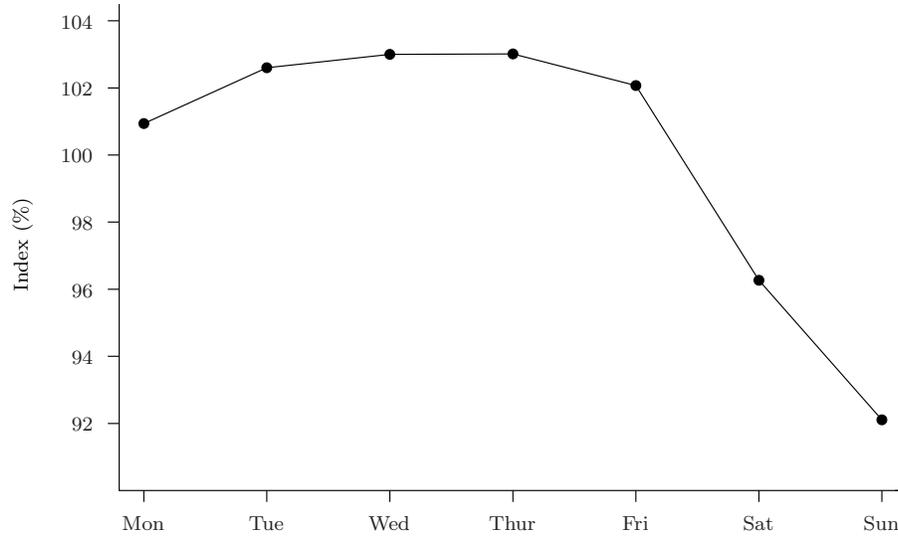


Figure 2: Daily index plots.

3.1 The piecewise linear regression model

Regression-based methods have been used extensively in load demand forecasting [4, 10, 20, 22]. These methods range from simple linear to multivariate linear regression models and work very well when the relationship between the dependent variable and the predictor variables is linear. They are usually fast, reliable and easy to implement with relatively robust solutions.

However, the relationship between electricity demand and temperature is nonlinear, as shown in Figure 4. This calls for the use of a multivariate linear regression model with three piecewise linear regression functions representing the winter, non-weather and summer-sensitive components.

The piecewise linear regression model used in this paper may be written as

$$z_t = \beta_0 + \beta_1 t + \beta_2(x_{pt} - t_w)x_{1t} + \beta_3(x_{pt} - t_s)x_{2t} + \sum_{d=1}^7 \alpha_d D_{dt} + \sum_{j=1}^{12} \tau_j M_{jt} + \mu H_t + \delta H_{t-1} + \lambda H_{t+1} + R_t, \quad (1)$$

where x_{pt} represents peak temperature (in degrees Celsius). The peak temperature is the temperature recorded at the hour of peak demand on day t , z_t denotes the DPD (in megawatts) observed on day t , t_w denotes the temperature where the winter-sensitive portion of demand joins the non-weather-sensitive demand component, t_s denotes the temperature where the summer-sensitive portion of demand joins the non-weather-sensitive demand component, and β_0 represents the mean DPD observed during the non-weather-sensitive period ($t_w \leq x_{pt} \leq t_s$). It should be noted that DPD during non-weather-sensitive days does not depend on temperature (x_{pt}). The variable t represents the trend component, H_t , H_{t-1} and H_{t+1} are dummy variables representing holiday, day before and day after a holiday, respectively. The day-of-the-week effect is represented by D_{dt} , where d

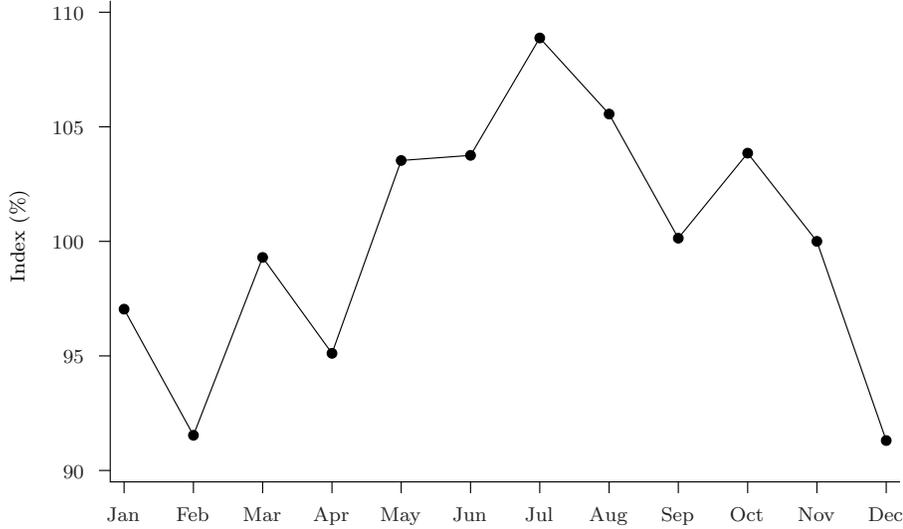


Figure 3: Monthly index plots.

