energies

MDPI

Article

Probabilistic Hourly Load Forecasting Using Additive Quantile Regression Models

Caston Sigauke [1,*], Murendeni Maurel Nemukula [2] and Daniel Maposa [2]

[1] Department of Statistics, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa
[2] Department of Statistics and Operations Research, University of Limpopo, Private Bag X1106, Sovenga 0727, South Africa; murendeni.nemukula@ul.ac.za (M.M.N.); danmaposa@gmail.com(D.M.)
* Correspondence: caston.sigauke@univen.ac.za; Tel.: +27-15-962-8135

Received: 13 July 2018; Accepted: 15 August 2018; Published: 23 August 2018

Abstract: Short-term hourly load forecasting in South Africa using additive quantile regression (AQR) models is discussed in this study. The modelling approach allows for easy interpretability and accounting for residual autocorrelation in the joint modelling of hourly electricity data. A comparative analysis is done using generalised additive models (GAMs). In both modelling frameworks, variable selection is done using least absolute shrinkage and selection operator (Lasso) via hierarchical interactions. Four models considered are GAMs and AQR models with and without interactions, respectively. The AQR model with pairwise interactions was found to be the best fitting model. The forecasts from the four models were then combined using an algorithm based on the pinball loss (convex combination model) and also using quantile regression averaging (QRA). The AQR model with interactions was then compared with the convex combination and QRA models and the QRA model gave the most accurate forecasts. Except for the AQR model with interactions, the other two models (convex combination model and QRA model) gave prediction interval coverage probabilities that were valid for the 90%, 95% and the 99% prediction intervals. The QRA model had the smallest prediction interval normalised average width and prediction interval normalised average deviation. The modelling framework discussed in this paper has established that going beyond summary performance statistics in forecasting has merit as it gives more insight into the developed forecasting models.

Keywords: additive quantile regression; Lasso; load forecasting; generalised additive models

1. Introduction

1.1. Context

In the literature, several modelling approaches are discussed in which hourly or half-hourly electricity demand data is modelled jointly and also modelling of hourly data separately [1,2]. Pros and cons of these different approaches are discussed in the literature. Modelling hourly data jointly helps in exploring the correlation structure of the intra-day relationships and can improve the accuracy of forecasts [1]. Wood et al. [2] argue that there are practical disadvantages of modelling hourly data individually which are the failure to capture the correlation between the hourly periods, the problem of interpretation due to lack of model continuity between the hourly periods and that the developed models will lack statistical stability. The authors further argue that over-fitting and the burden of model checking are significantly reduced if one model is fitted to the data. However, this modelling approach leads to the problem of the dimensional curse. Proponents of this modelling approach argue that the use of factor analysis can help in identifying a few factors that can account for most of the

variation in the covariance matrix of the data [3]. Dordonnat et al. [4] develop a regression model which takes the intra-day correlation structure to forecast electricity demand.

1.2. Literature Review on Related Problems

It is argued in the literature that electricity demand patterns change throughout the day. Soares and Medeiros [5] argue that modelling of hourly demand data separately avoids the intra-day correlations which are common with time series data. Ramanathan et al. [6] develop flexible multiple regression models for each hour of the day to forecast electricity demand. The authors included a dynamic error structure together with adaptive adjustments which allow for the correction of forecast errors of previous hours. The modelling approach by Ramanathan et al. [6] is extended by Fan and Hyndman [7] who use a semi-parametric additive modelling framework to forecast short-term half-hourly Australian electricity demand. Using regression splines to model temperature and lagged demand effects, Fan and Hyndman [7] model each half-hourly period separately. These authors argue that modelling hourly or half-hourly electricity demand data results in more accurate forecasts.

Work on short-term load forecasting in which hourly data is modelled separately is discussed in literature. Goude et al. [8] developed generalised additive models for forecasting electricity demand. The authors used hourly load data from 2260 substations across France. Individual models for each of the 24 h of the day were developed.The developed models produced accurate forecasts for the the short- and medium-term horizons. Additive quantile regression models for forecasting probabilistic load and electricity prices are developed by Gaillard et al. [9]. The work done by Gaillard et al. [9] is extended by Fasiolo et al. [10] who developed fast calibrated additive quantile regression models. An online load forecasting system for very-short-term load forecasts is proposed by Laouafi et al. [11]. The proposed system is based on a forecast combination methodology which gives accurate forecasts in both normal and anomalous conditions. Zhang et al. [12] developed a hybrid model to short-term load forecasting based on singular spectrum analysis and support vector machine, which is optimized by the heuristic method they refer to as the Cuckoo search algorithm. The new proposed model outperformed the other heuristic models used in the study.

Boroojeni et al. [13] proposed a model which captures the complex seasonalities of electricity demand including the non-seasonal cycles. The developed model was then used for both short-term and medium-term forecasting. A boosted artificial neural network technique was presented in Khwaja et al. [14]. The developed model was compared with other artificial neural networks based models. Results showed that the new proposed model produces the lowest forecast errors. Ekonomou et al. [15] propose a methodology for short-term load forecasting. In their paper, wavelets and neural networks are used. The developed models were then applied to real and simulated data sets. In a study by Pappas et al. [16], autoregressive integrated moving average (ARIMA) models were used in short-term load forecasting. The authors showed in their study that the ARIMA model was appropriate for modelling load data with periodic variations and performed poorly during blackouts or when unexpected peaks in load demand were experienced.

A two-stage approach which is presented as a pattern recognition problem is discussed in Gajowniczek and Zabkowski [17]. The stages involve forecasting and peak detection through the use of machine learning algorithms. It is found that the proposed modelling approach produces accurate forecasts and is capable of detecting about 96.3% of the peak loads. Chapagain and Kittipiyakul [18] present a modelling approach which includes atmospheric covariates in the modelling and forecasting of short-term electricity demand. The atmospheric covariates used are cloud cover, wind speed, rainfall, relative humidity, and solar radiation including snow fall. Empirical results from this study showed a significant improvement in the forecast accuracy compared to models without atmospheric variables. Divina et al. [19] show that the use of a stacking ensemble learning scheme results in combined forecasts which are more accurate compared to the forecasts from individual models. Nagbe et al. [20] developed a functional vector autoregressive state space model for short-term electricity demand.

The developed model was tested on real-life data sets and results showed that the modelling approach is adequate in forecasting electricity demand.

Short-term load forecasting using South African data is discussed in the literature. A regression-seasonal autoregressive integrated moving average (RegSARIMA) model for predicting short-term daily peak electricity demand is discussed in Chikobvu and Sigauke [21]. A comparative analysis is done with SARIMA and Holt–Winter's triple exponential smoothing models. Empirical results from this study show that the RegSARIMA model is capable of capturing important drivers of electricity demand. In another study, an additive regression model for forecasting daily winter peak electricity demand is presented in Sigauke and Chikobvu [22]. The authors show that electricity demand in South Africa is highly sensitive to cold temperatures compared to hot temperatures. A more recent study by Sigauke and Chikobvu [23] compares the performance of time series regression models in forecasting short-term daily peak electricity demand in South Africa. Temperature effects are smoothed using regression splines and linear splines. The model in which regression splines are used produced better forecasting results.

Joint modelling of hourly electricity demand using additive quantile regression with pairwise interactions including an application of quantile regression averaging (QRA) is not discussed in detail in the literature. The current study intends to bridge this gap. The study focuses on an application of additive quantile regression (AQR) models. A comparative analysis is then done with the generalised additive models (GAMs) which are used as benchmark models. In this study, we discuss an application of pairwise hierarchical interactions discussed in Bien et al. [24] and Laurinec [25] who showed that the inclusion of interactions improves forecast accuracy.

1.3. Contributions

From the literature discussed in Section 1.2, the contributions of the present study are as follows: this study has established that going beyond summary performance statistics has merit as it gives more insight into the forecasting models. QRA forecasts result in valid prediction interval coverage probabilities and narrow prediction interval widths. The inclusion of hierarchical pairwise interactions and a nonlinear trend variable improves forecast accuracy and that the modelling framework allows for residual autocorrelation in the joint modelling of hourly electricity data.

A discussion of the models is presented in Section 2, with Section 3 discussing the results of the study. The conclusions are given in Section 4.

