Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem

Moshoko Emily Lebotsa[1], Caston Sigauke[*1], Alphonce Bere[1], Robert Fildes[2], John E Boylan[2]

[1]Department of Statistics, University of Venda, Private Bag X5050
Thohoyandou, 0950, South Africa
[2]Lancaster Centre for Marketing Analytics and Forecasting, Department of Management Science
University of Lancaster, United Kingdom

March 28, 2018

Abstract

Short term probabilistic load forecasting is essential for any power generating utility. This paper discusses an application of partially linear additive quantile regression models for predicting short term electricity demand during the peak demand hours (i.e. from 18:00 to 20:00) using South African data for January 2009 to June 2012. Additionally the bounded variable mixed integer linear programming technique is used on the forecasts obtained in order to find an optimal number of units to commit (switch on or off. Variable selection is done using the least absolute shrinkage and selection operator. Results from the unit commitment problem show that it is very costly to use gas fired generating units. These were not selected as part of the optimal solution. It is shown that the optimal solutions based on median forecasts ($Q_{0.5}$ quantile forecasts) are the same as those from the 99^{th} quantile forecasts except for generating unit g_{8c}, which is a coal fired unit. This shows that for any increase in demand above the median quantile forecasts it will be economical to increase the generation of electricity from generating unit g_{8c}. The main contribution of this study is in the use of nonlinear trend variables and the combining of forecasting with the unit commitment problem. The study should be useful to system operators in power utility companies in the unit commitment scheduling and dispatching of electricity at a minimal cost particularly during the peak period when the grid is constrained due to increased demand for electricity.

Keywords: Lasso, mixed integer linear programming, quantile regression, short term peak load forecasting, unit commitment.

1 Introduction

1.1 Context

Load forecasting is deemed important as it provides prediction of electricity needed for future consumption [10] and is essential for every electricity utility in order to maintain the demand and supply

[*1] Corresponding author: Tel.: +27 15 962 8135. Email address: csigauke@gmail.com (C. Sigauke)

balance [34]. Load forecasts are also required for many other purposes such as system security, rate design, revenue projection as well as for scheduling activities [19,20,47]. Moreover, load forecasts are usually assessed periodically, ranging from hours, days, and weeks, months, up to a year, or even longer. However, the focus in this paper is on the short term load forecasts (STLF) which range from few hours up to a week [1]. Accurate and efficient forecasts are necessary for any power supplying company as they help in preventing problems such as under loading (i.e. load shedding, blackouts, etc.) or overloading (i.e. producing more capacity than needed), which are very costly to suppliers [34]. Accurate forecasts also help with the unit commitment decisions by reducing production costs [2].

1.2 An overview of the literature on load forecasting

Short term load forecasting (STLF) has been receiving a lot of attention [10]. Almost every year different models for STLF are developed, applied, reviewed and published. Short term forecasts are used to estimate the load demand up to a week ahead of a schedule [16]. Many utilities rely on them because they are useful when it comes to daily operation and scheduling of power systems [31]. For instance, obtained forecasts are used to schedule generation as well as transmission of units to consumers. A brief review on weather station and variable selection methods together with other methods used in short term forecasting as well as on the unit commitment problem (UC) follows.

South Africa (SA) is among developing countries around the world faced with electricity crisis. On the African continent, SA is considered as the most industrialized country with the highest electricity consumption [33]. The electricity demand pattern is complex in nature, due to the presence of several factors such as the economic, environmental, weather conditions and calendar effects [9,10]. However, many researchers use temperature in their forecasting models as it is one of the major drivers of electricity demand. A South African study that uses temperature is that of Chikobvu and Sigauke [5], their main aim being to assess its impact on the daily peak electricity demand. The results show that electricity demand is sensitive to cold weather in the country. Another study that examines meteorological effects is that of Taylor and MacSharry [42], in which the seasonality patterns and their effects were studied from European countries.

Other than meteorological effects, time factors also have a huge impact on electricity demand. Time factors include day-of-week, yearly seasonality, time of the day, etc. and generally are known as calendar effects and their impact in short term load forecasting are discussed in detail by [9]. For instance, load curves were used to explain electricity consumption patterns for different hours during the day and based on the load curves these two authors concluded that calendar effects mostly determine the daily life style of the consumers, i.e. highest peak load in the evening [9]. Moreover, the peak load demand is the most studied as highlighted by Hinman and Hickey [19].