denotes the days Tuesday up to Sunday with Monday as the base period. D_{dt} equals 1 if day d is found in observation t , or zero otherwise, with $t = 1, 2, \dots, n$. The monthly effect is denoted by M_{jt} , where j represents the months February up to December with January as the base month. M_{jt} equals 1 if month j is found in observation t , or zero otherwise, with $t = 1, 2, \dots, n$. Furthermore, $R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + \phi_5 R_{t-5} + \phi_7 R_{t-7} + \varepsilon_t$, where R_t is a stochastic disturbance term and ε_t is the innovation in the disturbance with

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

and

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

The model in (1) accounts for any residual correlation that may occur as a result of the week-to-week variation in peak demand and also for the day-to-day variation. The model is based on the following theoretical assumptions:

1. Peak demand on day t is highly correlated with peak demand on day $t + 1$.
2. There may be significant correlation between demand 2 days, 5 days and/or 7 days apart.

The derivations of the equations of the three demand-temperature lines are presented in the appendix at the end of the paper.

3.2 The multivariate adaptive regression splines (MARS) model

MARS is a non-parametric multivariate regression method which was developed in [9] and has been used to solve high-dimensional problems with complex model structures, such

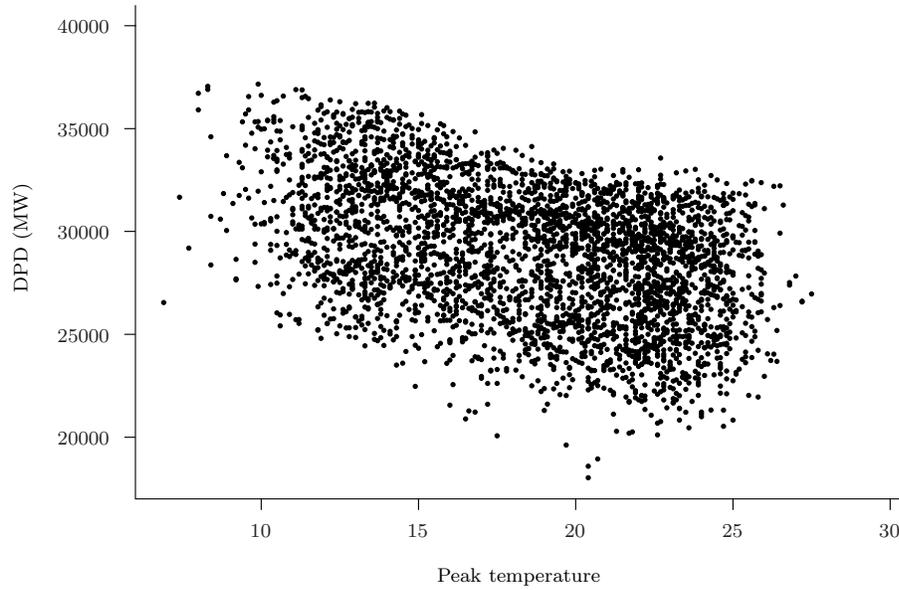


Figure 4: Scatter plot of daily peak demand against peak temperature (in °C).

as nonlinearities, interactions, multicollinearity and missing values [3, 6, 17, 27]. The method does not make any assumptions about the functional relationship between the response variable and the predictor variables. The MARS modelling approach overcomes the major drawbacks of using artificial neural networks which have long training processes, interpretive difficulties and an inability to determine the relative importance of potential input variables. In the MARS paradigm the modelling space is divided into subregions and then fits in each subregion simple linear regression models. The model building process occurs in two steps: the forward stepwise algorithm and the backward stepwise algorithm. In the forward stepwise step the MARS algorithm constructs a large number of basis functions which over-fits the data. In the backward stepwise step basis functions are deleted in order of least contribution using the *generalised cross validation* (GCV) criterion [5]. The general MARS model may be written as

$$f(x) = \alpha_0 + \sum_{m=1}^M \alpha_m B_m(x), \quad (2)$$

where

$$B_m(x) = \prod_{k=1}^{K_m} [s_{km} (x_{v(k,m)} - t_{km})]$$

is a basis function, α_0 and α_m are parameters, M is the number of basis functions, K_m is the number of knots, s_{km} takes on values of either 1 or -1 indicating the right or left sense of the associated step function, $v(k, m)$ is the label of the independent variable and t_{km} indicates the knot location. The MARS algorithm selects variables and values of those variables for knots of the hinge functions.