2. Theoretical Background

2.1. Quantile Regression

Developed by Koenker and Basset [26], quantile regression (QR) was introduced as a modelling framework for estimating conditional quantiles of the response variable. If Y denotes a random variable representing the response variable with corresponding covariates X, then the conditional quantile $q_{Y|X}(\tau)$, where $\tau \in (0,1)$ is defined as $q_{Y|X}(\tau) = \inf\{y \in \mathbb{R}, F_{Y|X}(y) \geq \tau\}$, where $F_{Y|X}$ represents the conditional distribution of Y given X. The conditional quantile $q_{Y|X}(\tau)$ is a solution to

$$q_{Y|X}(\tau) = \arg\min_{g} E\left[\rho_\tau(Y - g(x))|X\right], \tag{1}$$

where ρ_τ is the quantile loss also known as the pinball loss defined as $\rho_{\tau(s)} = s(\tau - \mathbf{I}(s < 0))$ and $\mathbf{I}(.)$ is an indicator function. Now, let $Y_t = X_t^T\beta + \varepsilon_t$ be a linear quantile regression where Y_t denotes hourly electricity demand, X_t the design matrix, β a vector of parameters and ε_t the error term; then, the estimates of β are given as

$$\hat{\beta}_\tau = \arg\min_{\beta \in \mathbb{R}^{æ}} \sum_{i=1}^{n} \rho_\tau\left(Y_t - X_t^T\beta\right). \tag{2}$$

2.2. Generalised Additive Models

Generalised additive models (GAMs) which were developed by Hastie and Tibshirani [27,28] are used in modelling predictors in regression-based models as a sum of smooth functions. The generalised additive model (GAM) is then written as [28–30]:

$$g(E(y_t)) = \beta_{0t} + \sum_{i=1}^{p} s_i(x_{ti}) + \varepsilon_t. \tag{3}$$

y_t follows some exponential family distribution, where g denotes a link function and usually the Gaussian link function is used, s_i are smooth functions and ε_t is the error term. The smooth function, s is written as

$$s(x) = \sum_{j=1}^{k} \beta_j b_j(x), \tag{4}$$

where β_j denotes the j^{th} parameter, and $b_j(x)$ represents the j^{th} basis function with the dimension of the basis denoted by k. There are several smoothing spline bases ranging from P-splines, thin plate regression splines, B-splines, cubic regression splines to cyclic cubic regression splines. In this study, we use P-splines and adaptive splines. We seek to find an optimal solution to the optimisation problem given in Equation (5):

$$\min \sum_{t=1}^{n} \left(y_t - \sum_{i=1}^{p} s_i(x_{ti}) \right)^2 + \sum_{i=1}^{p} \lambda_i \int (f''(x))^2 dx, \tag{5}$$

where λ_i is the i^{th} smoothing parameter.

2.3. The Proposed Models

2.3.1. Additive Quantile Regression Model

An additive quantile regression (AQR) model is a hybrid model that is a combination of GAM and QR models. AQR models were first applied to short-term load forecasting by Gaillard et al. [9] and extended by Fasiolo et al. [10]. Let y_t denote hourly electricity demand where $t = 1, \ldots, n$, n is the number of observations and let the number of days be denoted by n_d. Then, $n = 24n_d$, where 24 is the number of hours in a day and the corresponding p covariates, $x_{t1}, x_{t2}, \ldots, x_{tp}$. The AQR model is given in Equation (6) [9,10]:

$$y_{t,\tau} = \sum_{j=1}^{p} s_{j,\tau}(x_{tj}) + \varepsilon_{t,\tau}; \;\; \tau \in (0,1), \tag{6}$$

where $s_{j,\tau}$ are smooth functions and $\varepsilon_{t,\tau}$ is the error term. The smooth function, s, is written as

$$s_j(x) = \sum_{k=1}^{q} \beta_{kj} b_{kj}(x_{tj}), \tag{7}$$

where β_j denotes the j^{th} parameter, and $b_j(x)$ represents the j^{th} basis function with the dimension of the basis being denoted by q. The parameter estimates of Equation (6) are obtained by minimising the function given in Equation (8):

$$q_{Y|X}(\tau) = \sum_{t=1}^{n} \rho_\tau \left(y_{t,\tau} - \sum_{j=1}^{p} s_{j,\tau}(x_{tj}) \right), \tag{8}$$

where ρ_τ is the pinball loss function that is defined in Section 2.1. The AQR models are given in Equations (9) and (10):

$$y_{t,\tau} = \sum_{j=1}^{p} s_{j,\tau}(x_{tj}) + \sum_{k=1}^{K}\sum_{j=1}^{j} \alpha_{jk} s_j(x_{tj}) s_k(x_{tk}) + \varepsilon_{t,\tau},$$

$$\phi(B)\Phi(B^s)\varepsilon_{t,\tau} = \theta(B)\Theta(B^s)v_{t,\tau}, \tag{9}$$

$$\Rightarrow \phi(B)\Phi(B^s)\left[y_{t,\tau} - \left\{\sum_{j=1}^{p} s_{j,\tau}(x_{tj}) + \sum_{k=1}^{K}\sum_{j=1}^{J} \alpha_{jk} s_j(x_{tj}) s_k(x_{tk})\right\}\right] = \theta(B)\Theta(B^s)v_{t,\tau}. \tag{10}$$

A comparative analysis will be done with the GAM given in Equation (11) and discussed in Sigauke [31]:

$$y_t = \beta_{0t} + \sum_{i=1}^{p} s_i(x_{ti}) + \sum_{k=1}^{K}\sum_{j=1}^{J} \alpha_{jk} s_j(x_{tj}) s_k(x_{tk}) + \varepsilon_t,$$

$$\phi(B)\Phi(B^s)\varepsilon_t = \theta(B)\Theta(B^s)v_t, \tag{11}$$

$$\Rightarrow \phi(B)\Phi(B^s)\left[y_t - \left\{B_{0t} + \sum_{i=1}^{p} s_i(x_{ti}) + \sum_{k=1}^{K}\sum_{j=1}^{J} \alpha_{jk} s_j(x_{tj}) s_k(x_{tk})\right\}\right] = \theta(B)\Theta(B^s)v_t, \tag{12}$$

where y_t denotes hourly electricity demand, s_i denotes the smoothing function, x_{ti} represents the covariates, and ε_t denotes error terms which are assumed to be autocorrelated. Selection of variables is done using the least absolute shrinkage and selection operator (Lasso) for the hierarchical interactions method developed by Bien et al. [24] and implemented in the R package "hierNet" [32]. The objective is to include an interaction where both variables are included in the model. The restriction known as the strong hierarchy constraint is discussed in detail in Ben and Tibshirani [24] and Lim and Hastie [33].

2.3.2. Forecast Error Measures

There are several error measures for probabilistic forecasting which include among others the continuous rank probability score, the logarithmic score and the quantile loss that is also known as the pinball loss. In this paper, we use the pinball loss function which is relatively easy to compute and interpret [34]. The pinball loss function is given as

$$L(q_\tau, y_t) = \begin{cases} \tau(y_t - q_t), & \text{if } y_t > q_\tau, \\ (1-\tau)(q_\tau - y_t), & \text{if } y_t \leq q_\tau, \end{cases} \tag{13}$$

where q_τ is the quantile forecast and y_t is the observed value of hourly electricity demand.

2.3.3. Percentage Improvement

The percentage improvement between the best model $M_{j\text{best}}, j = 1, \ldots, k$ with the other models is computed as follows ([35]):

$$\text{Improvement}(\%) = \left(1 - \frac{\text{Pinball (best model)}}{\text{Pinball (other model)}}\right) \times 100. \tag{14}$$

Equation (14) is used to compute the percentage improvements of the best model developed from the other models.

2.3.4. Prediction Intervals

For each of the models, $M_j, j = 1, \ldots, k$, we compute the prediction interval widths (PIWs), which we shall abbreviate as $\text{PIW}_{ij}, i = 1, \ldots, n, j = 1, \ldots, k$ as follows:

$$\text{PIW}_{ij} = \text{UL}_{ij} - \text{LL}_{ij}, \tag{15}$$

where UL_{ij} and LL_{ij} are the upper and lower limits of the prediction interval, respectively. The analysis for determining the model which yields narrower PIW is done in this study using box and whisker plots, together with the probability density plots. A comparative analysis is done using the prediction intervals based on QRA [36].

