AOSIS

Density forecasting for long-term electricity demand in South Africa using quantile regression

CrossMark
click for updates

Authors:
Paul Mokilane[1,2]
Jacky Galpin[2]
V.S. Sarma Yadavalli[3]
Provesh Debba[1,2]
Renee Koen[1]
Siphamandla Sibiya[1]

Affiliations:
[1]Council for Scientific and Industrial Research, Pretoria, South Africa

[2]School of Statistics and Actuarial Science, University of the Witwatersrand, South Africa

[3]Department of Industrial and Systems Engineering, School of Engineering, University of Pretoria, South Africa

Corresponding author:
Paul Mokilane,
PMokilane@csir.co.za

Background: This study involves forecasting electricity demand for long-term planning purposes. Long-term forecasts for hourly electricity demands from 2006 to 2023 are done with in-sample forecasts from 2006 to 2012 and out-of-sample forecasts from 2013 to 2023. Quantile regression (QR) is used to forecast hourly electricity demand at various percentiles. Three contributions of this study are (1) that QR is used to generate long-term forecasts of the full distribution per hour of electricity demand in South Africa; (2) variabilities in the forecasts are evaluated and uncertainties around the forecasts can be assessed as the full demand distribution is forecasted and (3) probabilities of exceedance can be calculated, such as the probability of future peak demand exceeding certain levels of demand. A case study, in which forecasted electricity demands over the long-term horizon were developed using South African electricity demand data, is discussed.

Aim: The aim of the study was: (1) to apply a quantile regression (QR) model to forecast hourly distribution of electricity demand in South Africa; (2) to investigate variabilities in the forecasts and evaluate uncertainties around point forecasts and (3) to determine whether the future peak electricity demands are likely to increase or decrease.

Setting: The study explored the probabilistic forecasting of electricity demand in South Africa.

Methods: The future hourly electricity demands were forecasted at 0.01, 0.02, 0.03, … , 0.99 quantiles of the distribution using QR, hence each hour of the day would have 99 forecasted future hourly demands, instead of forecasting just a single overall hourly demand as in the case of OLS.

Results: The findings are that the future distributions of hourly demands and peak daily demands would be more likely to shift towards lower demands over the years until 2023 and that QR gives accurate long-term point forecasts with the peak demands well forecasted.

Conclusion: QR gives forecasts at all percentiles of the distribution, allowing the potential variabilities in the forecasts to be evaluated by comparing the 50th percentile forecasts with the forecasts at other percentiles. Additional planning information, such as expected pattern shifts and probable peak values, could also be obtained from the forecasts produced by the QR model, while such information would not easily be obtained from other forecasting approaches. The forecasted electricity demand distribution closely matched the actual demand distribution between 2012 and 2015. Therefore, the forecasted demand distribution is expected to continue representing the actual demand distribution until 2023. Using a QR approach to obtain long-term forecasts of hourly load profile patterns is, therefore, recommended.

Introduction

Electricity load is the amount of electricity that balances the amount generated with that drawn from the grid. In the absence of black-outs, load-shedding and the availability of electricity generated from renewable electricity sources, the electricity load is equivalent to the electricity demand. Therefore, in this study, the hourly electricity demand is defined as the amount of electricity (load) in kW sent out every hour by Eskom to meet consumers' demand.

The 1996 census showed that only 57.6% of the South African households had access to electricity for lighting (Statistics South Africa 1998). The 2001 census showed that this percentage went up to 70.2% (Lehohla 2005). The 2007 Community Survey indicated that 80.1% of the South African households had access to electricity for lighting (Statistics South Africa 2008). The 2011 census showed that this percentage went up to 84.7% (Statistics South Africa 2011). These censuses and surveys indicate that the high percentages of new households that were connected to the electricity grid between 1996 and 2007 would imply that residential electricity demand would

be expected to have increased during the same period. The percentage of new households connected to the grid stabilised between 2007 and 2011 and, therefore, residential electricity demand would be expected to have stabilised during this period. The shrinking South African economic growth between 2007 and 2015 could have contributed to a decline in electricity demand:

> South Africa experienced an average growth rate of approximately 5% in real terms between 2004 and 2007. However, the period 2008 to 2012 only recorded average growth of just above 2%. (Statistics South Africa n.d.)

The penetration of other sources of electricity such as renewables for example solar and wind, could also have contributed to a decline in electricity demand from Eskom. In addition, because of the lack of capacity in the generation of electricity experienced by Eskom in 2007 (Inglesi & Pouris 2010), some companies and households had to find other sources of electricity, which would have resulted in a decline in electricity demand from Eskom. Unfortunately, the actual size of the electricity demand market is still unknown because of the unavailability of certain types of data, such as renewable energy and other forms of electricity generation. The combined effect of all these changes in the demography, economy and usage patterns can be investigated using historical patterns, but contribute to uncertainties when trying to forecast future electricity demand.

Uncertainties occur in estimation, prediction or in forecasting. When statisticians develop predictions (forecasts) for an uncertain future, they need to quantify the uncertainties around these for those that have to make decisions in the face of those uncertainties. Sigauke (2014) indicates that uncertainties in future electricity demand could emanate from increased technologies making use of electricity, population growth, general randomness in individual usage of electricity, seasonal effects, prevailing economic patterns, change in weather conditions, escalating costs, use of power saving electrical appliances and the growing sources of renewable energies. The inherent uncertainties in predictions imply that forecasts should ideally be probabilistic; in other words, they should take the form of probability distributions over future quantities or events (Gneiting & Katzfuss 2014). Probabilistic forecasts could take the form of quantiles, prediction intervals or density forecasts to quantify uncertainties in predictions. They are an essential ingredient of optimal decision-making (Gneiting & Katzfuss 2014). It is important to quantify the uncertainties around the demand forecasts for planning purposes, to avoid building unnecessary infrastructure and to ensure that future electricity demand is met. Tay and Wallis (2000) define density forecasts of the realisation of a random variable at some future time as estimates of the probability distribution of the possible future values of that variable.

Hong, Wilson and Xie (2014) argue that forecasting is by nature a stochastic problem, but that most of the utilities are still developing and using point forecasts. They state that it would be better to use probabilistic forecasts that provide estimates of the full distribution of the possible future values as a way of quantifying the uncertainties in the forecasts.

In the late 1880s, when lighting was the sole end use of electricity, the forecasting of electricity demand was straightforward (Hong & Shahidehpour 2015). Power generating companies would count the number of light bulbs they installed and planned to install and they would then roughly estimate the level of demand in the evening. As electric appliances such as electric irons, radios, television sets, geysers, stoves and washing machines were invented and commonly used in many households, the complexity of forecasting electricity demand grew. The penetration of air conditioners into homes and offices to regulate temperature within comfort zones, and industrial uses of electricity became important drivers of electricity demand. These drivers of electricity demand add complexity in electricity demand forecasting and create uncertainties around the forecasts. Electricity forecasting methods have evolved from counting light bulbs and engineering approaches which were based on the use of charts and tables, to manually forecasting future demand, to sophisticated forecasting techniques. The availability of powerful computers and statistical software today enables forecasters to produce more accurate forecasts through sophisticated forecasting methods.

Electricity demand forecasts can be developed for short, medium- or long-term horizons, and they could be provided as point forecasts, which give one value at each time interval, or as probabilistic forecasts which give a full distribution of future values and therefore allow the assessment of uncertainties around the forecasts. Quantification of uncertainties around forecasts is even more important for long-term forecasts, because, as Sigauke and Chikobvu (2011) indicated, long-term decision-making in the electricity sector involves planning under substantial uncertainty.

In the literature to date, short-term electricity demand forecasting has attracted substantial attention because of its importance for power system control, unit commitment and electricity markets. Medium- and long-term forecasting have not received much attention, despite their value for system planning and budget allocation (Hyndman & Fan 2010). International literature on probabilistic load forecasting is very limited, and for load forecasting it is still dominated by short-term point forecasting.

There are some literature available on long-term forecasting of annual electricity demand as well as peak electricity demand in South Africa (Inglesi-Lotz 2011; Koen, Magadla & Mokilane 2014; Rasuba, Khuluse & Elphinstone 2010; Sigauke 2014; Sigauke & Chikobvu 2011; Ziramba 2008). The models used to forecast electricity demand in South Africa do not forecast the full distribution of demand and most of them are for short-term electricity demand (Sigauke 2014 among others). The objectives of the study were (1) to apply a quantile regression (QR) model to forecast hourly distribution of electricity demand in South Africa; (2) to investigate variabilities in the forecasts and evaluate

uncertainties around point forecasts and (3) to determine whether the future peak electricity demands are likely to increase or decrease.

Methodology framework

Weron and Misiorek (2004) indicate that forecasting models could be classified into two broad streams: those that use statistical methods (e.g., multiple regression, autoregressive (AR), autoregressive integrated moving average [ARIMA], autoregressive generalised autoregressive conditional heteroscedasticity [AR-GARCH], jump diffusion, factor models, regime switching models, multilevel models, mixed models and semi-parametric models) and those that use computational intelligence techniques (such as fuzzy techniques, support vector machines and, in particular, artificial neural networks [ANNs]).

Statistical methods differ from ANN in that the former forecast the current value of a variable by using mathematical combination of the previous values of that variable and sometimes the previous values of exogenous factors (Weron & Misiorek 2004). Weron and Misiorek (2004) pointed out that the reviewers of ANN-based forecasting systems have concluded that much work still needs to be conducted before they are accepted as established forecasting techniques. ANN is considered a black-box modelling approach. In electricity demand forecasting, statistical models are attractive because physical interpretation may be attached to their components, and hence allow forecasters to understand behaviour (Weron & Misiorek 2004).