The GCV criterion is a measure of the goodness of fit which takes into account the residual error and the model complexity. In its simplest form the GCV criterion may be written

as

$$\text{GCV}(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2}, \quad (3)$$

where N is the sample size, and $C(M)$ is the cost-penalty measure of a model containing M basis functions. The numerator measures the lack of fit on the M basis function model $\hat{f}_M(x_i)$ and the denominator represents the penalty for the model complexity $C(M)$. As in [9], the complexity cost function may be written as

$$C(M) = \text{trace}(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) + 1,$$

where \mathbf{B} is the $M \times N$ data matrix of the M (nonconstant) basis functions ($B_{ij} = B_i(x_j)$). The best model is one with the lowest GCV criterion value.

The three general MARS models used in this paper are

$$\begin{aligned} z_t = & a_0 + c_1 \max\{0, x_{pt} - t_w\} + c_2 \max\{0, x_{pt} - t_s\} \\ & + \sum_{d=1}^7 \alpha_d D_{dt} + \sum_{j=1}^{12} \tau_j M_{jt} + \mu H_t + \delta H_{t-1} + \lambda H_{t+1} + \varepsilon_t, \end{aligned} \quad (4)$$

$$\begin{aligned} z_t = & \omega_0 + c_3 \max\{0, T_t - T_{\text{ref}}\} + c_4 \max\{0, T_{\text{ref}} - T_t\} \\ & + \sum_{d=1}^7 \alpha_d D_{dt} + \sum_{j=1}^{12} \tau_j M_{jt} + \mu H_t + \delta H_{t-1} + \lambda H_{t+1} + \varepsilon_t \end{aligned} \quad (5)$$

and

$$z_t = \beta_0 + c_5 \max\{0, T_t - T_{\text{ref}}\} + c_6 \max\{0, T_{\text{ref}} - T_t\} + \varepsilon_t, \quad (6)$$

where a_0, ω_0, β_0 and c_1, \dots, c_6 are constants, and where the parameters have meanings as declared in §3.1.

4 Results and discussion

The forecast results obtained via the piecewise linear regression model and the MARS models are presented in this section.

4.1 Piecewise linear regression model

Three different piecewise linear functions for modeling the peak demand (z_t) and peak temperature (x_{pt}) relationship were proposed in (1). The values of t_w and t_s are taken as 17.5°C and 24°C, respectively. These values were determined from a visual inspection of the graph in Figure 4. Piecewise linear regression models were fitted for various reference temperatures in the interval 17°C – 24°C, without any significant improvements in the

results. The reference temperature (T_{ref}) has been selected as 20.5°C , as mentioned before. The resulting piecewise linear function is

$$z_t = \beta_0 + \beta_1 t + \beta_2(x_{pt} - 17.5)x_{1t} + \beta_3(x_{pt} - 24)x_{2t} + \sum_{d=1}^7 \alpha_d D_{dt} + \sum_{j=1}^{12} \tau_j M_{jt} + \mu H_t + \delta H_{t-1} + \lambda H_{t+1} + R_t, \quad (7)$$

where $R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + \phi_5 R_{t-5} + \phi_7 R_{t-7} + \varepsilon_t$, as already stated.

Out of the 3636 data points, which span the period 1 January 2000 to 14 December 2009, the 3592 data points (representing the period 1 January 2000 to 31 October 2009) were used for developing the models, while the remaining 44 observations were used for validation purposes. Table 1 shows the parameter estimates of the best piecewise linear regression model,

$$z_t = 26274 + 1.9t - 232.8(x_{pt} - 17.5)x_{1t} + 21.0(x_{pt} - 24)x_{2t} - 881.7D_{5t} - 2279.0D_{6t} - 2551.6D_{7t} - 1813.4H_t - 810.9H_{t-1} - 248.7H_{t+1} + 0.819z_{t-1} + 0.013z_{t-2} + 0.056z_{t-5} + 0.094z_{t-7}. \quad (8)$$

The coefficient of t is positive, showing a positive linear trend. The dummy variable x_{1t} is negative, showing that if the peak temperature decreases by one degree from 17.5°C , electricity demand increases by 232.8 MW. The coefficient of x_{2t} in (1) is positive, showing that if the temperature increases by one degree from 24°C , electricity demand increases by 21 MW. This shows that electricity demand is more sensitive to winter conditions than to summer conditions. All the coefficients of the dummy variables representing *Friday*, *Saturday*, *Sunday*, *holiday*, *day before holiday* and *day after holiday* are negative. This shows that there is a decrease in demand during these periods. Of the three days of the week, the largest decrease occurs on a Sunday. During holidays, demand for electricity decreases significantly compared to a day before and after a holiday. The smallest decrease is experienced on days after holidays.