2.3.5. Evaluation of Prediction Intervals

A prediction interval with nominal confidence (PINC) of $100(1-\alpha)\%$ is defined as the probability that the forecast $\hat{y}_{t,\tau}$ lies in the prediction interval $(\text{LL}_{ij}, \text{UL}_{ij})$. PINC is given in Equation (16) [37]:

$$\text{PINC} = P\left(\hat{y}_{t,\tau} \in (\text{LL}_{ij}, \text{UL}_{ij})\right) = 100(1-\alpha)\%. \tag{16}$$

Various indices are used to evaluate the reliability of prediction intervals (PIs). In this paper, we use the prediction interval coverage probability (PICP), the prediction interval normalised average width (PINAW) and the prediction interval normalised average deviation (PINAD) that are discussed in Sun et al. [37] and Shen et al. [38]. The PICP is given in Equation (17):

$$\text{PICP} = \frac{1}{m}\sum_{i=1}^{m} \text{l}_{ij}, \tag{17}$$

where m is the number of forecasts and I is a binary variable that is defined as

$$I_{ij} = \begin{cases} 1, & \text{if } y_i \in \left(\text{LL}_{ij}, \text{UL}_{ij}\right), \\ 0, & \text{if otherwise.} \end{cases} \tag{18}$$

The PICP is valid if it is greater than or equal to the predetermined level of confidence [37,38]. The PINAW is an index that is used to check if the required value is covered by the prediction interval and is given as follows [37,38]:

$$\text{PINAW} = \frac{1}{m(\max(y_{ij}) - \min(y_{ij}))}\sum_{i=1}^{m}\left(\text{UL}_{ij} - \text{LL}_{ij}\right), j = 1, \ldots, k. \tag{19}$$

If the PICP is valid and accurate, then the PINAW is usually small [37,38]. However, PINAW can be used to compare different models and then determine the one that possesses the smallest percentage value. Another index which is used to assess the deviation of the target value from the prediction interval is the PINAD, which is given in Equation (20) [37,38]:

$$\text{PINAD} = \frac{1}{m}\sum_{i=1}^{m}\frac{D_{ij}}{\max(y_{ij}) - \min(y_{ij})}, \tag{20}$$

where

$$D_{ij} = \begin{cases} \text{LL}_{ij} - y_{ji}, & \text{if } y_{ji} < \text{LL}_{ij}, \\ 0, & \text{if } \text{LL}_{ij} \leq y_{ji} \leq \text{UL}_{ij}, \\ y_{ji} - UL_{ij}, & \text{if } y_{ji} > \text{UL}_{ij}. \end{cases}$$

2.3.6. Forecast Error Distribution

For each of the models, $M_j, j = 1, \ldots, k$, we extract the residuals $\varepsilon_{tj} = y_{tj} - \hat{y}_{tj}$ and then compute the under and over predictions. Probability density and box plots of forecast errors including summary statistics are used in the analysis of over and under predictions.

2.3.7. Forecast Combination

QRA is based on forecasting the response variable against the combined forecasts which are treated as independent variables. Let $y_{t,\tau}$ be hourly electricity demand as discussed in Section 2.3.1 and let there be M methods used to predict the next observations of $y_{t,\tau}$, which shall be denoted by $y_{t+1}, y_{t+2}, \ldots, y_{t+M}$. Using $m = 1, \ldots, M$ methods, the combined forecasts will be given by

$$\hat{y}_{t,\tau}^{\text{QRA}} = \beta_0 + \sum_{j=1}^{k} \beta_j \hat{y}_{tj} + \varepsilon_{t,\tau}, \tag{21}$$

where $\hat{y}_{tj}$ represents forecasts from method j, $\hat{y}_{t,\tau}^{\text{QRA}}$ is the combined forecasts and $\varepsilon_{t,\tau}$ is the error term. We seek to minimise

$$\arg\min_{\beta} \sum_{t=1}^{n} \rho_\tau \left(\hat{y}_t^{\text{QRA}} - \beta_0 - \sum_{j=1}^{k} \beta_j \hat{y}_{tj} \right). \tag{22}$$

In matrix form, we have

$$\arg\min_{\beta \in \mathbb{R}} \sum_{t=1}^{n} \rho_\tau \left(\hat{y}_t^{\text{QRA}} - x_t^T \beta \right),$$

which reduces to

$$\arg\min_{\beta \in \mathbb{R}^p} \sum_{t: \hat{y}_t^{\text{QRA}} > x_t^T \beta} \tau \left(\hat{y}_t^{QRA} - x_t^T \beta \right) + \sum_{t: \hat{y}_t^{\text{QRA}} < x_t^T \beta} (1 - \tau) \left(\hat{y}_t^{\text{QRA}} - x_t^T \beta \right).$$

The QRA forecasts will be compared with forecasts based on weighted average of the forecasts given in Equation (23)

$$\hat{y}_{t,\tau}^{c} = \sum_{m=1}^{M} \omega_{mt} \hat{y}_{mt,\tau}, \tag{23}$$

where ω_{mt} is weight assigned to the forecast m.

3. Description of the Case Study

The modelling framework discussed in Section 2 is then applied to a real-life data set. Hourly load data from Eskom, South Africa's power utility company is used. The data is from all the sectors of the South African economy, i.e., industrial, commercial, agricultural including the residential sectors. In this study, hourly temperature data from the South African Weather Services is used. The temperature data is from 28 meteorological stations. Other variables (predictors) used are lagged demand at lags 1, 12 and 24; including factor variables, hour = 1, hour = 2, ..., hour = 24; month which takes values, month = 1 for January, month = 2 for February, ..., month = 12 for December; daytype taking values daytype = 1 for Monday, daytype = 2 for Tuesday, ..., daytype = 7 for Sunday, variable holiday which takes value 1 if a day is a holiday and also value 1 for a day before and after a holiday. In addition, a nonlinear trend variable is also used.

4. Empirical Results

4.1. Exploratory Data Analysis

The summary statistics of hourly electricity demand for the sampling period January 2010 to December 2012 is given in Table 1. The distribution of hourly load is non-normal since it is skewed to the left and platykurtic as shown by the skewness value of −0.243 and a kurtosis value of 2.05 given in Table 1.

Table 1. Summary statistics for hourly electricity demand (MW).

Descriptive Statistics	Mean	Median	Max	Min	St. Dev.	Skewness	Kurtosis
Load	27,798	28,496	36,664	18,739	3337	−0.2433	2.050

Figure 1 shows the time series plot of hourly electricity demand together with density, normal quantile to quantile (Q–Q) and box plots that all show departure from normality of the data. The distribution of the sampling data is bimodal.

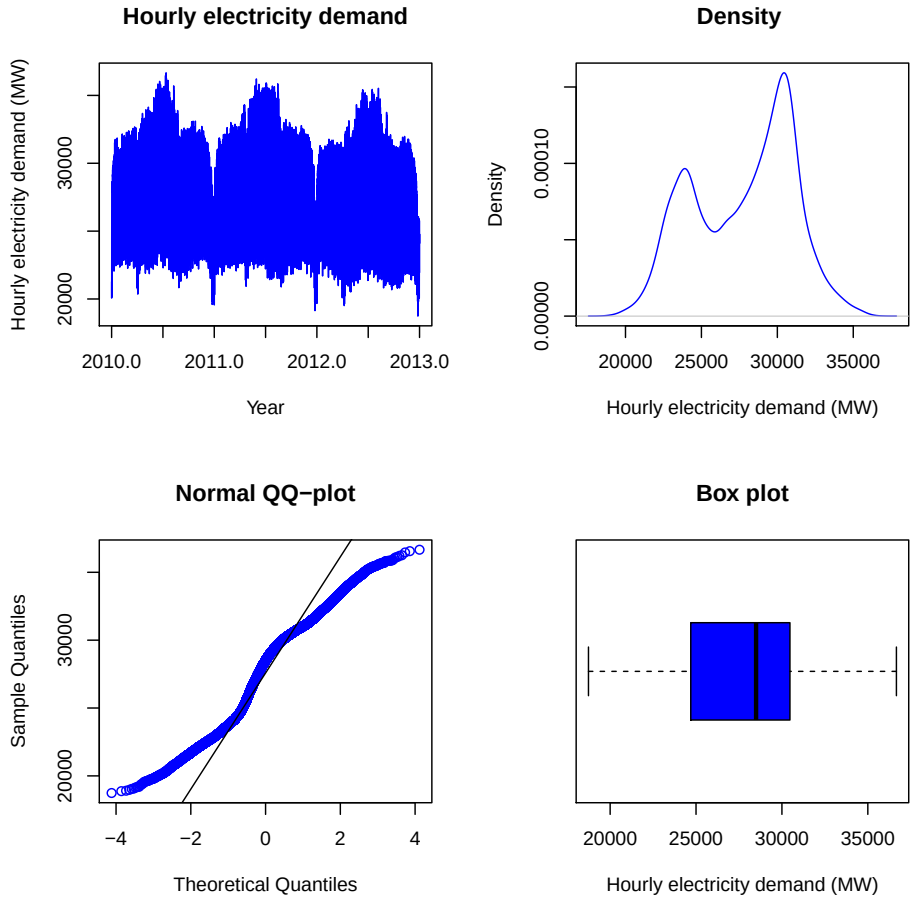

Figure 1. Hourly electricity demand from January 2010 to 31 December 2012.