Suganthi and Samuel (2012) give a comprehensive review of demand forecasting models which are commonly used in the energy sector. Electricity demand data consist of a sequence of observations collected over equally spaced time periods (hourly) with no missing data. The observations are serially correlated. Statistical modelling approaches for forecasting electricity demand can be divided into three main groups. Firstly, there are approaches which consider demand as a univariate time series, that is, a load forecasting process which results in one forecasted value at each step, or point forecasts. Secondly, there are approaches which take each intraday period as a separate parametric regression and estimate each model's parameters separately, ignoring the intraday correlation in the process. Thirdly, there are approaches which consider each intraday period as a separate parametric regression model and estimate model parameters together in a way that takes the intraday correlations into consideration.

Within the univariate time series framework, the stochastic nature of electricity demand as a function of time has frequently been modelled with seasonal autoregressive integrated moving average (SARIMA) and state space models (Taylor, De Menezes & McSharry 2006). Mostly, electricity data exhibit not only non-constant mean and variance, but also multiple seasonalities corresponding to daily, weekly, monthly and yearly periodicity. The assumption of homoscedasticity in SARIMA models is also inappropriate for the forecasting of electricity demand. Furthermore, SARIMA models are used for point forecasting and cannot forecast the full demand distribution.

SARIMA models could be extended to a SARIMA-GARCH model to account for the possibility of heteroscedasticity. A GARCH modelling approach could be used to capture potential conditional heteroscedasticity in electricity data (Byström 2005; Taylor 2006). However, this modelling approach does not accommodate exogenous drivers of electricity demand, and is used for point forecasting.

Seasonal autoregressive integrated moving average with exogenous variables (SARIMAX) models, also known as regression-SARIMA, have been used in load forecasting in order to incorporate important drivers of demand such as calendar variables and temperature (Bunn 1982; Suganthi & Samuel 2012; Weron 2007). This method uses an ordinary least squares regression (OLS) model which may be affected by outliers and could underestimate the peaks as it models the mean of the distribution.

Structural time series (STS) models have also been successfully used in demand forecasting. STS modelling was developed by Harvey (1990) and it involves the decomposition of a time series into trend, seasonality, cycle and irregular (noise) components. This modelling approach can accommodate drivers of electricity demand like temperature, but is also used for point forecasting.

Hyndman and Fan (2010) propose a semi-parametric additive model in the regression framework, but which includes nonlinear relationships and serially correlated errors. The proposed models allow for nonlinear and non-parametric terms using the framework of additive models. The authors applied this method to develop long-term probabilistic load forecasts.

OLS regression models model the relationship between covariates X and the conditional mean of a response variable Y given $X = x$. Koenker and Bassett (1978) argue that what the regression curve does, is to give a summary for the averages of the distributions corresponding to the set of Xs. One could go further and compute several different regression curves corresponding to the various percentage points of the distribution and thus get a more complete picture of the set (Koenker & Bassett 1978). Ordinarily this is not done, and so regression often gives a rather incomplete picture. In forecasting electricity demand, least squares regression models the mean of the electricity demand as the dependent variable. OLS regression determines coefficients α_0 and α_i which minimise $\sum_i^n [y_i - (\alpha_0 + \alpha_i x_i)]^2$.

To apply an OLS regression model, the data must meet stringent assumptions, such as that the residuals should be normally distributed, the observations should be independent and the variance of the residuals should be homoscedastic. As we are dealing with time series data the observations are

not independent, and for electricity demand data, the variance is heteroscedastic. Therefore, some assumptions of OLS are violated in the time series data used in this study.

In this article, QR is proposed for developing long-term probabilistic forecasts. QR was developed as an extension of OLS regression for estimating rates of change in all parts of the distribution of the response variable (Cade & Noon 2003). QR offers a comprehensive strategy for completing the regression picture and it has been applied in ecology (Cade & Noon 2003). QR has been widely used in financial economics where the data are volatile and extremes are important to study. Gibbons and Faruqui (2014) applied QR methodology for forecasting the annual peak electricity demand.

Cornec (2014) proposes QR as a way of estimating the distribution of forecasts, and uses the dispersion of the estimated quantiles for calculating an uncertainty index. QR imposes no normality assumption, allowing, for example, for fat-tailed distributions, which is useful for forecasting extreme events. In electricity demand forecasting, QR can be used to model the median, the 1st, 5th, 10th, 90th, 95th and 99th percentiles or all quantiles to describe the full distribution of forecasted electricity demand at each hour. QR attempts to find the coefficients a_0 and a_i which minimises $\sum_i^n f(y_i - (\alpha_0 + \alpha_i x_i))$.

QR does not require any distribution assumptions regarding the population and can estimate the parameters non-parametrically (Koenker & Bassett 1982). A linear model for the τth quantile is given by:

$$y_i = x_i^T \beta + \epsilon_i, i = 1, \ldots, n. \qquad \text{[Eqn 1]}$$

where x_i^T is the transposed (indicated by T) design matrix (matrix of covariates), β is the regression coefficients and the τth quantile of $\square_i$ is assumed to be zero. The standard QR model is given by:

$$Q_\tau(Y|X) = x_i^T \beta, \qquad \text{[Eqn 2]}$$

where

$$Q_\tau(Y|X) = \inf\{y : F_\tau(y|x_i) \geq \tau\}, \qquad \text{[Eqn 3]}$$

is the conditional τth quantile of the response (y_i) given the covariate (x_i) and $Q_\tau(Y|X)$ is non-decreasing function of τ for any given x. β is the vector of parameters and is the marginal change in the quantile because of the marginal change in x_i

In estimating the QR model for a given quantile, we follow the ideas of Koenker (2005) and Yue and Rue (2011) who used the standard approach of Koenker and Bassett (1978) to estimate their QR model. QR minimises the tilted absolute function $\rho_\tau(.)$, which they called the check-function (Maistre, Lavergne & Patilea 2017), which asymmetrically weights residuals from the model to a degree that depends upon τ.

$$\rho_\tau(\epsilon) = \begin{cases} (1-\tau)\epsilon, & \epsilon < 0 \\ \tau\epsilon, & \epsilon \geq 0, \end{cases} \quad 0 < \tau < 1; \qquad \text{[Eqn 4]}$$

$\rho_\tau(\epsilon)$ is a continuous piecewise linear function and non-differentiable at $\square = 0$ but differentiable everywhere else (has directional derivative in all directions) (Yue & Rue 2011). This check-function ensures that all ρ_τ are positive and the scale is based on the probability τ. A linear model is estimated by solving:

$$\hat{\beta}(\tau) = argmin_{\beta \epsilon \mathcal{R}^p} \sum_i^n \rho_\tau(y_i - x_i^T \beta). \qquad \text{[Eqn 5]}$$

This concept is extendable to any quantile, such as the 75th, 90th and 99th percentile. The QR estimator for β at quantile τ minimises the objective function:

$$Q_\tau(\beta) = min_{\beta \epsilon \mathcal{R}^p} \left\{ \sum_{i:y_i \geq x_i^T\beta}^n \tau|y_i - x_i^T\beta| + \sum_{i:y_i \geq x_i^T\beta}^n (1-\tau)|y_i - x_i^T\beta| \right\}.$$
$$\text{[Eqn 6]}$$

This is a non-differentiable function and there is no closed-form solution for $\hat{\beta}$; instead these parameters can be found using a linear programming algorithm (Gibbons & Faruqui 2014). The minimisation is done for each subsection defined by ρ_τ, where the estimate of the τth quantile function is achieved with the parametric function $x_i^T \beta$.

Features that characterise QR and differentiate it from other regression methods are:

- QR computes several different regression curves corresponding to the various percentile points of the distribution and thus provides a more complete picture of the relationship between the response variable and the covariates.
- Heteroscedasticity can be detected and, if the data are heteroscedastic, median regression estimators can be used instead of mean regression estimators.
- Median regression is more robust to outliers than other regression methods that use mean estimators, and it is semi-parametric, therefore, avoiding assumptions about the parametric distribution of the error process.

QR is, therefore, considered to be more suitable than other methods, given the type of data used, as well as its ability to provide the full distribution of forecasted electricity demand.

Data and analysis

The hourly electricity demand data for South Africa for the period 1997–2015 was provided by Eskom. For developing the long-term forecasting model, a transformed series was developed using a logarithmic transformation. The logarithmic transformation is convenient for turning a highly skewed variable into one that is more approximately normal (Benoit 2011).

Hourly demands were forecasted from 2006 to 2023. The data from 2013 to 2015 were withheld in order to validate the model. The forecasts from 2006 to 2012 were then used as in-sample forecasts, whereas the forecasts from 2013 to 2023 were out of sample forecasts.

Various time-related variables were used as covariates, namely day, public holidays, months, weekends, December break and seasons (see Table 1 for details). Lagged demand variables were included in the model to test whether suspected lagged demand effects from the high degree of diurnal activity in electricity usage were significant, that is, whether the South African consumers of electricity typically exhibit consistent daily patterns of usage (as in Farland 2013). For example, in the afternoon when people return from work, around 19:00, they start cooking, watch TV and take bath and at this time the household electricity demand could go up. Fourier series or harmonic terms were

TABLE 1: Variables used in the quantile regression.