Par	c	H_t	H_{t-1}	H_{t+1}	Friday	Saturday	Sunday
Coef	26274	-1813.4	-810.9	-248.7	-881.7	-2279	-2551.6
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Par	t	x_{1t}	x_{2t}	ϕ_1	ϕ_2	ϕ_5	ϕ_7
Coef	1.9	-232.8	21	0.819	0.013	0.056	0.094
	(0.000)	(0.000)	(-0.4963)	(0.000)	(-0.4679)	(0.000)	(0.000)

Table 1: Parameter estimates for the piecewise linear regression model.

4.2 The MARS models

The values of the various parameters in the MARS models in (4)–(6) are estimated from the data in this section. The forecasting results obtained via the models are also interpreted.

4.2.1 Model 1

The model in (1) is a simple MARS model which was used to determine the reference temperature separating the winter periods from the summer periods of the DPD-temperature relationship. The DPD is the dependent variable in this model with the peak temperature as the regressor variable. The best MARS model achieved a GCV value of 8.66744×10^6 and the reference temperature was found to be 20.9°C . The complete model may be written as

$$\begin{aligned} z_t &= \beta_0 + c_5 \max\{0, x_{pt} - T_{\text{ref}}\} + c_6 \max\{0, T_{\text{ref}} - x_{pt}\}. \\ &= 27833.6 - 125.423 \max\{0, x_{pt} - 20.9\} + 384.209 \max\{0, 20.9 - x_{pt}\}. \end{aligned} \quad (9)$$

A plot of the value of z_t as a function of x_{pt} may be found in Figure 5. If temperature decreases by a degree from 20.9°C , the DPD increases by 384.209MW . Similarly, an increase by one degree above 20.9°C results in a DPD decrease of 125.423MW . This shows that the DPD is more sensitive to low temperatures. This model was used to determine the number of heating degree days and also the number of cooling degree days.

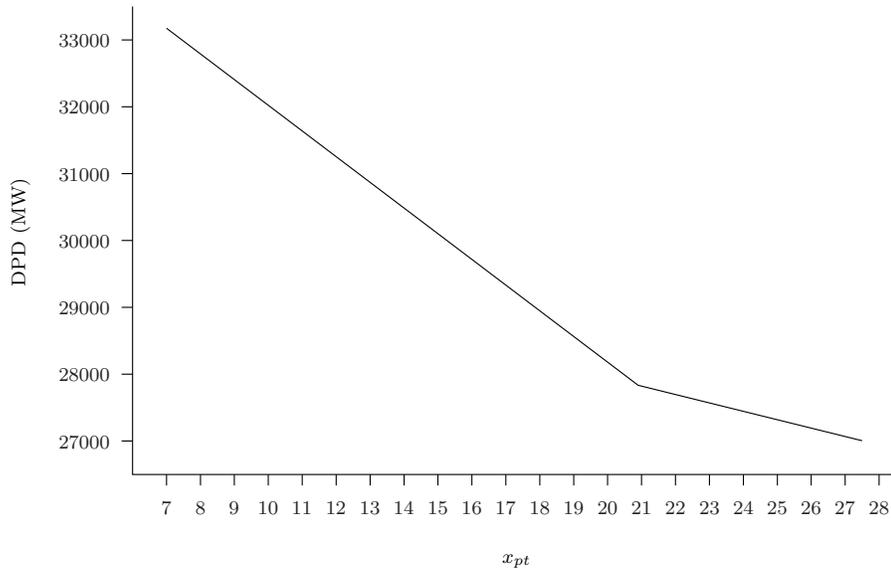


Figure 5: DPD as a function of temperature according to MARS model 1.

4.2.2 Model 2

Out of the 24 predictor variables, the MARS algorithm selected eight variables as the most important. These variables are shown in Table 2 in order of their importance. The piecewise linear GCV value was 9.02477×10^5 .

Variable	Importance	GCV value
t	100.000	7.27027×10^6
x_{pt}	58.25802	3.06371×10^6
D_{7t}	34.58035	1.66394×10^6
D_{6t}	31.78196	1.54569×10^6
H_t	18.85743	1.12892×10^6
H_{t-1}	14.80247	1.04201×10^6
D_{5t}	12.35992	9.99758×10^5
H_{t+1}	8.76643	9.51416×10^5

Table 2: Important predictor variables according to the MARS models.