A plot of hourly electricity demand with a superimposed nonlinear trend is shown in Figure 2. A penalised cubic regression spline $\pi(t) = \sum_{t=1}^{n} \left(y_t - f(x_t)\right)^2 + \lambda \int \left(f''(x)\right)^2 dx$ is used as the nonlinear trend function, with λ as the smoothing parameter and is estimated by generalised cross-validation

(GCV) approach. The fitted values are extracted and used as input values for the nonlinear trend variable in the GAM and AQR models.

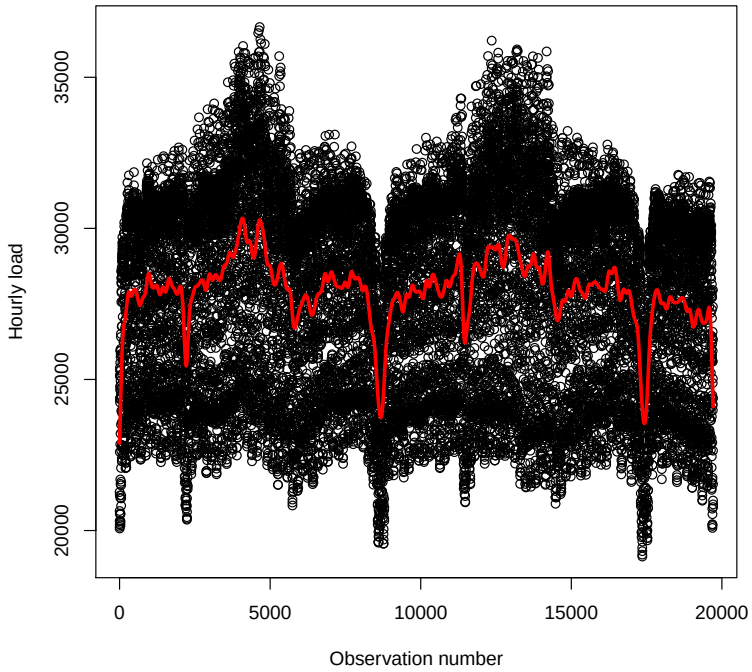

Figure 2. Plot of hourly electricity demand from 1 January 2010 to 31 December 2012 superimposed with a nonlinear trend.

4.2. Forecasting Electricity Demand When Covariates Are Known in Advance

4.2.1. Forecasting Results

The data used is hourly electricity demand from 1 January 2010 to 31 December 2012 giving us $n = 26{,}281$ observations. The data is split into training data, 1 January 2010 to 2 April 2012, i.e., $n_1 = 19{,}708$ and testing data, from 2 April 2012 to 31 December 2012, i.e., $n_2 = 6573$, which is 25% of the total number of observations. The smoothed effect of the variable "hour" which is given in Figure 3 shows that daily peak electricity demand occurs around 7:00 p.m. The period 5:00 p.m. to 9:00 p.m. is then considered as the peak period in which electricity demand is expected to exceed a certain high threshold, which is likely to cause problems for the system operators due to grid instability and severe stress on the system.

The models considered are M_1 (GAM), M_2 (GAMI), which are GAM models without and with interactions, respectively, and M_3 (AQR), M_4 (AQRI) which are additive quantile regression models without and with interactions, respectively. The four models M_1 to M_4 are then combined based on the pinball losses, resulting in M_5 and also combined using QRA, resulting in M_6.

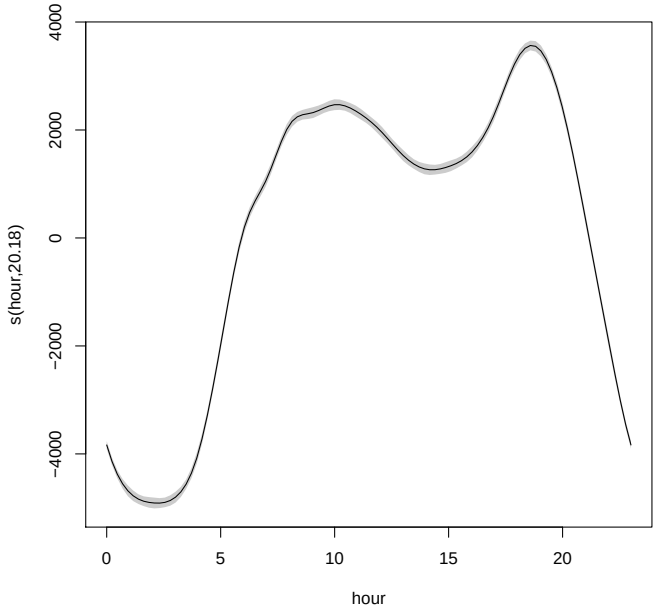

Figure 3. Smoothed effects of variable "hour".

4.2.2. Out of Sample Forecasts

After correcting for residual autocorrelation, we then use the model for out of sample forecasting (testing). A comparative analysis of the models given in Table 2 shows that M_4 is the best model out of the four models, M_1 to M_4, based on the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The forecasts from the four models are then combined based on the pinball losses. The weights assigned to the forecasts from the models M_1 to M_4 are 0.0174, 0.0946, 0.326 and 0.562, respectively. The model for combining the forecasts based on the pinball losses is M_5. Model M_6, i.e., the model based on QRA has the lowest MAE and MAPE values as shown in Table 2. Model 4 has more under predictions compared to over predictions, and M_5 has more over predictions compared to under predictions, while, for model 6, the under and over-predictions are almost the same.

Table 2. Model comparisons.

	M_1	M_2	M_3	M_4	M_5	M_6
RMSE	736.2	662.4	731.5	648.8	596.1	577.7
MAE (NW)	568.7	516.2	549.5	499.7	459.4	445.2
MAPE (%)	2.15	1.93	2.04	1.86	1.70	1.65
Under predictions				3319	3279	3280
Over predictions				3251	3291	3286

Using models M_4, M_5 and M_6, we then compute the average pinball losses. The average losses suffered by the models based on the pinball losses are given in Table 3 with model M_6 having the smallest average pinball loss.

Table 3. Average pinball losses for M_1 to M_6 (2 April 2012 to 31 December 2012).

	M_1	M_2	M_3	M_4	M_5	M_6
Average Pinball loss	284.363	258.087	274.768	249.842	229.723	222.584

In order to test the effectiveness of the forecasting models M_4 to M_6, we present, in Figure 4, box plots of the pinball loses of the models.

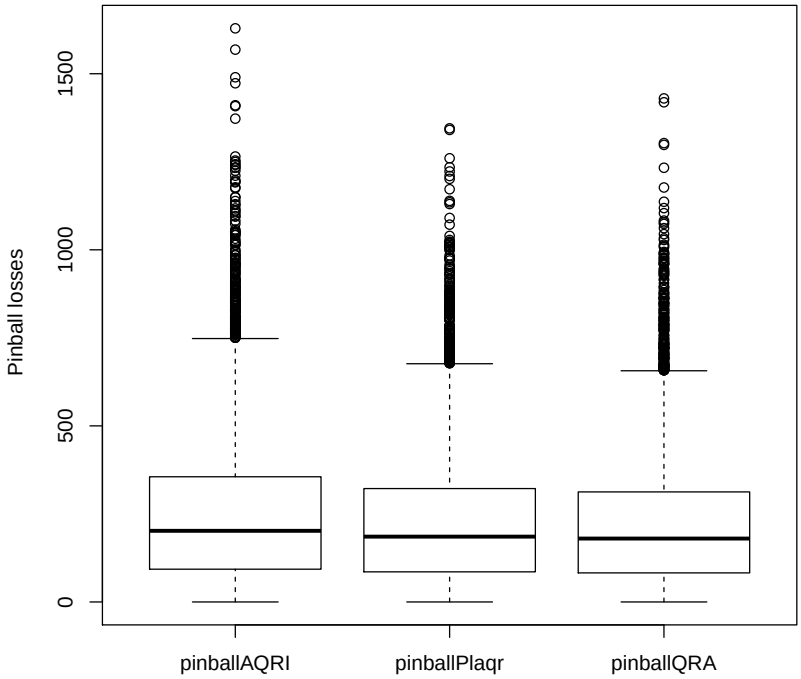

Figure 4. Plot of pinball losses for models M_4 (pinballAQRI), M_5 (pinballPlaqr) and M_6 (pinballQRA) (2 April 2012 to 31 December 2012).