Variable	Type of variable	Scale	Created dummy variables
Demand	Dependent	Continuous	-
Newyear	Independent	Dichotomous	1 if day is 01 January; 0 otherwise
Humanrights	Independent	Dichotomous	1 if day is 21 March; 0 otherwise
FreedomDay	Independent	Dichotomous	1 if day is 27 April; 0 otherwise
WorkersDay	Independent	Dichotomous	1 if day is 01 May; 0 otherwise
YouthDay	Independent	Dichotomous	1 if day is 16 June; 0 otherwise
HeritageDay	Independent	Dichotomous	1 if day is 24 September; 0 otherwise
ReconciliationDay	Independent	Dichotomous	1 if day is 16 December; 0 otherwise
ChristmasDay	Independent	Dichotomous	1 if day is 25 December; 0 otherwise
GoodwillDay	Independent	Dichotomous	1 if day is 26 December; 0 otherwise
Month1	Independent	Dichotomous	1 if Month is January; 0 otherwise
Month2	Independent	Dichotomous	1 if Month is February; 0 otherwise
Month3	Independent	Dichotomous	1 if Month is March; 0 otherwise
Month4	Independent	Dichotomous	1 if Month is April; 0 otherwise
Month5	Independent	Dichotomous	1 if Month is May; 0 otherwise
Month6	Independent	Dichotomous	1 if Month is June; 0 otherwise
Month7	Independent	Dichotomous	1 if Month is July; 0 otherwise
Month8	Independent	Dichotomous	1 if Month is August; 0 otherwise
Month9	Independent	Dichotomous	1 if Month is September; 0 otherwise
Month10	Independent	Dichotomous	1 if Month is October; 0 otherwise
Month11	Independent	Dichotomous	1 if Month is November; 0 otherwise
Sin6	Independent	Continuous	Sine term of Fourier series with Period 6
Cos6	Independent	Continuous	Cosine term of Fourier series with Period 6
Sin12	Independent	Continuous	Sine term of Fourier series with Period 12
Cos12	Independent	Continuous	Cosine term of Fourier series with Period 12
Sin18	Independent	Continuous	Sine term of Fourier series with Period 18
Cos18	Independent	Continuous	Cosine term of Fourier series with Period 18
Sin24	Independent	Continuous	Sine term of Fourier series with Period 24
Cos24	Independent	Continuous	Cosine term of Fourier series with Period 24
Lag70128	Independent	Continuous	The 1st time lag
Lag70152	Independent	Continuous	The 2nd time lag
Lag70176	Independent	Continuous	The 3rd time lag
Lag70200	Independent	Continuous	The 4th time lag
Lag70224	Independent	Continuous	The 5th time lag
Lag70248	Independent	Continuous	The 6th time lag
Sun	Independent	Dichotomous	1 if day is Sunday; 0 otherwise
Mon	Independent	Dichotomous	1 if day is Monday; 0 otherwise
Tues	Independent	Dichotomous	1 if day is Tuesday; 0 otherwise
Wed	Independent	Dichotomous	1 if day is Wednesday; 0 otherwise
Thurs	Independent	Dichotomous	1 if day is Thursday; 0 otherwise
Fri	Independent	Dichotomous	1 if day is Friday; 0 otherwise
Fribtwn	Independent	Dichotomous	1 if Friday preceded by a holiday; 0 otherwise
Monbtwn	Independent	Dichotomous	1 if Monday preceded a holiday; 0 otherwise
Lngwknd	Independent	Dichotomous	Long weekend
Dec_closure	Independent	Dichotomous	1 if period between 16 December and 01 January; 0 otherwise
Winter_schoolholiDay	Independent	Dichotomous	1 if period is during school closure in June/July; 0 otherwise
Easter	Independent	Dichotomous	1 if day is Easter; 0 otherwise
Winter	Independent	Dichotomous	1 if period is between June and August

Sin, Sine term of Fourier series; Cos, Cosine term of Fourier series; Lag, time lag; Fribtwn, Friday preceded by holiday; Monbtwn, Monday preceded by a holiday; Newyear, New year; Humanrights, Human rights day; FreedomDay, Freedom day; WorkersDay, Workers day; YouthDay, Youth day; HeritageDay, Heritage day; ReconciliationDay, Reconciliation day; ChristmasDay, Christmas day; GoodwillDay, Day of Goodwill; Sun, Sunday; Mon, Monday; Tues, Tuesday; Wed, Wednesday; Thurs, Thursday; Fri, Friday; Lngwknd, Long weekend; Dec_closure, December closure; Winter_SchoolholiDay, Winter school holiday.

used, where applicable, to capture the cycles inherent in the demand data.

The future hourly electricity demands were forecasted at 0.01, 0.02, 0.03, … , 0.99 quantiles of the distribution using QR, hence each hour of the day would have 99 forecasted future hourly demands, instead of forecasting just a single overall hourly demand as in the case of OLS. To avoid graphs that are too busy and difficult to read, only the 1st, 50th and the 99th percentile graphs are shown and discussed.

The uncertainties in the forecasts are captured by the interval between the 1st and 99th percentiles of the demand distribution, as this is the interval into which 98% of the possible future hourly demands are expected to fall. The wider the interval, the more uncertain we are about the forecasted hourly demand as the variability between the forecasts would be very high. The forecasts at the 50th percentile (median) are important because they could be used as point forecasts, namely, our best guess of demand at that certain hour.

The density functions give the full distribution of the hourly electricity demand. The probability of the hourly demand between the two demand points say, '*a*' and '*b*', is the area under the demand density function between the two points. The area could be calculated by integrating the density function between the two points. The probability of exceedance relates to the probability of electricity demand exceeding the specified hourly demand and this could also be calculated by integrating the density function from the specified demand and upwards. The forecasted density demand functions between 2013 and 2023 are compared. If the density curves shift towards the higher demands over time, then an increase in future hourly demand is expected. If the shift is towards smaller demands over time, then the decrease in future hourly demand is expected.

The South African electricity demand has two daily peaks, especially noticeable in winter, namely a morning demand peak at around 08:00 and an afternoon demand peak at around 19:00. As the winter peak represents the highest annual demand, this is important for planning electricity generation. Therefore, it is important to examine the demand densities of both morning and afternoon peak forecasts over the years. The morning peak density demand is generated from all possible demand forecasts at 07:00, 08:00 and 09:00, whereas the afternoon peak density demand is generated from all possible demand forecasts at 18:00, 19:00 and 20:00. The probability of the future peak electricity demand exceeding a certain value was then calculated by integrating the density functions.

The performance of the QR model is evaluated by comparing the predicted demand density functions with the actual demand density functions. The predicted demand density is generated from all demand forecasts from 1st to 99th percentile of the demand distribution. If the forecasted demand density function closely tracks the actual demand density, then it shows that the model is forecasting well and

it is, therefore, reliable. The model is also evaluated by observing the closeness of the actual demand distribution to the predicted lower and upper 99% interval. If the interval is narrow, then the predictions exhibit sharpness. The mean absolute percentage error (MAPE) is used to compare the forecasts at the 50th percentile with the actual demands; this is mainly to determine how far the point forecasts are from the actual demands.

Figures 1 and 2 depict the time series of the actual and logarithmic hourly demands, respectively, for the hourly electricity demand over the 1997–2015 period. During this period the highest demand reached was 36 826 kW in 2011, whereas the minimum was 13 533 kW in 1998.

The historical demand data from 1997 to 2015 indicate that the demand for electricity increased steadily between 1997 and 2007 (Figure 1). During this period, South Africa experienced accelerated economic growth and a large number of new households were connected to the grid as government wanted to make electricity accessible to all South Africans. The electricity demand from Eskom stabilised between 2007 and 2012 and started declining in the latest 4 years until 2015 (see Figure 1). The decline in electricity demand from Eskom could partly be attributed to the shrinking economic growth between 2007 and 2015 and the growth of renewable sources of electricity.
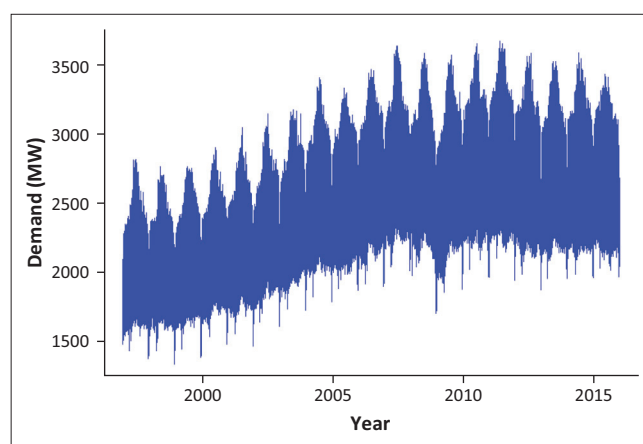

FIGURE 1: Electricity demand between 1997 and 2015 in South Africa.
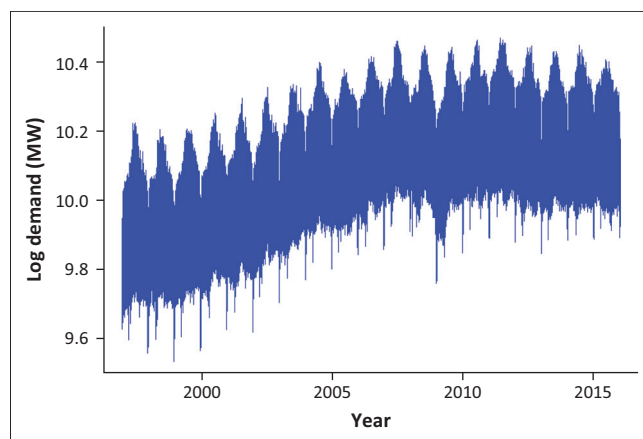

FIGURE 2: Adjusted electricity demand between 1997 and 2015 in South Africa.