The final model is

$$\begin{aligned}
\text{DPD} = & 24676.4 - 1.74 \max\{0, t - 2735\} - 2.99 \max\{0, 2735 - t\} \\
& - 235.38 \max\{0, x_{pt} - 19.2\} + 411 \max\{0, 19.2 - x_{pt}\} + 2598.11(D_{7t} = 0) \\
& - 2367.12(D_{6t} = 1) + 2722.32(H_t = 0) + 924.86(D_{5t} = 0) \\
& + 500.43(H_{t-1} = 0) + 1280.82(H_{t+1} = 0)(H_{t-1} = 0). \tag{10}
\end{aligned}$$

The coefficient of basis function 1 is negative, meaning that if the trend is above 2375, electricity demand decreases at a rate of 1.74 MW and when its below this knot, it decreases at a rate of 2.99MW. The trend component shows that the DPD is increasing at a decreasing rate. The coefficient of basis function 3 is negative, implying that if the peak temperature increases by 1°C from 19.2°C, the DPD decreases by 235.381MW and if the peak temperature decreases by 1°C below this knot, the DPD increases by 411MW. The coefficient for basis function 5 is positive, meaning if the day of the week is not a Sunday, then the DPD increases by 2598.11MW, but if the day is Saturday, there is a decrease in the DPD of 2367.12MW. The DPD increases by 2722.32MW if day t is not a holiday and increases by 500.434MW if it is not a day before a holiday. There is one bivariate interaction between a day before and after a holiday. If day t is not a day before or after a holiday, the DPD increases by 1280.82MW. If day t is not a Friday, there is an increase in DPD of 924.862MW.

4.2.3 Model 3

The third model is a MARS model for the *Average Daily Energy Sent Out* (ADESO) with *average daily temperature* (ADT) as the predictor variable. The model identifies the winter-sensitive, weather-neutral and summer-sensitive periods. The resulting model is

$$\text{ADESO} = 564863 + 7332.94 \max\{0, 22 - \text{ADT}\} + 3714.8 \max\{0, \text{ADT} - 16\}. \tag{11}$$

The piecewise linear GCV value was 3.82422×10^9 . The value achieved by the model is displayed in the graphical plot of ADESO against ADT, shown in Figure 6.

If the average daily temperature is less than or equal to 16°C, we use

$$\text{ADESO} = 564863 + 7333 \max\{0, 22 - \text{ADT}\}. \tag{12}$$

That is, if the temperature decreases by 1°C (*e.g.* from 16°C to 15°C), electricity demand increases by 7333MW, which is about a 1.2% increase.

If average daily temperature is greater than or equal to 22°C, we use

$$\text{ADESO} = 564863 + 3715 \max\{0, \text{ADT} - 16\}. \quad (13)$$

If the temperature increases by 1°C (*e.g.* from 22°C to 23°C), electricity demand increases by 3715MW, which is about a 0.6% increase.

For an average daily temperature between 16°C and 22°C, we use the full model

$$\text{ADESO} = 564863 + 7333 \max\{0, 22 - \text{ADT}\} + 3715 \max\{0, \text{ADT} - 16\}. \quad (14)$$

If the temperature decreases by 1°C in the range 16°C–22°C (*e.g.* from 22°C to 21°C), electricity demand increases by 3618MW, which is about a 0.6% increase. The MARS plot