Weather effects are generally known to have a major impact on load demand thus making selection of weather stations an important aspect to consider in load demand forecasting. The fact that station selection depends on the geographical location, climate as well as the industrial structure of a country or region of study was highlighted by Janicki [26]. One of the few papers that focused more on the weather stations selection is that of Hong et al. [22]. In this paper the authors propose a new weather station selection framework which they applied to two of the electricity utilities in the United States (US). Their focus was more on developing an algorithm that will help in determining the number of weather stations to consider as well as selecting the stations to use for a particular electricity utility.

Another important aspect in model building is to be able to identify the best set of input variables. For instance, over the past years many different variable selection methods have been used to select important variables to include when forecasting generally. There are many different available methods that can be used to select important variables in a study. These include the dimension reduction, shrinkage as well as the subset selection methods [27]. The most commonly used methods are subset selection which include techniques such as the stepwise criterion and many more.

A study that applied one of the variable selection methods mentioned above is that of Fan and Hyndman [10]. The authors in this paper used a stepwise variable selection technique was used to select a combination set of variables that were then used in models that were used to predict 48 half-hourly load demand in Australia. The authors used the stepwise backward selection method which involves including all the variables at first and removing one at a time while keeping the rest in the model. Then they continued by checking the performance of the model with removed variable using mean absolute percentage error (MAPE). Thus if the model resulted with a lower MAPE value, it was then selected as the best model for that period.

Although stepwise variable selection methods have been and are still widely used, they have limitations. Some of these limitations include their inability to deal with multicollinearity and less computational efficiency. These limitations are mostly ignored in many studies. For instance, according to Olusegun et al. [35] the stepwise selection criterion selects variables based on the correlation between a response and a set of explanatory variables only. Somehow the correlation within the explanatory variables is not considered which can lead to having one or more variables with the same characteristics in the model, due to their multicollinearity. Shrinkage methods are better equipped to handle multicollinearity.

Several methods have been proposed in the literature on variable selection in regression based models. In this study we use least absolute shrinkage and selection operator (Lasso) via hierarchical interactions. For modelling time series data with multiple seasonalities Bien et al. [3] developed a Lasso for hierarchical pairwise interactions in regression based models. In another study a method which satisfies the strong hierarchy for learning linear interaction models is presented in [32]. Results show that the developed method is comparable with past methods. The method caters for both continuous and categorical variables.

In a study short term load forecasting model based on the semi-parametric additive technique were developed [10]. The model was used to investigate the relationship between the load demand and the standard explanatory variables: calendar effects, temperature effects, as well as lagged load observations were included in the model to forecast the half-hourly load demand up to a week ahead. Another study also proposed the use of a more general semi-parametric technique known as generalized additive model (GAM) [34]. The GAM model was first developed by Hastie and Tibshirani [17] and discussed in detail in Wood [44]. These authors used the model to forecast the French hourly load data for over 5 years. The focus was to model effects that drove the French load consumption and also to compare the proposed model with the operational one. The effects included weather conditions, economic growth, weekly and yearly seasonality. The empirical results showed that the proposed model performs better and was able to capture the effects that the French's operational model could not capture [34]. Using functional approximation, [12] developed a day-ahead hourly electricity forecasting model. The developed model uses weather forecasts and also captures temporal patterns. The model provides accurate forecasts as well as prediction error bands which are narrow.

On the other hand, quantile regression (QR) models are used a lot in forecasting by different energy sectors, e.g. wind power forecasting, price forecasting, etc. A recent study [14], introduced an optimal forecast quantile regression (OFQR) model which was used to forecast the annual peak electricity demand in 32 zones in United States (US). The main aim was to compare the relative performance of the OFQR model with the ordinary least square (OLS) regression model, which is a standard method used by almost every utility. The OFQR model provided more accurate forecasts compared to OLS.