Table 1 provides a summary of all the variables considered in the modelling of hourly demand. The demand data have Periods 6, 12, 18 and 24 as shown in Figure 1-A2. The Fourier series terms were formed using these periods.

Results

Model assessment

For each hour, the demand at the 1st to 99th percentiles was forecasted from 2006 to 2023, which translated into the full demand distribution being forecasted.

The actual and predicted demand densities between 2012 and 2015 illustrated in Figure 3 were used to assess the model fit. If the predicted demand distribution is close to the actual demand distribution, then the forecasts are considered to be reliable. The closer the 1st and the 99th percentile points are to the actual demand, the better it is and this indicates the sharpness of the forecasts (Figures 4–6). The sharpness of the forecasts refers to how tightly the predicted distribution covers the actual distribution.

The MAPE between the hourly demand forecasts at the 50th percentile and the actual hourly demand over the period of 4 years were below 5% and the overall MAPE was 2.77%, as shown in Table 2-A1. Lewis (1982) indicates that a MAPE of less than 10% can be classified as a highly accurate forecast. The QR model therefore provides very good demand forecasts at the 50th percentile.

Further confirmation of the suitability of fit of the QR model is obtained by comparing the actual values to the full range of quantile predictions.
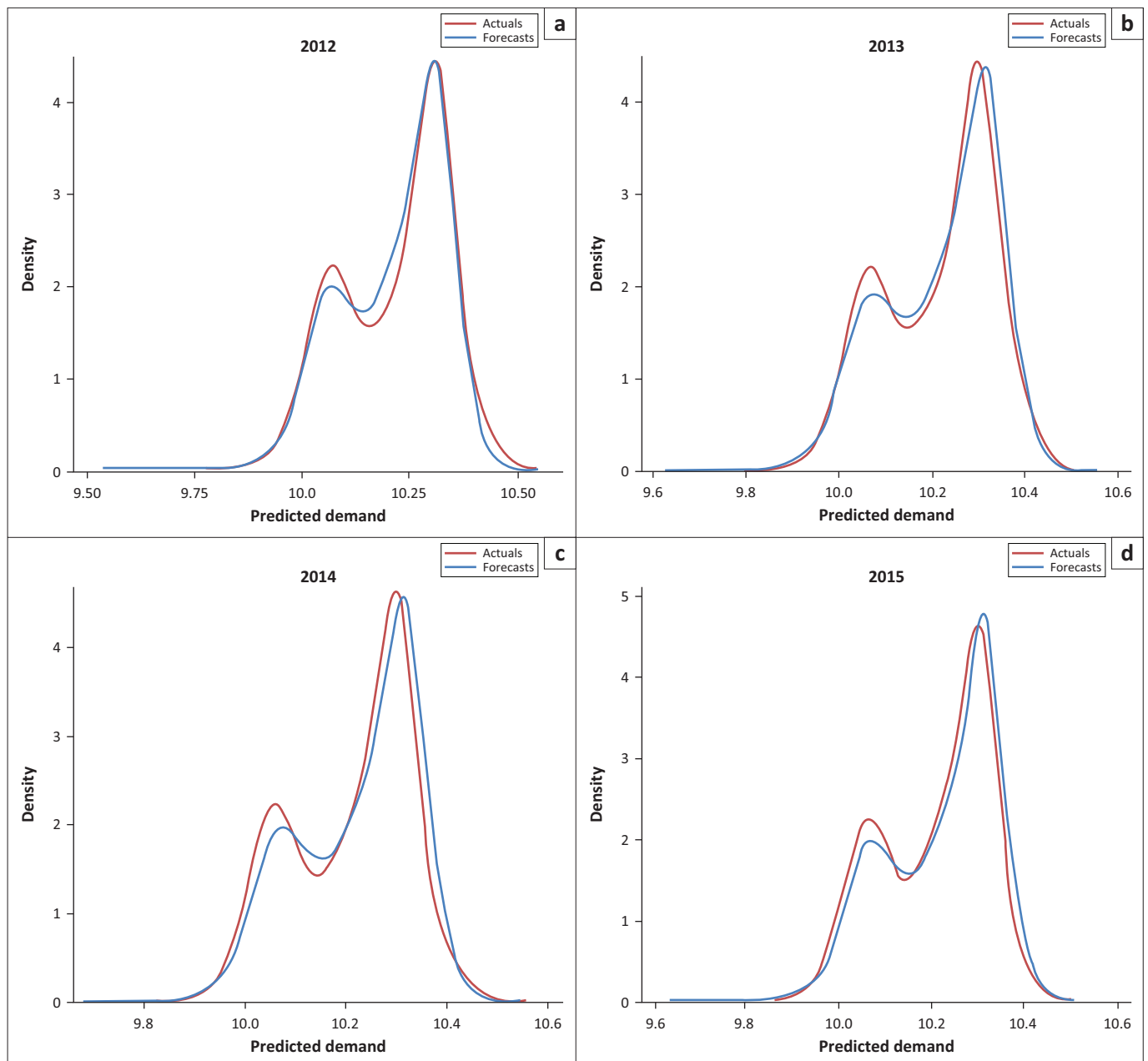

FIGURE 3: Comparison of actual and forecasted demand distributions: (a) 2012; (b) 2013; (c) 2014 and (d) 2015.
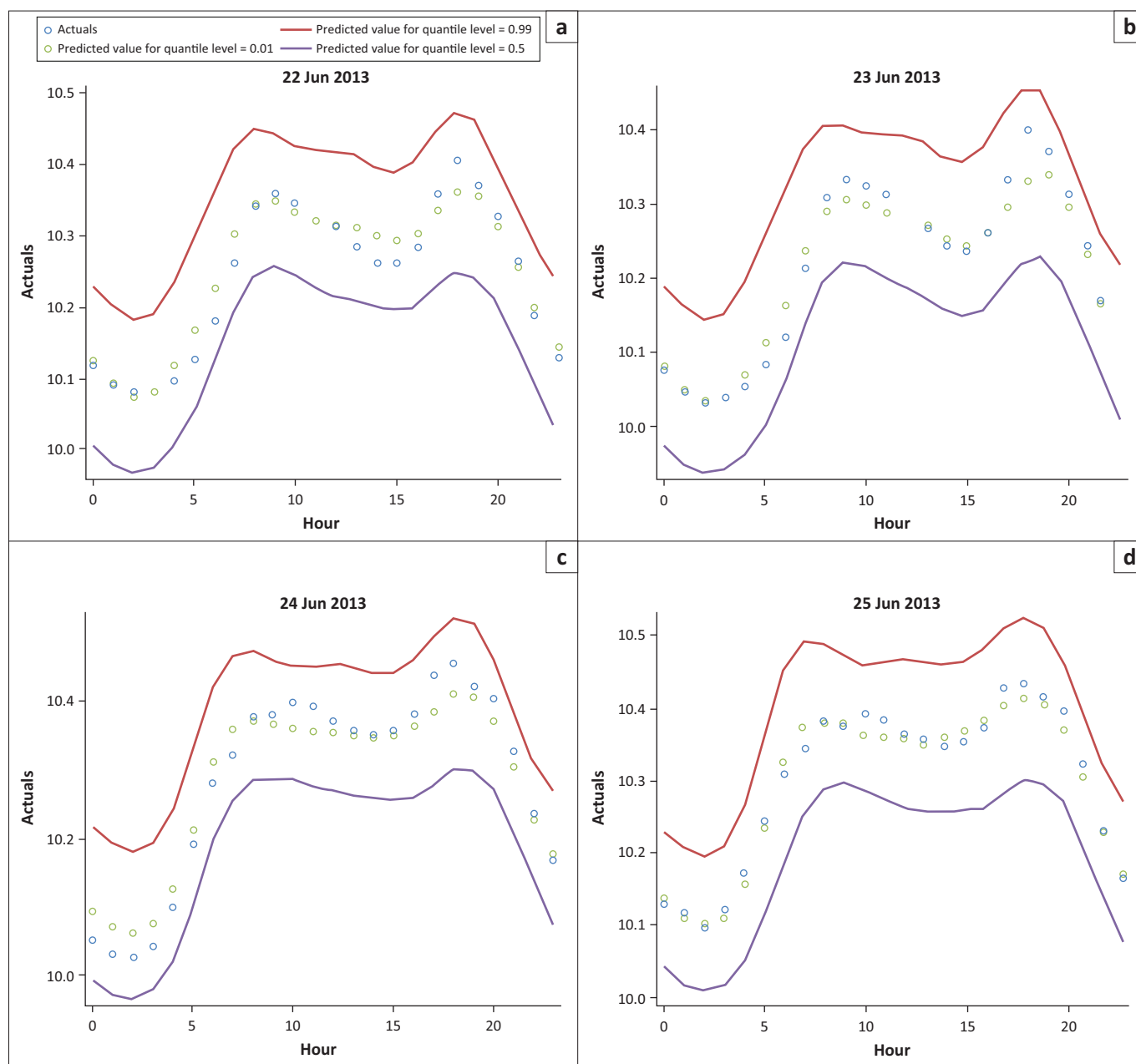
FIGURE 4: Distribution of forecasted daily electricity demand: (a) 22 June 2013, (b) 23 June 2013, (c) 24 June 2013, (d) 25 June 2013.

Model estimates

For illustration purposes, estimates for only three of the QR models (at 0.01, 0.5 and 0.99 quantile levels) are given in Table 1-A1. At the 5% level of significance, some variables in Table 1-A1 are significant at the certain percentile of the demand distribution, but insignificant at others. For example, at the 5% level of significance, the variable 'Month1' is significant at the 1st percentile, but not significant at the 99th percentile of the distribution.