4.2.3. Evaluation of Prediction Intervals

Empirical prediction intervals (PIs) are constructed using the forecasts from the models M_4 to M_6. The constructed PIs are then used to find PIWs, PINAWs, PINADs and calculation of the number of forecasts below and above the PIs from each model. Summary statistics of the PIWs for the models M_4 to M_6 for PINC value of 95% are given in Table 4. The distributions of the PIWs for the three models are all leptokurtic since they are greater than 3. They are all skewed to the right since the values of their skewness are all positive. This shows that heavy-tailed distributions would be appropriate to fit the distributions of the PIWs of the three models. Model M_5 has the smallest standard deviation, which indicates narrower PIW compared to M_4 and M_6.

Table 4. Model comparisons.

	Mean	**Median**	**Minimum**	**Maximum**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Range**
M_4	2100.9	2023	287	5617	686.98	0.7256	3.7217	5330
M_5	2419.1	2435	1883	3560	117.72	1.4898	12.3368	1667
M_6	2300.0	2263	795	4438	418.11	0.6776	4.0304	3643

Boxplots of widths of the PIs for the forecasting models M_4, M_5 and M_6 are given in Figure 5. The figure shows that the PI from model M5 are narrower compared to those from M4 and M6.

Figure 6 shows the density plots of the PIWs of M_4, M_5 and M_6. The distribution of the PIWs of M_5 is bimodal and all the densities show that the distributions are skewed to the right.

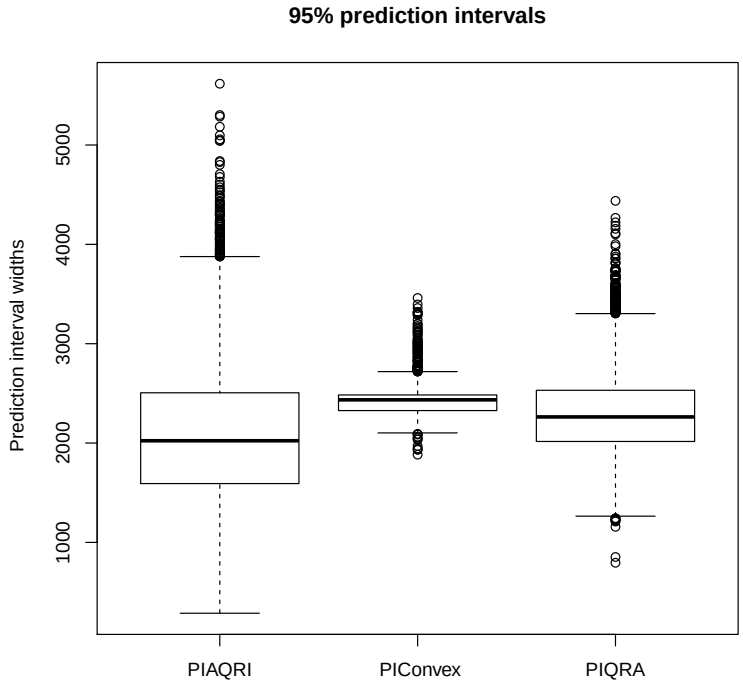

Figure 5. Prediction interval widths for models M_4 (PIAQRI), M_5 (PIConvex) and M_6 (PIQRA).

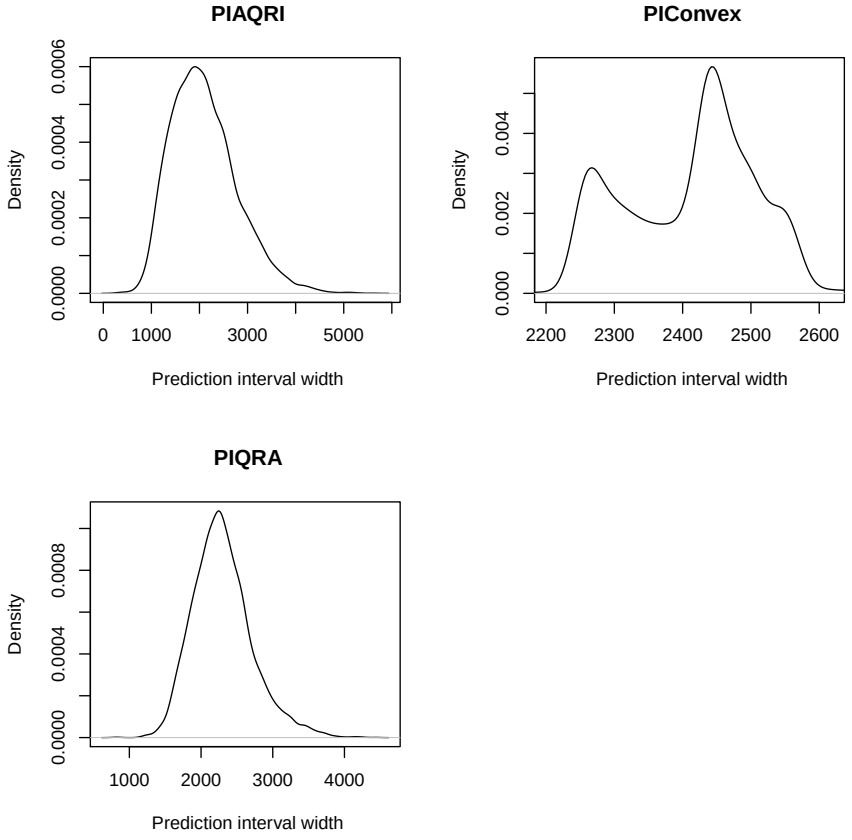

Figure 6. Density plots of the prediction interval widths for models M_4 (PIAQRI), M_5 (PIConvex) and M_6 (PIQRA).

In order to choose the best model based on the analysis of the PIWs, we need to calculate the PICPs, PINAWs and PINADs including a count of the number of forecasts below and above the PIs. This is done for various PINC values, which are 90%, 95% and 99%, respectively. A comparative evaluation of the models using PI indices for PINC values of 90%, 95% and 99% are given in Table 5. Models M_5 and M_6 have valid PICPs for the three PINC values, with M_6 having the highest PICP. Model M_6 has the smallest PINAD values and fewer number of forecasts falling below and above the PIs. Model M_4 has the smallest PINAW value for all three of the PINC values. All three of the models could be used in the construction of PIs. Although M_4 does not give a valid PICP, the PINAW and PINAD are reasonably small. The performance of model M_6 seems to be the best amongst these three models. However, this analysis is not enough and, as a result, we need further analysis using residuals of the three models.

Table 5. Comparative evaluation of models using prediction interval (PI) indices. Below LL = number of forecasts below the lower prediction limit, Above UL = number of forecasts above the upper prediction limit.

PINC	Model	PICP (%)	PINAW (%)	PINAD (%)	Below LL	Above UL
90%	M_4	84.41	10.63	0.2353	462	563
	M_5	90.46	11.73	0.1671	310	317
	M_6	90.80	11.07	0.1347	301	304
95%	M_4	91.19	12.52	0.1186	236	343
	M_5	95.16	14.41	0.0756	156	162
	M_6	95.31	13.70	0.0573	151	157
99%	M_4	97.35	16.43	0.03127	36	138
	M_5	99.1	19.87	0.0110	30	31
	M_6	99.22	17.75	0.005986	31	20

4.2.4. Residual Analysis

Table 6 gives summary statistics of the residuals from the models M_4, M_5 and M_6. Model M_6 has the smallest standard deviation with a median of zero, showing that it is the best model for predicting hourly electricity demand. All three of the models have positive skewness, an indication of a large number of positive errors, which is a reflection of underestimation of predicted hourly electricity demand. Model 6 has the smallest skewness value. A failure to predict high electricity demand is shown by high values of kurtosis [16]. The kurtosis values of all three of the models are greater than 3.

Table 6. Model comparisons.

	Mean	Median	Minimum	Maximum	Standard Deviation	Skewness	Kurtosis
M_4	44.16	7	−2507	3258	647.36	0.3761	3.9266
M_5	28.55	−1	−2520	2690	595.49	0.2442	3.7702
M_6	14.98	0	−2273	2860	577.56	0.1997	3.9223

The error distributions of each of the forecasting models M_4, M_5 and M_6 are given in Figure 7, which shows that the number of positive errors dominates negative errors, an indication that the distribution of errors for each of the three models is positively skewed. Model M_6 is the best fitting model since it has the smallest distribution of errors.

Figure 8 shows the boxplots of the hourly errors from the three models. The range of the errors from M6 is narrower compared to the ones from M4 and M5.