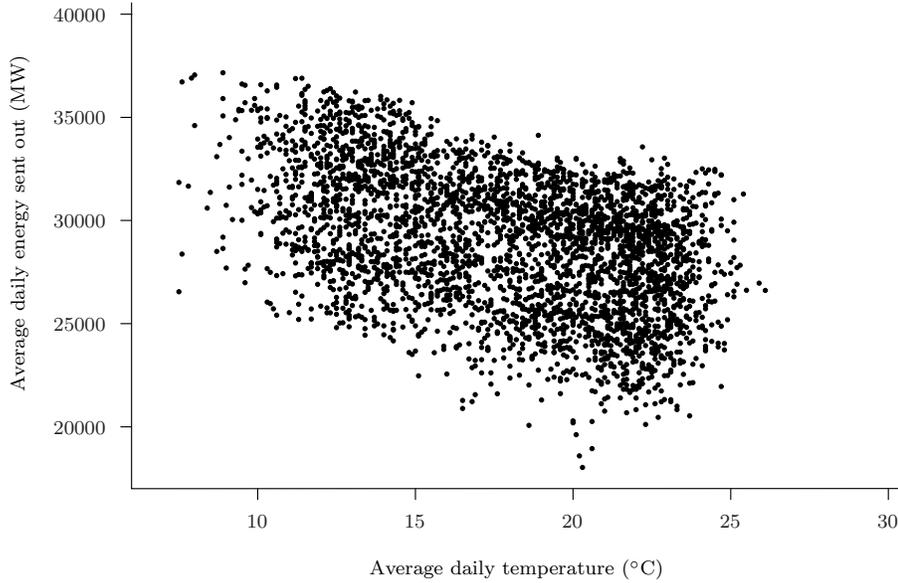


Figure 6: Scatter plot of average daily energy sent out against average daily temperature.

is shown in Figure 7. The modelling space is divided into three subspaces, separated by two knots at 16°C and 22°C, as shown in Figure 7.

4.3 Evaluating the goodness of fit of the models

The *root mean squared error* (RMSE) was used to evaluate the goodness of fit achieved by the piecewise regression model and the MARS model for peak load demand forecasting in the out-of-sample predictions for the period 1 November to 14 December 2009. As mentioned, the training period was 1 January 2000 to 31 October 2009. The RMSE was calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (z_{at} - z_{ft})^2}{n}}, \quad (15)$$

where n is the number of out-of-sample forecast data points and $z_{at} - z_{ft}$ represents the forecast errors. The terms z_{at} and z_{ft} are the actual DPD and its future forecast,

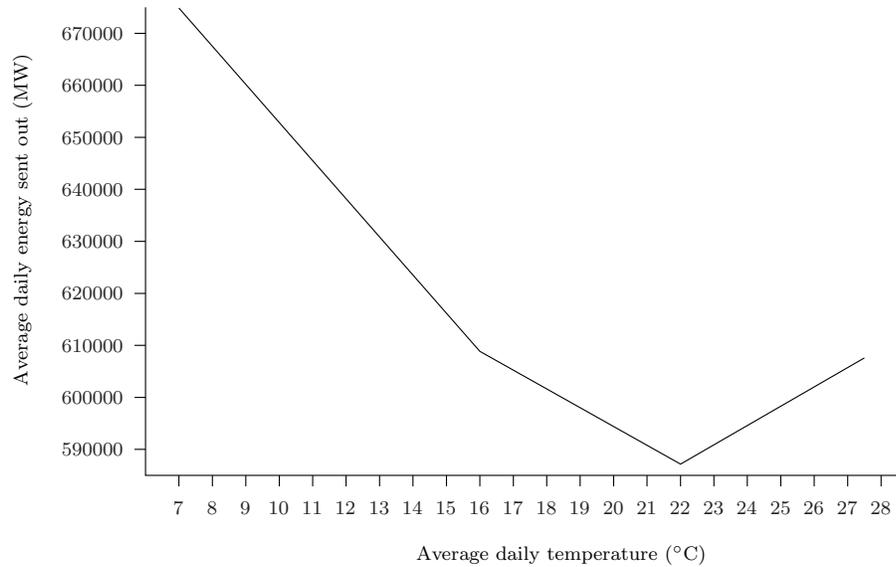


Figure 7: *MARS plot of model 3.*

respectively. The goodness of fit results for the piecewise linear and MARS models are shown in Table 3. It may be seen from the table that the MARS models outperform the piecewise linear regression model convincingly.

Model	R-Squared Adjusted	RMSE Validation (Testing Period)
Piecewise Linear	0.91626	940.843
MARS Model 2	0.98096	446.013

Table 3: *Goodness of fit results for the piecewise and MARS models.*

The forecasts using MARS model 2, approximate 95% prediction intervals and the actual DPD values for the first seven days of November 2009 are given in Table 4. The actual peak demand falls within the prediction interval for all seven days. The MARS model seems to be useful for making short-term forecasts of daily peak demand.

5 Conclusions

A MARS model was developed for predicting daily electricity peak demand and the performance of the model was compared to that of a piecewise linear regression model. There were 3 636 data points, spanning the period 1 January 2000 to 14 December 2009. Of these, 3 592 data points were used for developing the models, while the remaining 44 observations were reserved for validation purposes. The MARS model outperformed the piecewise linear regression model convincingly and is easy to explain to management. The model is capable of clustering together categories of variables that have similar effects on the dependent variable.