Probabilistic forecasts of the daily profiles over the years

For illustration purposes, 4 days in June (22–25 June) were selected and their results from 2013 to 2015 were discussed. (Note that June falls in the high-demand winter period of the year.) The different panels in Figure 4 give the hourly

electricity demand of each of the 4 days in 2013. For each day, the green circles represents hourly demands at the 1st percentile level of the distribution; these are the points below which 1% of all possible future hourly demands are expected to fall. The grey line represents future hourly demands at the 99th percentile of the demand distribution; these are the points above which 1% of the future hourly demands would fall and below which 99% of all possible future hourly demands are expected to fall. The blue circles represent the forecasted hourly demands at the 50th percentile and the red circles represent the actual hourly demands. Figures 5 and 6 give the hourly demand forecasts for the same 3 days in 2014 and 2015, respectively.

Figures 4–6 confirm that the demand forecasts at the 50th percentile are close to the actual hourly electricity demand in winter.
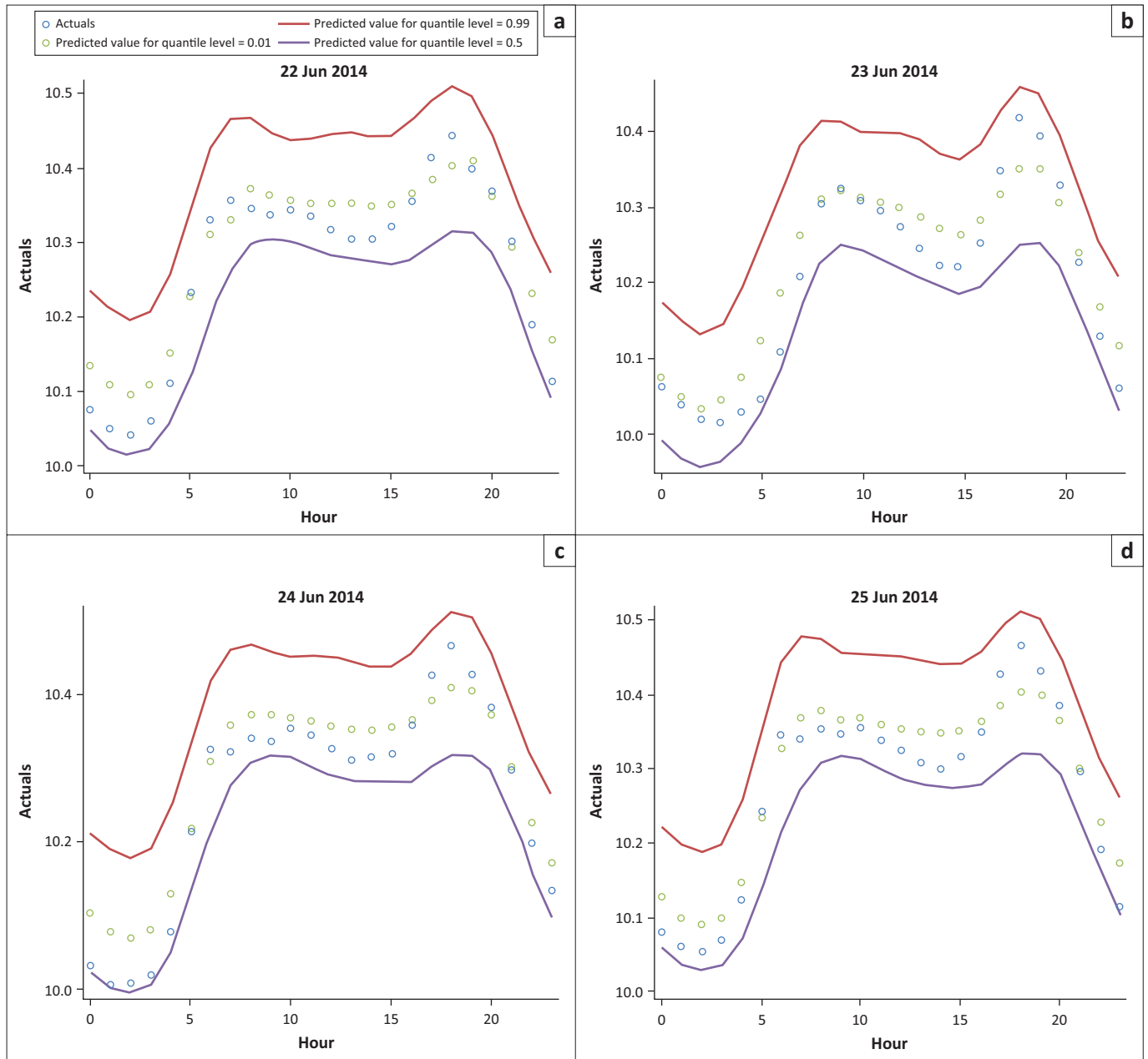
FIGURE 5: Distribution of forecasted daily electricity demand: (a) 22 June 2014, (b) 23 June 2014, (c) 24 June 2014, (d) 25 June 2014.

In addition to just using the 50th percentile as a 'best guess' or point forecast, information contained in the other quantile forecasts produce probabilistic information which may also be useful in the planning process. It can be seen from Figures 4–6 that the interval between the 1st and the 99th quantiles was in fact fairly narrow, and therefore the uncertainties around the point forecasts were not too large.

Using probabilistic forecasts of hourly demand distributions for comparing demand over the years

The hourly electricity demand distribution in South Africa is bimodal as shown in Figure 7. By comparing the demand density functions over the years, insight into expected shifts in patterns can be obtained. The forecasted demand distributions obtained from the QR models for the period

investigated suggest that the hourly electricity demand from Eskom is more likely to shift towards lower demands over the years until 2023 (Figure 7). The apparent year to year decline in electricity demand from Eskom between 2012 and 2015, among others, could be attributed to the increase in the number of households and companies generating their own electricity through renewable energies, the shrinking economic growth and the increase in electricity prices. The renewable electricity market in South Africa is growing.

In addition, the forecasted hourly density demands (in Figures 7–9) could be used to calculate the probabilities of exceedance, for example, the probability of hourly demand exceeding 32 860 kW (exp[10.4]). This probability of exceedance can be calculated from the area under the density curve comprising all demands ranging from 32 860 kW and
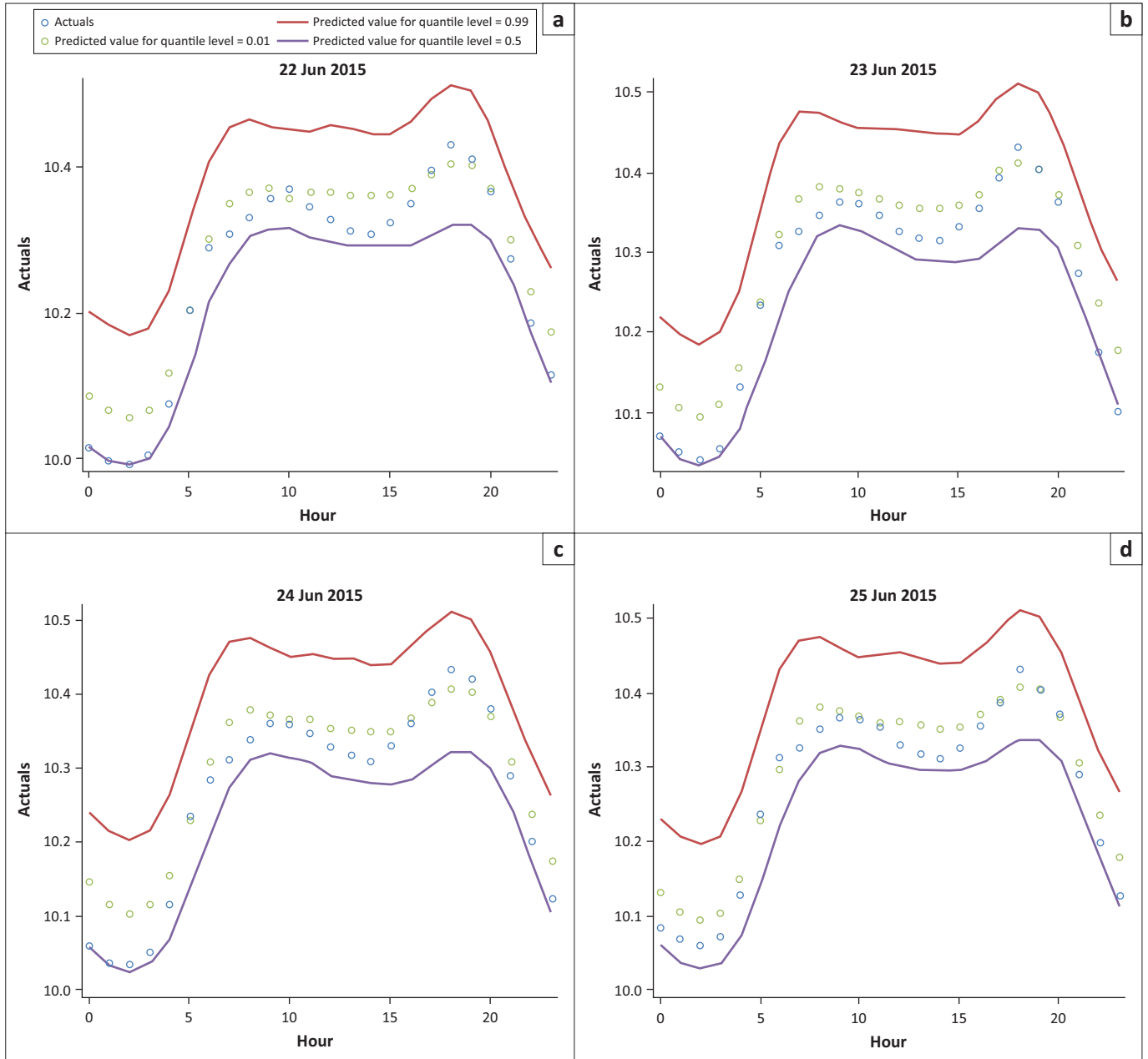
FIGURE 6: Distribution of forecasted daily electricity demand: (a) 22 June 2015, (b) 23 June 2015, (c) 24 June 2015, (d) 25 June 2015.
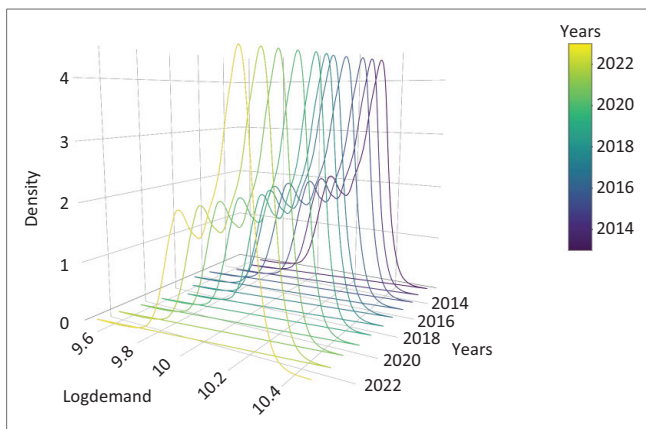