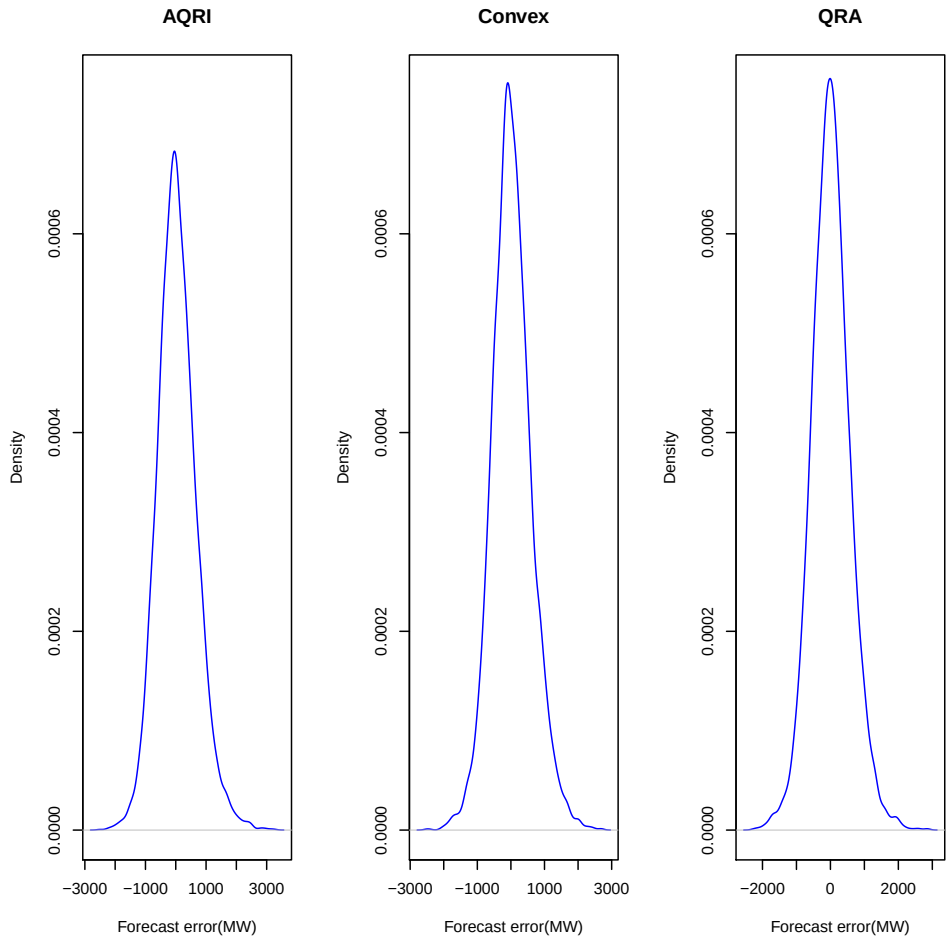

Figure 7. The error distribution of forecasting techniques for M4(AQRI), M5(convex) and M6(QRA).

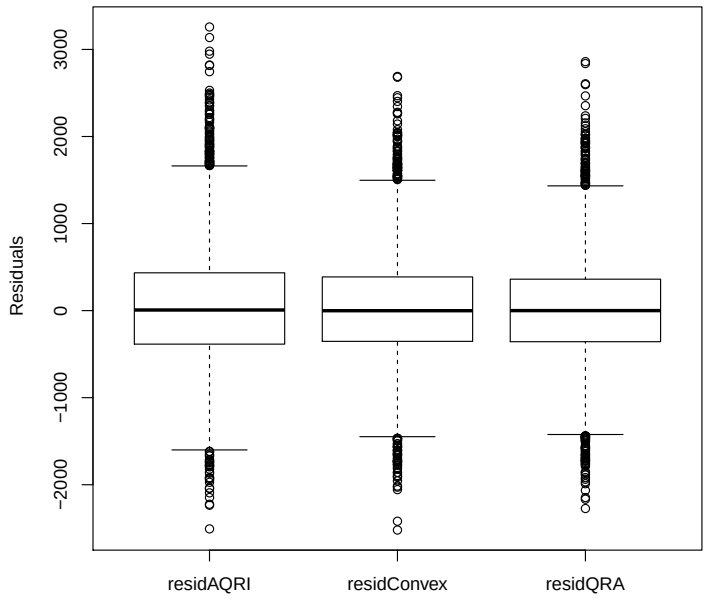

Figure 8. Box plots of residuals from models M_4 (residAQRI), M_5 (residConvex) and M_6 (residQRA).

4.2.5. Plots of out of Sample Forecasts

From the analysis of the PIWs and residual analysis, M_6 is the best fitting model and can be used for predicting hourly electricity demand. The plot of actual demand superimposed with forecasted demand from model M_6 (2 April to 31 December 2012) given in Figure 9 shows that the forecasts follow hourly electricity demand very well.

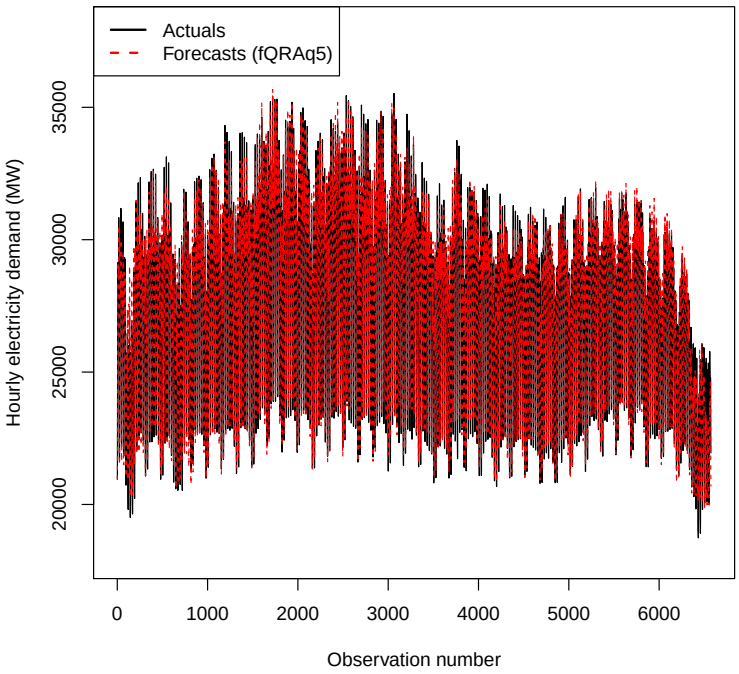

Figure 9. Plot of actual demand superimposed with forecasted demand from M_6 (2 April to 31 December 2012).

The density plots from M_6 (QRA forecasts) and M_5 (convex combination forecasts) models superimposed with actual hourly electricity demand are given in Figure 10. In both plots, the fit of the densities is fairly good.

A summary of the accuracy measures for the months April to December 2012 for each of the first 168 forecasts of each month is given in Table A1 in Appendix A while Appendix B shows in Figures A1–A3 hourly load superimposed with forecasts together with their respective densities.

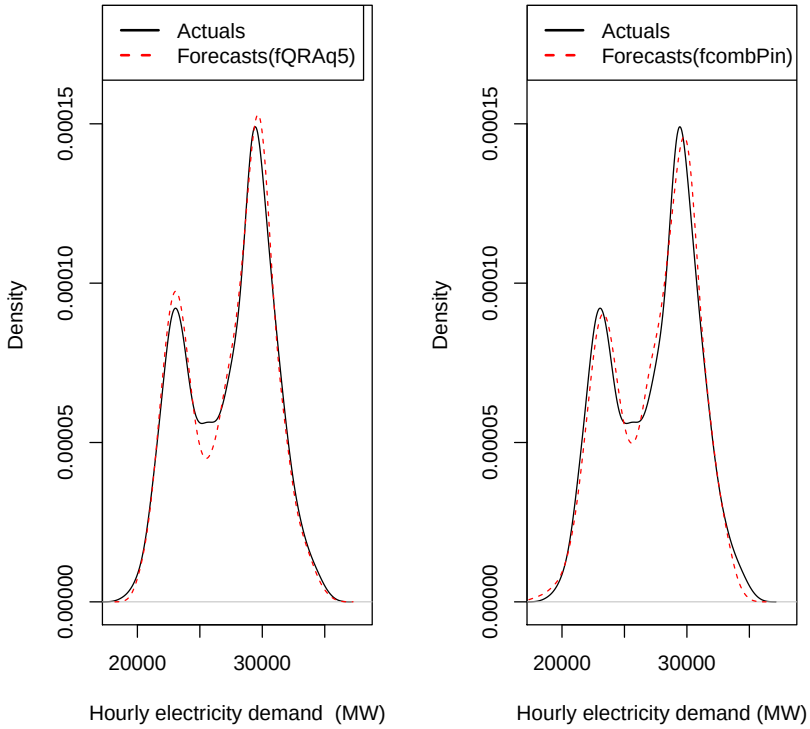

Figure 10. Density plots of actual demand superimposed with density plots from M_6 and M_5 models (2 April to 31 December 2012).