Future research may include a sensitivity analysis with respect to daily and seasonal peak

electricity demand performed for each of the provinces of South Africa and the development of a two-stage stochastic integer recourse model with the objective of optimising electricity distribution. Another interesting area to investigate would be the development of a hybrid model integrating the MARS approach with neural network techniques and also the use of other adaptive techniques, such as *classification and regression trees* (CARTs), TreeNet and Random Forests.

Date	Day (t)	Forecast (z_{ft})	Prediction Interval Approx 95%	Actual DPD (z_{at})	Actual Peak Temp (x_{pt})
Sun, Nov 1	3593	29219.75	(28009.53, 30485.30)	29187	20.2
Mon, Nov 2	3594	31767.53	(30046.46, 33593.91)	31640	18.9
Tue, Nov 3	3595	31866.16	(29941.48, 33924.75)	31882	20.2
Wed, Nov 4	3596	31751.65	(29664.31, 33999.46)	31641	18.3
Thur, Nov 5	3597	31678.71	(29492.76, 34047.10)	31832	20.2
Frid, Nov 6	3598	30878.02	(28669.81, 33272.95)	30704	22.1
Sat, Nov 7	3599	29516.36	(27348.04, 31875.72)	29650	20.8

Table 4: Forecast and Actual DPD for the first seven days of November 2009.

Acknowledgments

The authors are grateful to Eskom for providing the data. The authors also thank the Department of Statistics and Operations Research, School of Mathematical and Computer Sciences, University of Limpopo, and the Department of Mathematical Statistics and Actuarial Science, University of the Free State for the use of their resources. Finally, the authors are indebted to numerous people who assisted in making constructive suggestions which have improved the quality of this paper.

References

- [1] AMARAL LF, SOUZA RC & STEVENSON M, 2008, *A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting*, International Journal of Forecasting, **24**, pp. 603–615.
- [2] AMIN-NASERI MR & SOROUGH AR, 2008, *Combined use of unsupervised and supervised learning for daily peak load forecasting*, Energy Conversion and Management, **49**, pp. 1302–1308.
- [3] CHEN L-A, 1997, *Multivariate regression splines*, Computational Statistics and Data Analysis, **26**, pp. 71–82.
- [4] CHOU S-M, LEE T-S, SHAO YE & CHEN I-F, 2004, *Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines*, Expert Systems with Applications, **27**, pp. 133–142.
- [5] CRAVEN P & WAHBA G, 1979, *Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation*, Numerische Mathematik, **31**, pp. 317–403.
- [6] DE GOOLJER JG & RAY BK, 2003, *Modeling vector nonlinear time series using POLYMARS*, Computational Statistics & Data Analysis, **42**, pp. 73–90.
- [7] FEINBERG E & GENETHLIOU D, 2005, *Load Forecasting*, pp. 269–285 in CHOW J, WU F & MOMOH J (EDS.), *Applied mathematics for restructured electric power systems optimization*, Control and Computational Intelligence, Springer, Berlin.