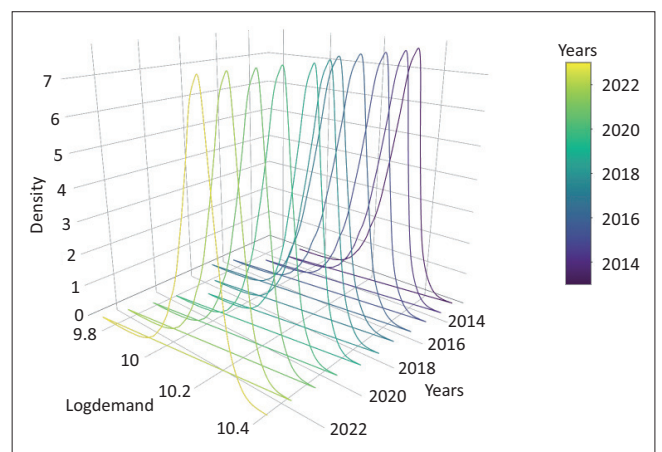**FIGURE 7:** Distribution of forecasted hourly demand between 2013 and 2023.

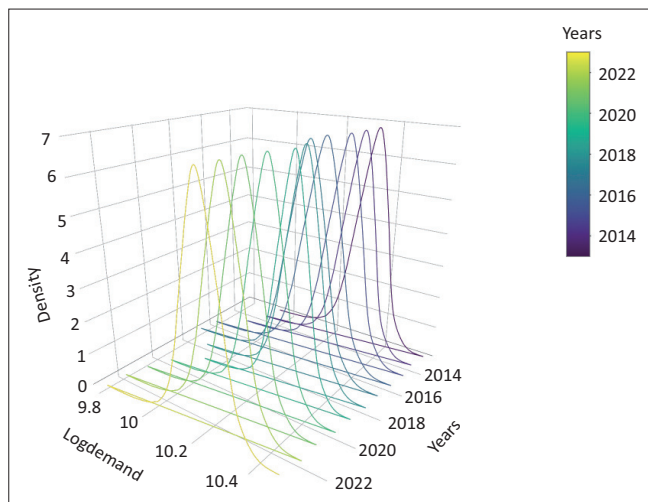**FIGURE 8:** The morning peak demand distributions.

FIGURE 9: The afternoon peak demand distributions.

above, and the area can be computed by integrating the density functions in Figure 7. The probability of demand exceeding 32 860 kW is less likely in 2023 than it is in any previous year while 2015 had the highest probability of demand exceeding 32 860 kW in the period between 2013 and 2023 as shown in Figure 7.

Finally, the forecasts obtained from QR can be used to investigate the expected future peak demands and their probability of exceedance over the years. The annual peak demands are very important for planning purposes, as these represent the maximum that would need to be supplied in an hour and if the power generating company could meet the daily peak hourly demand, it could meet any hourly demand. Figures 8 and 9 suggest that the morning and afternoon peak demand distributions are more likely to shift towards lower demands over the years until 2023.

Conclusions and discussion

The daily electricity demand in South Africa generally has two peaks, more noticeable during winter than summer seasons. The morning demand peak occurs at around 08:00 and the afternoon demand peak at around 19:00. OLS would most likely underestimate the peaks as it models the mean of the demand distribution while QR models demand at all percentiles of the demand distribution and therefore can provide better peak forecasts. In addition, as QR gives the full hourly demand distribution, the uncertainties around the forecasts are quantifiable. While the best guess of the future hourly electricity demand can be obtained from forecasted demands at the 50th percentile, QR gives forecasts at all percentiles of the distribution, allowing the potential variabilities in the forecasts to be evaluated by comparing the 50th percentile forecasts with the forecasts at other percentiles. Additional planning information, such as expected pattern shifts and probable peak values, could also be obtained from the forecasts produced by the QR model, while such information would not easily be obtained from other forecasting approaches.

The first important finding presented in this article is that the demand forecasts at the 50th percentile from the QR model closely estimate the actual hourly demands (see the red and blue circles in Figure 4–6 and the MAPE values in Table 2-A1 in Appendix 1). The second important finding is that the distributions of hourly demand and the peak daily demand in South Africa are shifting towards lower demands over the years until 2023 as shown in Figures 7–9. The third finding is that QR allows the assessment of uncertainties around point forecasts. This was illustrated by calculating the probability of forecasted hourly density demands (in Figures 7–9) exceeding 32 860 kW. The probability of demand exceeding 32 860 kW was found to be less likely in 2023 than in any previous year.

The forecasted electricity demand distribution closely matched the actual demand distribution between 2012 and 2015 as shown in Figure 3. Therefore, the forecasted demand distribution is expected to continue representing the actual demand distribution until 2023. Using a QR approach to obtain long-term forecasts of hourly load profile patterns is, therefore, recommended.

Acknowledgements

Competing interests

The authors declare that they have no financial or personal relationship(s) that may have inappropriately influenced them in writing this article.

Authors' contributions

P.M. was the project leader. Did the analysis and wrote the paper. P.D, J.G. and V.S.S. were supervisors for P.M.'s PhD and gave guidance and critically read the paper. All authors edited the article. R.K. gave guidance, assisted in the writing of conclusions, proofread and did the final editing of the paper. S.S. helped with cleaning the data and searching for some literature.

References

Benoit, K., 2011, 'Linear regression models with logarithmic transformations', *London School of Economics, London* 22(1), 23–36.

Bunn, D.W., 1982, 'Short-term forecasting: A review of procedures in the electricity supply industry', *Journal of the Operational Research Society* 33(6), 533–545. https://doi.org/10.1057/jors.1982.116

Byström, H.N., 2005, 'Extreme value theory and extremely large electricity price changes', *International Review of Economics & Finance* 14(1), 41–55. https://doi.org/10.1016/S1059-0560(03)00032-7

Cade, B.S. & Noon, B.R., 2003, 'A gentle introduction to quantile regression for ecologists', *Frontiers in Ecology and the Environment* 1(8), 412–420. https://doi.org/10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2

Cornec, M., 2014, 'Constructing a conditional GDP fan chart with an application to French business survey data', *OECD Journal: Journal of Business Cycle Measurement and Analysis* 2013(2), 109–127.

Farland, J., 2013, 'Zonal and regional load forecasting in the New England wholesale electricity market: A semiparametric regression approach', Masters thesis 1911 – February 2014, 1120, viewed 16 August 2017, from http://scholarworks.umass.edu/theses/1120.

Gibbons, C. & Faruqui, A., 2014, 'Quantile regression for peak demand forecasting', viewed 15 August 2017, from https://ssrn.com/abstract=2485657.

Gneiting, T. & Katzfuss, M., 2014, 'Probabilistic forecasting', *Annual Review of Statistics and Its Application* 1, 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

Harvey, A.C., 1990, *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge, UK.

Hong, T. & Shahidehpour, M., 2015, Load forecasting case study, EISPC, US Department of Energy, Charlotte, NC.

Hong, T., Wilson, J. & Xie, J., 2014, 'Long-term probabilistic load forecasting and normalization with hourly information', *IEEE Transactions on Smart Grid* 5(1), 456–462. https://doi.org/10.1109/TSG.2013.2274373.

Hyndman, R.J. & Fan, S., 2010, 'Density forecasting for long-term peak electricity demand', *IEEE Transactions on Power Systems* 25(2), 1142–1153. https://doi.org/10.1109/TPWRS.2009.2036017

Inglesi, R. & Pouris, A., 2010, 'Forecasting electricity demand in South Africa: A critique of Eskom's projections', *South African Journal of Science* 106(1–2), 50–53.

Inglesi-Lotz, R., 2011, 'The evolution of price elasticity of electricity demand in South Africa: A Kalman filter application', *Energy Policy* 39(6), 3690–3696. https://doi.org/10.1016/j.enpol.2011.03.078

Koen, R., Magadla, T. & Mokilane, P., 2014, 'Developing long-term scenario forecasts to support electricity generation investment decisions', In *43rd Annual Conference of the Operations Research Society of South Africa*, p. 9, Operations Research Society of South Africa, Pretoria, South Africa.