5. Discussion

The modelling approach discussed in this study allows for easy interpretability and accounting for residual autocorrelation in the joint modelling of hourly electricity data. A comparative analysis was then done with the generalised additive models (GAMs). In both modelling frameworks, variable selection was done using Lasso via hierarchical interactions. Four models considered were GAMs and AQR models with and without interactions. The AQR model with pairwise interactions was found to be the best fitting model. The forecasts from the four models were then combined using an algorithm based on the pinball loss (convex combination model) and also using quantile regression averaging (QRA). The AQR model with interactions was then compared with the convex combination and QRA models and the QRA model gave the most accurate forecasts. Except for the AQR model with interactions, the other two models' convex combination model and the QRA model gave prediction interval coverage probabilities which were valid for the 90%, 95% and the 99% prediction intervals. The QRA model had the smallest prediction interval normalised average width and prediction interval normalised average deviation.

6. Conclusions

This study discussed an application of short-term hourly electricity demand forecasting in South Africa using additive quantile regression (AQR) models without and with pairwise interactions which satisfy the strong hierarchy in Lasso via hierarchical interactions. This modelling approach allows for a detailed analysis which goes beyond the performance statistics in forecasting. This approach has merit in that it gives more insight in the developed models.

Author Contributions: Conceptualization, C.S.; Methodology, C.S., M.M.N. and D.M.; Software, C.S., M.M.N. and D.M.; Validation, C.S., M.M.N. and D.M.; Formal Analysis, C.S., M.M.N. and D.M.; Investigation, C.S., M.M.N. and D.M.; Data Curation, C.S., M.M.N. and D.M.; Writing—Original Draft Preparation, C.S.; Writing—Review and Editing, M.M.N. and D.M.; Project Administration, C.S.; Funding Acquisition, C.S.

Funding: This research was funded by the National Research Foundation of South Africa, Grant No. 93613.

Acknowledgments: The authors are grateful to Eskom, South Africa's power utility company for providing the data.

Abbreviations

The following abbreviations are used in this manuscript:

AQR	Additive Quantile Regression
GAM	Generalised additive model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
PI	Prediction Interval
PICP	Prediction Interval Coverage Probability
PINAD	Prediction Interval Normalised Average Deviation
PINAW	Prediction Interval Normalised Average Width
PINC	Prediction Interval with Nominal Confidence
QR	Quantile Regression
QRA	Quantile regression averaging
RMSE	Root Mean Square Error

Appendix A. Summary of the Accuracy Measures for the Months April to December 2012

A summary of the accuracy measures for the months April to December 2012 for each of the first 168 forecasts of each month is given in Table A1. The best forecasts are in October and the worst are in April.

Table A1. Forecast accuracy measures root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) for the forecasts of April to December 2012.

	RMSE	MAE (MW)	MAPE (%)
April	945.5214	781.6429	3.151406
May	620.7605	488.6548	1.891559
June	665.0797	537.5238	1.898156
July	392.0611	329.3393	1.181808
August	642.6158	538.2321	1.903814
September	750.3948	618.5476	2.264714
October	345.0181	271.0595	1.010533
November	394.3301	302.9048	1.146244
December	468.6219	369.5595	1.395704

Appendix B. Hourly Load with Forecasts for the Months April–December 2012

Hourly load superimposed with forecasts for the first 168 forecasts of each month of the months April to December of 2012 together with their respective densities is given in Figures A1–A3.

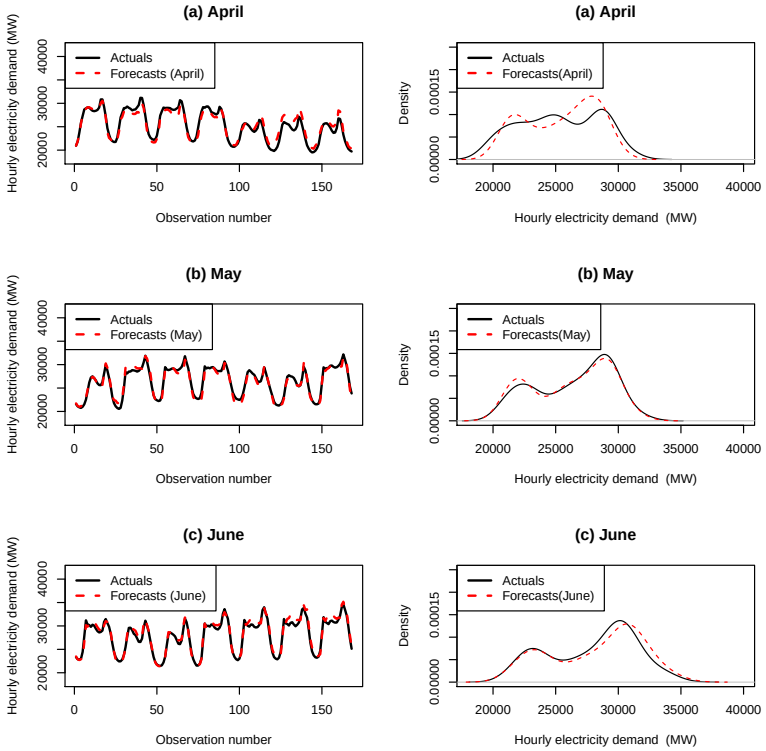

Figure A1. Hourly load superimposed with forecasts for the first 168 forecasts of each month of the months April to June 2012 together with their respective densities.

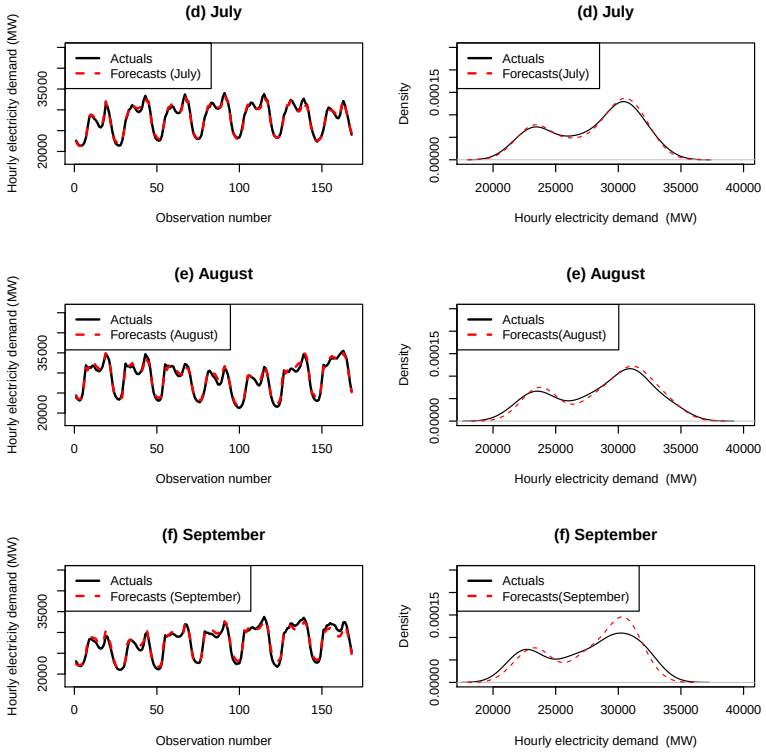

Figure A2. Hourly load superimposed with forecasts for the first 168 forecasts of each month of the months July to September 2012 together with their respective densities.

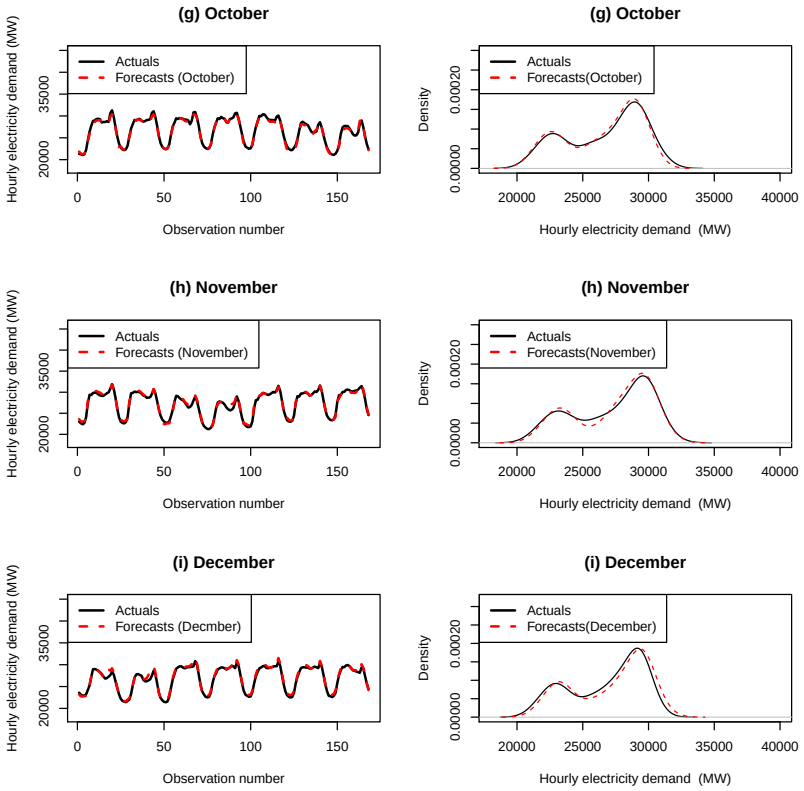

Figure A3. Hourly load superimposed with forecasts for the first 168 forecasts of each month of the months October to December 2012 together with their respective densities.