In another study, a QR model was also used to forecast electricity demand [41]. The data used in their study was collected from 3639 households in Ireland at both aggregated and disaggregated levels. The proposed QR model was compared with three other benchmark methods. Other authors also developed additive quantile regression models for forecasting both probabilistic load and electricity prices as part of the global energy forecasting competition of 2014 (GEFCom2014) [13]. A summary of the methods used in GEFCom2014 are given in [20]. The proposed new methodology of [13] ranked first in both tracks of the competition. The work done by [13] is extended by Fasiolo et al. [11] who developed fast calibrated additive quantile regression models. To implement the developed models, [11] developed a new r statistical package "*qgam*". The same covariates used in [13] were also used in [11]. In both papers variable selection techniques are not discussed. In another study [18] used kernel support vector quantile regression and copula theory for short-term load probability density forecasting. Two criteria for evaluating the accuracy of the prediction intervals are proposed, the prediction interval normalized average width (PINAW) and the prediction interval coverage probability (PICP). Results from this study show that the Gaussian kernel gives the most accurate forecasts compared to the linear and polynomial kernels respectively. In a more recent study, [48] developed a Gaussian process quantile regression model for short-term load probability density forecasting. The authors argue that this modelling framework provides accurate point forecasts as well as giving probabilistic descriptions of the prediction intervals.

The present study discusses an application of partially linear additive quantile regression (PLAQR) models in forecasting hourly electricity demand during the peak period (i.e from 18:00 to 20:00) in South Africa (SA). PLAQR models are a combination of generalized additive models (GAMs) developed by [17] and quantile regression (QR) models [29,30] where the conditional quantile function comprises a linear parametric component and a nonparametric additive component [23]. Among the first to introduce partially linear models include among others Engle et al. [8] who analyzed the relationship between electricity usage and temperature. A two-step approach for estimating a PLAQR model is discussed in Hoshino [23] and applied to a real data set. In a another study [28] discussed an application of double-penalized quantile regression partially linear additive models. A simulation study and an application to a real data set were used to evaluate the developed models. In a related study, Bayesian partially linear additive quantile regression models are used in simulation studies by [24]. To summarize, the state-of-the-art in modelling is to use hybrid models.

1.3 Literature review on the unit commitment problem

In the literature very few papers combine short term load forecasting with the unit commitment (UC) problem. The management of every electricity utility makes decisions of UC based on the forecasts obtained [16]. Hong and Fan [20] defines UC problem as an optimization problem that determines an optimal number of generating units to be scheduled in order to meet the load demand at a particular time. The main objective of UC is to determine which generating units to commit (switching on or off) in order to meet electricity demand with minimal costs of production and appropriate reserve requirements [45]. Such an application is useful as it helps prevent losses and minimize fuel consumption.

In a study by Kurban and Filik [31], a method that combined short term load forecasting with UC in order to reduce production costs in one of the thermal plants based in Kutahya region, Turkey was proposed. The authors proposed two models which were used to forecast electricity demand and then used the forecasts obtained to find solutions to a UC problem using the Lagrange relaxation method. The empirical results indicated that accurate load forecasting is essential for UC.

In another study considered the stochastic model for the long term solution on security-constrained unit commitment (SCUC) problem is that of Wu et al. [46]. SCUC is an "extension of the basic UC problem that includes additional factors such as fuel, emission and transmission constraints" [45]. These factors were not included in the early stochastic problems (SP) [6]. The proposed model was used to capture the uncertainties of generation components (i.e. generation units and transmission lines) as well as the inaccuracy of the load forecasts.

In order to meet forecast demand it is important to plan in advance the start-up and shut down schedules of power generating units. This has to be done in an optimal way so as to minimize total generation and start-up costs while ensuring that load demand and reserve margin requirements are met. This calls for probabilistic load forecasts which help system operators in planning the scheduling of generating units. A review that surveys the latest optimization methods to solve the UC problem in deterministic and stochastic cases is presented in Saravanan et al. [38]. The paper presented the planning of the running time for different production units so as to meet the constraints under various scenarios. But the authors underlined the fact that it is very difficult to meet all of the constraints in one optimization technique. They also categorized the optimization methods into conventional and non-conventional ones. They focused on the novel hybrid algorithms through classical and non-classical methods. Finally, the authors tabulated UC papers for the last decade, in the interest of new researchers in the field, and briefly discussed a variety of techniques such as hybrid Lagrangian relaxation, particle swarm optimization (PSO) and hybrid genetic algorithms.