Koenker, R., 2005, *Quantile regression*, Cambridge University Press, New York, NY.

Koenker, R. & Bassett Jr, G., 1978, 'Regression quantiles', *Econometrica: Journal of the Econometric Society* 46(1), 33–50. https://doi.org/10.2307/1913643

Koenker, R. & Bassett Jr, G., 1982, 'Robust tests for heteroscedasticity based on regression quantiles', *Econometrica: Journal of the Econometric Society* 50(1), 43–61. https://doi.org/10.2307/1912528

Lehohla, P., 2005, 'Census 2001: Stages in the life cycle of South Africans', viewed 12 October, 2007, http://www.statssa.gov.za/publications/C2001Stages/C2001Stages.pdf

Lewis, C.D., 1982, *Industrial and business forecasting methods: A radical guide to exponential smoothing and curve fitting*, London, Butterworth Scientific.

Maistre, S., Lavergne, P. & Patilea, V., 2017, 'Powerful nonparametric checks for quantile regression', *Journal of Statistical Planning and Inference* 180, 13–29. https://doi.org/10.1016/j.jspi.2016.08.006

Rasuba, M., Khuluse, S. & Elphinstone, C.D., 2010, 'Forecasts for electricity demand in South Africa (2010–2035) using the CSIR sectoral regression model', viewed 04 September 2017, from http://www.energy.gov.za/irp/irp%20files/CSIR_model_IRP%20forecasts%202010_final_v2.pdf.

Sigauke, C., 2014, *Modelling electricity demand in South Africa*, University of the Free State, Doctoral dissertation, University of the Free State, viewed 26 August, 2017 from, http://scholar.ufs.ac.za:8080/xmlui/handle/11660/1569.

Sigauke, C. & Chikobvu, D., 2011, 'Prediction of daily peak electricity demand in South Africa using volatility forecasting models', *Energy Economics* 33(5), 882–888.

Statistics South Africa, 1998, *The people of South Africa population census, 1996*, Pretoria, Statistics South Africa.

Statistics South Africa, 2008, *Community survey, 2007: Basic results municipalities*, Pretoria, Statistics South Africa.

Statistics South Africa, 2011, *Census 2011 methodology and highlights of key results*, Pretoria, Statistics South Africa.

Statistics South Africa, n.d., *Economic growth*, viewed August 16 2017, from http://www.statssa.gov.za/?page_id=735&id=1

Suganthi, L. & Samuel, A.A., 2012, 'Energy models for demand forecasting – A review', *Renewable and Sustainable Energy Reviews* 16(2), 1223–1240. https://doi.org/10.1016/j.rser.2011.08.014

Tay, A.S. & Wallis, K.F. 2000, 'Density forecasting: A survey', *Journal of Forecasting* 19(4), 235–254. https://doi.org/10.1002/1099-131X(200007)19:4<235::AID-FOR772>3.0.CO;2-L

Taylor, J.W., 2006, 'Density forecasting for the efficient balancing of the generation and consumption of electricity', *International Journal of Forecasting* 22(4), 707–724. https://doi.org/10.1016/j.ijforecast.2006.02.001

Taylor, J.W., De Menezes, L.M. & McSharry, P.E., 2006, 'A comparison of univariate methods for forecasting electricity demand up to a day ahead', *International Journal of Forecasting* 22(1), 1–16. https://doi.org/10.1016/j.ijforecast.2005.06.006

Weron, R., 2007, *Modeling and forecasting electricity loads and prices: A statistical approach* (Vol. 403), John Wiley & Sons.

Weron, R. & Misiorek, A., 2004, 'Modeling and forecasting electricity loads: A comparison', *International Conference EEM-04 proceedings,* September 20–22, 2004, Poland, pp. 135–142.

Yue, Y.R. & Rue, H., 2011, 'Bayesian inference for additive mixed quantile regression models', *Computational Statistics & Data Analysis* 55(1), 84–96. https://doi.org/10.1016/j.csda.2010.05.006

Ziramba, E., 2008, 'The demand for residential electricity in South Africa', *Energy Policy* 36(9), 3460–3466. https://doi.org/10.1016/j.enpol.2008.05.026

Appendix 1

TABLE 1-A1: Selected quantile regression models.

Variables	1st percentile level			50th percentile level			99th percentile level		
	Estimates	95% confidence limits	Pr > |t|	Estimates	95% confidence limits	Pr > |t|	Estimates	95% confidence limits	Pr > |t|
Intercept	8.1435	8.0515 8.2355	<0.0001	9.3791	9.351 9.4071	<0.0001	9.7052	9.6011 9.8093	<0.0001
Newyear	-0.1577	-0.1909 -0.1246	<0.0001	-0.1189	-0.1328 -0.105	<0.0001	-0.0308	-0.0661 0.0045	0.0872
Humanrights	-0.0626	-0.0999 -0.0252	0.001	-0.051	-0.0614 -0.0407	<0.0001	-0.0236	-0.0376 -0.0095	0.001
FreedomDay	-0.0612	-0.1149 -0.0076	0.0251	-0.0315	-0.0412 -0.0219	<0.0001	-0.0443	-0.0602 -0.0285	<0.0001
WorkersDay	-0.0386	-0.055 -0.0221	<0.0001	-0.0699	-0.0773 -0.0625	<0.0001	-0.064	-0.0817 -0.0463	<0.0001
YouthDay	-0.0903	-0.1129 -0.0676	<0.0001	-0.0566	-0.0695 -0.0436	<0.0001	-0.0485	-0.0641 -0.033	<0.0001
HeritageDay	-0.0711	-0.0924 -0.0498	<0.0001	-0.0532	-0.0618 -0.0447	<0.0001	-0.0355	-0.0604 -0.0107	0.0051
Reconciliation Day	0.0251	0.0002 0.05	0.0478	-0.0008	-0.0093 0.0077	0.8546	0.0287	0.0059 0.0515	0.0137
ChristmasDay	-0.1253	-0.1635 -0.0872	<0.0001	-0.0844	-0.0981 -0.0706	<0.0001	-0.0586	-0.0782 -0.0389	<0.0001
GoodwillDay	-0.1014	-0.1377 -0.0652	<0.0001	-0.0941	-0.105 -0.0831	<0.0001	-0.0882	-0.1077 -0.0686	<0.0001
Month1	-0.0703	-0.0784 -0.0622	<0.0001	-0.0095	-0.0117 -0.0073	<0.0001	0.0026	-0.0006 0.0058	0.1065
Month2	0.003	-0.0022 0.0082	0.2596	0.0078	0.0059 0.0096	<0.0001	0.0037	0.0008 0.0066	0.0116
Month3	-0.0015	-0.0074 0.0043	0.6131	-0.0006	-0.0026 0.0014	0.5641	0.0083	0.0048 0.0117	<0.0001
Month4	-0.0063	-0.0132 0.0006	0.0741	-0.0102	-0.0121 -0.0084	<0.0001	0.0103	0.0063 0.0143	<0.0001
Month5	-0.0087	-0.016 -0.0014	0.0203	0.0019	-0.0009 0.0047	0.1837	0.0376	0.032 0.0431	<0.0001
Month6	-0.0196	-0.0314 -0.0079	0.001	0.0212	0.0172 0.0251	<0.0001	0.0453	0.0378 0.0528	<0.0001
Month7	-0.0158	-0.0283 -0.0033	0.0133	0.0115	0.0072 0.0158	<0.0001	0.0336	0.0244 0.0428	<0.0001
Month8	-0.0499	-0.06 -0.0398	<0.0001	-0.0016	-0.0053 0.0021	0.3961	0.0199	0.0102 0.0295	<0.0001
Month9	0.0097	0.0054 0.014	<0.0001	0.0127	0.0111 0.0144	<0.0001	0.0155	0.0118 0.0192	<0.0001
Month10	0.0168	0.0125 0.021	<0.0001	0.0105	0.0088 0.0121	<0.0001	0.0057	0.0027 0.0088	0.0002
Month11	0.0106	0.0057 0.0156	<0.0001	0.0106	0.0088 0.0125	<0.0001	0.0113	0.0083 0.0144	<0.0001
Sin6	0.0048	0.0035 0.0061	<0.0001	0.0113	0.0109 0.0118	<0.0001	0.0178	0.0169 0.0188	<0.0001
Cos6	-0.0079	-0.0092 -0.0065	<0.0001	-0.0059	-0.0063 -0.0054	<0.0001	-0.0026	-0.0036 -0.0017	<0.0001
Sin12	-0.0594	-0.0611 -0.0577	<0.0001	-0.0689	-0.0694 -0.0684	<0.0001	-0.0681	-0.0694 -0.0669	<0.0001
Cos12	0.0004	-0.0011 0.0019	0.5896	-0.0124	-0.0128 -0.0119	<0.0001	-0.0218	-0.0228 -0.0208	<0.0001
Sin18	-0.0013	-0.0025 -0.0001	0.033	0.0003	-0.0002 0.0007	0.2228	-0.0004	-0.0012 0.0004	0.305
Cos18	0.0008	-0.0005 0.002	0.2318	0.0002	-0.0003 0.0006	0.5197	0.0008	0.0001 0.0016	0.0358
Sin24	-0.0648	-0.0663 -0.0632	<0.0001	-0.0746	-0.0751 -0.074	<0.0001	-0.0719	-0.0731 -0.0708	<0.0001
Cos24	-0.0779	-0.0792 -0.0766	<0.0001	-0.0822	-0.0827 -0.0817	<0.0001	-0.0846	-0.0862 -0.083	<0.0001
Lag70128	0.0086	-0.0196 0.0367	0.5496	-0.0835	-0.0945 -0.0724	<0.0001	-0.1164	-0.1331 -0.0997	<0.0001
Lag70152	-0.0002	-0.0282 0.0279	0.9912	-0.0421	-0.0537 -0.0306	<0.0001	0.0071	-0.0147 0.0289	0.524
Lag70176	-0.091	-0.1134 -0.0686	<0.0001	-0.1039	-0.1165 -0.0913	<0.0001	-0.1735	-0.194 -0.1529	<0.0001
Lag70200	-0.0629	-0.0871 -0.0386	<0.0001	-0.0194	-0.0309 -0.0078	0.001	0.076	0.0543 0.0977	<0.0001
Lag70224	0.4198	0.3919 0.4476	<0.0001	0.4573	0.444 0.4705	<0.0001	0.2638	0.2464 0.2811	<0.0001
Lag70248	-0.08	-0.1024 -0.0575	<0.0001	-0.127	-0.1378 -0.1161	<0.0001	-0.0007	-0.0166 0.0152	0.9309
Sun	-0.012	-0.0166 -0.0073	<0.0001	-0.0164	-0.0181 -0.0146	<0.0001	-0.0201	-0.0227 -0.0174	<0.0001
Mon	0.0215	0.0172 0.0258	<0.0001	0.0096	0.0074 0.0118	<0.0001	0.0211	0.0173 0.0249	<0.0001
Tues	0.0314	0.0264 0.0363	<0.0001	0.0198	0.018 0.0217	<0.0001	0.0146	0.0116 0.0176	<0.0001
Wed	0.033	0.0281 0.038	<0.0001	0.0153	0.0133 0.0174	<0.0001	0.0089	0.0055 0.0123	<0.0001
Thurs	0.0341	0.0298 0.0385	<0.0001	0.0159	0.0141 0.0177	<0.0001	0.0105	0.0069 0.0141	<0.0001
Fri	0.0314	0.0274 0.0354	<0.0001	0.0155	0.0137 0.0173	<0.0001	0.0062	0.0035 0.0088	<0.0001
Fribtwn	-0.1295	-0.1758 -0.0833	<0.0001	-0.0673	-0.0738 -0.0608	<0.0001	-0.0348	-0.0686 -0.001	0.0438
Monbtwn	-0.0115	-0.0416 0.0187	0.4561	-0.0289	-0.0357 -0.0222	<0.0001	0.0006	-0.0136 0.0148	0.933
Lngwknd	-0.0579	-0.0666 -0.0492	<0.0001	-0.0264	-0.0288 -0.0239	<0.0001	-0.0175	-0.0222 -0.0128	<0.0001
Dec_closure	-0.1011	-0.1117 -0.0906	<0.0001	-0.0563	-0.0593 -0.0533	<0.0001	-0.0405	-0.0449 -0.0361	<0.0001
Winter_SchoolholiDay	-0.0069	-0.0117 -0.0021	0.0051	0.0045	0.0028 0.0062	<0.0001	0.0078	0.0044 0.0112	<0.0001
Easter	-0.0826	-0.0996 -0.0656	<0.0001	-0.0502	-0.0546 -0.0458	<0.0001	-0.0495	-0.0578 -0.0412	<0.0001
Winter	0.0411	0.029 0.0532	<0.0001	0.0341	0.0303 0.038	<0.0001	0.042	0.0332 0.0508	<0.0001