- [8] FRANCO G & SANSTAD AH, 2008, *Climate change and electricity demand in California*, Climatic Change (Supplement 1), **87**, pp. S139–S151.
- [9] FRIEDMAN JH, 1991, *Multivariate adaptive regression splines*, The Annals of Statistics, **19(1)**, pp. 1–141.
- [10] GHOSH S, 2008, *Univariate time-series forecasting of monthly peak demand of electricity in northern India*, International Journal of Indian Culture and Business Management, **1**, pp. 466–474.
- [11] GOIA A, MAY C & FUSAI G, 2010, *Functional clustering and linear regression for peak load forecasting*, International Journal of Forecasting, in press.
- [12] HAHN H, MEYER-NIEBERG S & PICKL S, 2009, *Electric load forecasting methods: Tools for decision making*, European Journal of Operational Research, **199**, pp. 902–907.
- [13] HEKKENBERG M, BENDERS RMJ, MOLL HC & SCHOOT AJM, 2009, *Indications for a changing electricity demand pattern: The temperature dependence of electricity demand in the Netherlands*, Energy Policy, **37**, pp. 1542–1551.
- [14] HEKKENBERG M, MOLL HC & SCHOOT AJM, *Dynamic temperature dependence patterns in future energy demand models in the context of climate change*, Energy, **34**, pp. 1797–1806.
- [15] MAKRIDAKIS S, WHEELWRIGHT S & HYNDMAN, 1998, *Forecasting methods and applications*, 3rd Edition, Wiley, New York (NY).
- [16] MIRASGEDIS S, SARAFIDIS Y, GEORGOPOULOU E, LALAS DP, MSCHOVITIS M, KARAGIANNIS F & PAKONSTANTINO D, 2006, *Models for mid-term electricity demand forecasting incorporating weather influences*, Energy, **31**, pp. 208–227.
- [17] MUNOZ J & FELICISIMO AM, 2004, *Comparison of statistical methods commonly used in predictive modeling*, Journal of Vegetation Science, **15**, pp. 285–292.
- [18] PILLI-SIHVOLA K, AATOLA P, OLLIKAINEN M & TUOMENVIRTA H, 2010, *Climate change and electricity consumption – Witnessing increasing or decreasing use and costs*, Energy Policy, **38(5)**, pp. 2409–2419.
- [19] PSILOGLOU BE, GIANNAKOPOULOS C, MAJITHIA S & PETRAKIS M, 2009, *Factors affecting electricity demand in Athens, Greece and London, UK: A comparative assessment*, Energy, **34**, pp. 1855–1863.
- [20] ISMAIL Z, JAMALUDDIN F & JAMALUDDIN F, 2008, *Time series regression model for forecasting Malaysian electricity load demand*, Asian Journal of Mathematics and Statistics, **1(3)**, pp. 139–149.
- [21] ISMAIL Z, YAHYA A & MAHPOL KA, 2009, *Forecasting peak load electricity demand using statistics and rule based approach*, American Journal of Applied Sciences, **6(7)**, pp. 1618–625.
- [22] RAMANATHAN R, ENGLE R, GRANGER CWJ, VAHID-ARAGHI F & BRACE C, 1997, *Short-run forecasts of electricity loads and peaks*, International Journal of Forecasting, **13**, pp. 161–174.
- [23] SOARES LJ & MEDEIROS MC, 2008, *Modelling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data*, International Journal of Forecasting, **24**, pp. 630–644.
- [24] TAYLOR J, DE MENEZES L & MCSHARRY P, 2006, *A comparison of univariate methods for forecasting electricity demand up to a day ahead*, International Journal of Forecasting, **22**, pp. 1–16.
- [25] TAYLOR JW, 2008, *An evaluation of methods for very short-term load forecasting using minute-by-minute British data*, International Journal of Forecasting, **24**, pp. 645–658.
- [26] TRUONG N-V, WANG L & WONG PKC, 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**, pp. 511–518.
- [27] TSEKOURAS GJ, DIALYNAS EN, HATZIARGYRIOU ND & KAVATZA S, 2007, *A non-linear multivariable regression model for midterm energy forecasting of power systems*, Electric Power Systems Research, **77**, pp. 1560–1568.

Appendix

This appendix contains a brief derivation of the equations of the three demand-temperature lines from

$$z_t = \beta_0 + \beta_1 t + \beta_2(x_{pt} - t_w)x_{1t} + \beta_3(x_{pt} - t_s)x_{2t} + \sum_{d=1}^7 \alpha_d D_{dt} + \sum_{m=1}^{12} \tau_m M_{mt} + \mu H_t + \delta H_{t-1} + \lambda H_{t+1} + R_t. \quad (16)$$

Let

$$\sum_{d=1}^7 \alpha_d D_{dt} = \sum_{m=1}^{12} \tau_m M_{mt} = \mu H_t = \delta H_{t-1} = \lambda H_{t+1} = 0. \quad (17)$$

Substituting (17) into (16) we obtain three equations for winter-sensitive, summer-sensitive and non-weather sensitive months.

For winter-sensitive months, *i.e.* $x_{pt} < t_w$, $x_{1t} = 1$, $x_{2t} = 0$, we obtain

$$\begin{aligned} E(z_t) &= \beta_0 + \beta_1(x_{pt} - t_w)(1) + \beta_2(x_{pt} - t_s)(0) \\ &= \beta_0 + \beta_1(x_{pt} - t_w) \\ &= (\beta_0 - \beta_1 t_w) + \beta_1 x_{pt}. \end{aligned} \quad (18)$$

For summer-sensitive months, *i.e.* $x_{pt} > t_s$, $x_{1t} = 0$, $x_{2t} = 1$, we obtain

$$\begin{aligned} E(z_t) &= \beta_0 + \beta_1(x_{pt} - t_w)(0) + \beta_2(x_{pt} - t_s)(1) \\ &= \beta_0 + \beta_2(x_{pt} - t_s) \\ &= (\beta_0 - \beta_2 t_s) + \beta_2 x_{pt}. \end{aligned} \quad (19)$$

Finally, for non-weather sensitive months, *i.e.* $t_w \leq x_{pt} \leq t_s$, $x_{1t} = x_{2t} = 0$, we obtain

$$E(z_t) = \beta_0, \quad (20)$$

where β_0 represents the mean daily peak demand observed during the non-weather sensitive period.