References

1. Maciejowska, K.; Weron, R. Forecasting of daily electricity prices with factor models: Utilizing intra-day and inter-zone relationships. *Comput. Stat*. **2017**, *30*, 805–819. [CrossRef]
2. Wood, S.N.; Goude, Y.; Shaw, S. Generalized additive models for large datasets. *J. R. Stat. Soc*. **2015**, *64*, 139–155. [CrossRef]
3. Tsay, R.S. *Analysis of Financial Time Series*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley Online Library: Hoboken, NJ, USA, 2005.
4. Dordonnat, V.; Koopman, S.J.; Ooms, M. Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. *Comput. Stat. Data Anal*. **2012**, *56*, 3134–3152. [CrossRef]
5. Soares, L.J.; Medeiros, M.C. Modeling and Forecasting Short-term Electric Load Demand: A Two-Step Methodology. 2016. Available online: https://pdfs.semanticscholar.org/734b/3f6565243912784ad7b1a7421acb7188c9ca.pdf (accessed on 28 December 2016).
6. Ramanathan, R.; Engle, R.; Granger, C.W.J.; Vahid-Araghi, F.; Brace, C. Short-run forecasts of electricity loads and peaks. *Int. J. Forecast*. **1997**, *13*, 161–174. [CrossRef]
7. Fan, S.; Hyndman, R.J. Short-term load forecasting based on a semi-parametric additive model. *IEEE Trans. Power Syst*. **2012**, *27*, 134–141. [CrossRef]
8. Goude, Y.; Nedellec, R.; Kong, N. Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE Trans. Smart Grid* **2014**, *5*, 440–446. [CrossRef]
9. Gaillard, P.; Goude, Y.; Nedellec, R. Additive models and robust aggregation for GEFcom2014 probabilistic electric load and electricity price forecasting. *Int. J. Forecast*. **2016**, *32*, 1038–1050. [CrossRef]
10. Fasiolo, M.; Goude, Y.; Nedellec, R.; Wood, S.N. Fast Calibrated Additive Quantile Regression. 2017. Available online: https://github.com/mfasiolo/qgam/blob/master/draftqgam.pdf (accessed on 13 March 2017).

11. Laouafi, A.; Mordjaoui, M.; Haddad, S.; Boukelia, T.E.; Ganouche, A. Online electricity demand forecasting based on effective forecast combination methodology. *Electr. Power Syst. Res.* **2017**, *148*, 35–47. [CrossRef]
12. Zhang, X.; Wang, J.; Zhang, K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electr. Power Syst. Res.* **2017**, *146*, 270–285. [CrossRef]
13. Boroojeni, K.G.; Amini, M.H.; Bahrami, S.; Iyengar, S.S.; Sarwat, A.I.; Karabasoglu, O. A novel multi-time-scale modelling for electric power demand forecasting: From short-term to medium-term horizon. *Electr. Power Syst. Res.* **2017**, *142*, 58–73. [CrossRef]
14. Khwaja, A.S.; Zhang, X. Anpalagan, A.; Venkatesh, B. Boosted neural networks for improved short-term electric load forecasting. *Electr. Power Syst. Res.* **2017**, *143*, 431–437. [CrossRef]
15. Ekonomou, L.; Christodoulou, C.A.; Mladenov, V. A short-term load forecasting method using artificial neural networks and wavelet analysis. *Int. J. Power Syst.* **2016**, *1*, 64–68.
16. Pappas, S.S.; Ekonomou, L.; Moussas, V.C.; Karampelas, P.; Katsikas, S.K. Adaptive load forecasting of the Hellenic electric grid. *J. Zhejiang Univ. Sci. A* **2008**, *9*, 1724–1730. [CrossRef]
17. Gajowwniczek, K.; Zabkowski, T. Two-stage electricity demand modeling using machine learning algorithms. *Energies* **2017**, *10*, 1547. [CrossRef]
18. Chapgain, K.; Kittipiyakul, S. Performance analysis of short-term electricity demand with atmospheric variables. *Energies* **2018**, *11*, 818. [CrossRef]
19. Divina, F.; Gilson, A; Gomeéz-Vela, F.; Torres, M.G.; Torres, J.F. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies* **2018**, *11*, 949. [CrossRef]
20. Nagbe, K.; Cugliari, J.; Jacques, J. Short-term electricity demand forecasting using a functional state space model. *Energies* **2018**, *11*, 1120. [CrossRef]
21. Chikobvu, D.; Sigauke, C. Regression-SARIMA modelling of daily peak electricity demand in South Africa. *J. Energy S. Afr.* **2012**, *23*, 23–30.
22. Sigauke, C.; Chikobvu, D. Short-term peak electricity demand in South Africa. *Afr. J. Bus. Manag.* **2012**, *6*, 9243–9249. [CrossRef]
23. Sigauke, C.; Chikobvu, D. Peak electricity demand forecasting using time series regression models: An application to South African data. *J. Stat. Manag. Syst.* **2016**, *19*, 567–586. [CrossRef]
24. Bien, J.; Taylor, J.; Tibshirani, R. A lasso for hierarchical interactions. *Ann. Stat.* **2013**, *41*, 1111–1141. [CrossRef] [PubMed]
25. Laurinec, P. Doing Magic and Analyzing Seasonal Time Series with GAM, (Generalized Additive Model) in R. 2017. Available online: https://petolau.github.io/Analyzing-double-seasonal-time-series-with-GAM-in-R/ (accessed on 23 February 2017).
26. Koenker, R.; Bassett, G. Regression quantiles. *Econ. J. Econ. Soc.* **1978**, *46*, 33–50. [CrossRef]
27. Hastie, T.; Tibshirani, R. Generalized additive models (with discussion). *Stat. Sci.* **1986**, *1*, 297–318. [CrossRef]
28. Hastie, T.; Tibshirani, R. *Generalized Additive Models*; Chapman & Hall: London, UK, 1990.
29. Wood, S.N. *Generalized Additive Models: An Introduction with R*; Chapman & Hall: London, UK, 2006.
30. Wood, S.N. *Generalized Additive Models: An Introduction with R*; Chapman & Hall: London, UK, 2017.
31. Sigauke, C. Forecasting medium-term electricity demand in a South African electric power supply system. *J. Energy S. Afr.* **2017**, *28*, 54–67. [CrossRef]
32. Bien, J.; Tibshirani, R. R Package "HierNet", Version 1.6. 2015. Available online: https://cran.r-project.org/web/packages/hierNet/hierNet.pdf (accessed on 22 May 2017).
33. Lim, M.; Hastie, T. Learning interactions via hierarchical group-lasso regularization. *J. Comput. Graph. Stat.* **2015**, *24*, 627–654. [CrossRef] [PubMed]
34. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913. [CrossRef]
35. Abuella, M.; Chowdhury, B. Hourly probabilistic forecasting of solar power. In Proceedings of the 49th North American Power Symposium, Morgantown, WV, USA, 17–19 September 2017.
36. Liu, B.; Nowotarski, J.; Hong, T.; Weron, R. Probabilistic load forecasting via quantile regression averaging of sister forecasts. *IEEE Trans. Smart Grid* **2017**, *8*, 730–737. [CrossRef]

37. Sun, X.; Wang, Z.; Hu, J. Prediction interval construction for byproduct gas flow forecasting using optimized twin extreme learning machine. *Math. Probl. Eng*. **2017**. [CrossRef]
38. Shen, Y.; Wang, X.; Chen, J. Wind power forecasting using multi-objective evolutionary algorithms for wavelet neural network-optimized prediction intervals. *Appl. Sci*. **2018**, *8*, 185. [CrossRef]