Sin, Sine term of Fourier series; Cos, Cosine term of Fourier series; Lag, time lag; Fribtwn, Friday preceded by holiday; Monbtwn, Monday preceded by a holiday;

Newyear, New year; Humanrights, Human rights day; FreedomDay, Freedom day; WorkersDay, Workers day; YouthDay, Youth day; HeritageDay, Heritage day; ReconciliationDay, Reconciliation day; ChristmasDay, Christmas day; GoodwillDay, Day of Goodwill; Sun, Sunday; Mon, Monday; Tues, Tuesday; Wed, Wednesday; Thurs, Thursday; Fri, Friday; Lngwknd, Long weekend; Dec_closure, December closure; Winter_SchoolholiDay, Winter school holiday.

TABLE 2-A1: Mean absolute percentage error – Forecasts at the 50th percentile.

| Hours	2012	2013	2014	2015	Average
0	2.92	2.23	3.29	3.93	3.09
1	2.75	1.83	3.02	3.67	2.82
2	2.67	1.65	2.82	3.58	2.68
3	2.58	1.54	2.64	3.37	2.53
4	2.46	1.72	2.35	3.02	2.39
5	3.85	3.67	3.68	3.84	3.76
6	4.30	4.15	3.99	4.32	4.19
7	2.76	3.44	3.88	4.62	3.68
8	2.15	2.55	2.89	3.47	2.77
9	2.15	2.26	2.54	2.65	2.40
10	2.35	1.88	1.97	2.38	2.15
11	2.59	1.79	1.91	2.48	2.19
12	2.47	1.72	2.21	2.83	2.31
13	2.50	1.79	2.49	3.15	2.48
14	2.55	1.79	2.55	3.20	2.52
15	2.58	1.75	2.30	2.96	2.40
16	2.41	1.94	2.14	2.93	2.36
17	2.85	3.23	2.80	3.60	3.12
18	3.34	4.29	3.67	3.67	3.74
19	2.19	2.42	2.12	2.55	2.32
20	2.42	2.22	1.98	2.83	2.36
21	2.19	2.21	2.22	3.31	2.48
22	2.15	2.22	2.84	3.81	2.76
23	2.60	2.19	3.23	3.93	2.99
Average	2.66	2.35	2.73	3.34	2.77

Appendix 2
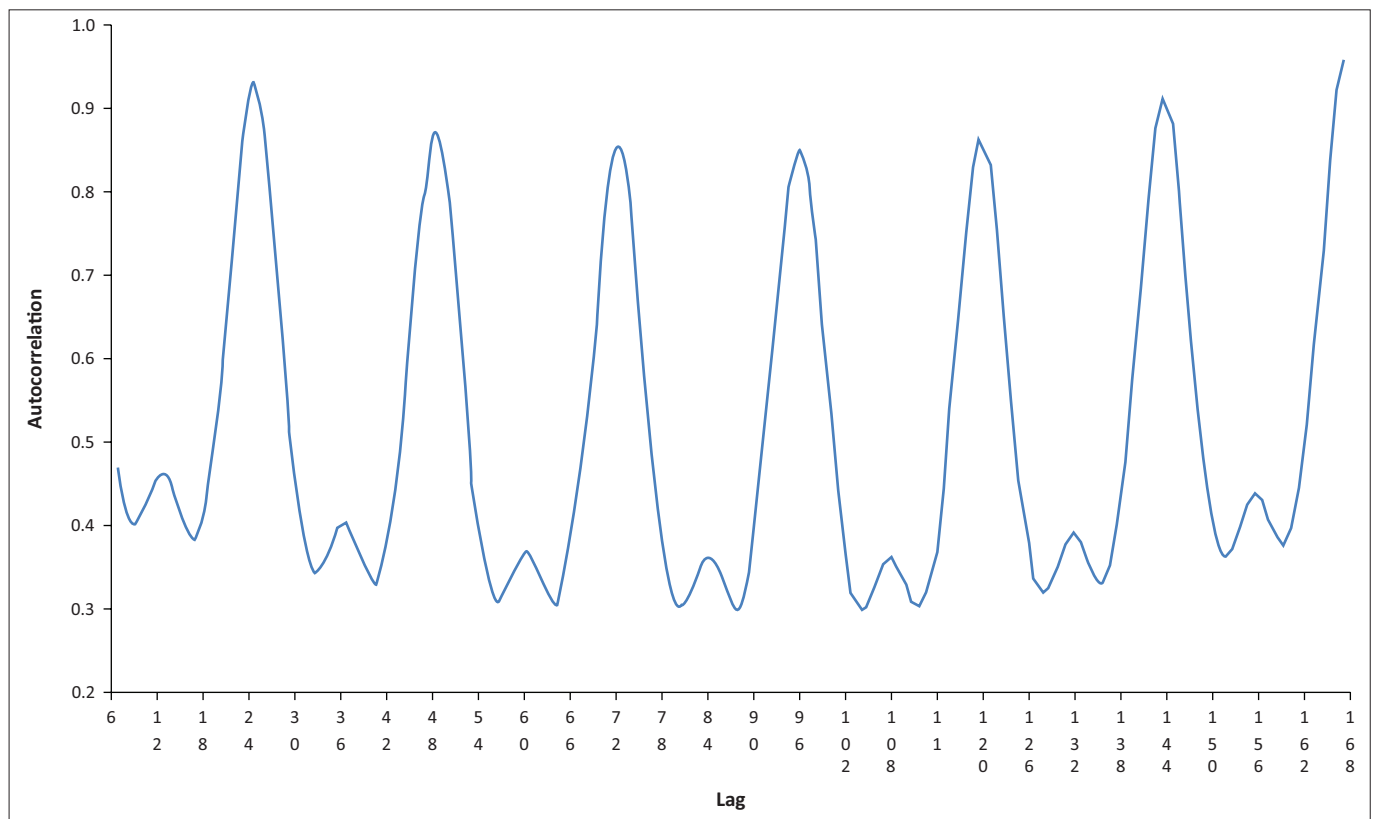
FIGURE 1-A2: The autocorrelation plot: Correlation of hourly electricity demand data.