A review of past and present findings was done in order to provide a detailed understanding of the UC [45]. The UC aims at scheduling the most cost-effective association of electricity production from a couple of units, during a specific period of time, under some units and transmission restrictions [45]. The author underlined that the UC is a large-scale sequential problem often involving a large number of generating units, making it difficult to obtain, in a reasonable period of time, an optimal solution. The paper affirmed that on the US markets, the most used, among several optimization techniques, are mixed integer linear programming (MILP) and Lagrangian relaxation (LR). The two approaches are briefly discussed and compared in the paper. The author concluded in admitting that the UC problem is far from fully solved, and many researchers preferred new hybrid techniques that combine different optimization techniques. An open software components used in short-term load forecasting combined with the unit commitment problem was presented by Short et al. [40]. The empirical results showed a reduction in costs and improved efficiency as a result of balancing market interactions in the presence of prediction inaccuracies in the proposed modelling framework.

1.4 Contributions

From the literature review discussed above the contributions made in this study are as follows. Firstly, the methodology proposed in this study uses nonlinear trend variables in the developed partially linear additive quantile regression models. It is shown that the inclusion of these variables significantly improves the forecast accuracy. Secondly weather stations in this study are selected based on cluster analysis. They were selected in such a way that they represent the different parts of South Africa (SA). For each cluster, the data were aggregated to get the maximum, minimum and average daily temperature. Coastal and inland temperature data were also used as separate variables in the models. The analysis was done on the overall aggregated as well on the non-aggregated temperature data set. The third contribution involves the use of the bounded variable mixed integer linear programming (BVMILP) in solving the unit commitment problem during the peak period.

2 Models

Partially linear additive quantile regression (PLAQR) models including variable selection and estimation of parameters are discussed in this section.

2.1 Partially linear additive quantile regression models

Generalized additive models (GAMs) which allow flexibility in modelling of predictors as a sum of smooth functions were developed by [17]. Quantile regression (QR) on the other hand was developed by [29] as a modelling framework for estimating the conditional median, including the full range of other conditional quantiles. The partially linear additive quantile regression models are a combination of GAMs and QR models. A PLAQR model is given in Eq. (1) [23]:

$$y_{th} = \beta_{0h,\tau} + \sum_{j=1}^{p_1} s_{jh,\tau}(x_{thj}) + \sum_{j=1}^{p_2} \beta_{hj,\tau} z_{thj} + \varepsilon_{th,\tau}, \tau \in (0,1) \tag{1}$$

where $y_{th},\ t = 1, ..., n$ is the response variable which is electricity demand on day t at hour h, $p = p_1 + p_2$ is the total number of input variables which includes linear and non linear variables, $h = 1, ..., 24$, x_{thj} are continuous variables, $s_{jh,\tau}$ are smooth functions, z_{thj} are linear variables, $\beta_{hj,\tau}$ are parameters and $\varepsilon_{th,\tau}$ is the quantile error term. Eq. (1) can be written in matrix form as:

$$\mathbf{Y} = \mathbf{s}_\tau(\mathbf{X}) + \mathbf{Z}^{\mathbf{T}}\beta_\tau + \varepsilon_\tau, \quad \tau \in (0,1). \tag{2}$$

Let $\mathbf{Y} = F(\mathbf{X}, \mathbf{Z}, \tau)$ then $\tau = P(\mathbf{Y} \leq F(\mathbf{X}, \mathbf{Z}, \tau)$. The parameter estimates of Eq. (2) are obtained by minimizing the following function

$$Q_{Y|X,Z}(\tau) = \sum_{i=1}^{n} \rho_\tau \left(Y_i - Z_i^T \beta - s(X) \right), \tag{3}$$

where $\rho_\tau(u) = \{\tau \boldsymbol{I}(u \geq 0) + (1-\tau)\boldsymbol{I}(u < 0)\} |u| = \{\tau - \boldsymbol{I}(u < 0)\} u$ is the quantile loss function. Since the residuals, $\varepsilon_{th,\tau}$ of the hourly load models are autocorrelated it is important that the autocorrelation is significantly reduced before the models are used for forecasting. In this study we use the following procedure in modelling residual autocorrelation.

1. We estimate the parameters in Eq. (3) and extract residuals, ε_{th}. An appropriate SARIMA(p,d,q) $\times$ (P,D,Q)[s] is then fitted,

2. The fitted values of the residuals of the SARIMA(p,d,q) $\times$ (P,D,Q)[s] are subtracted from y_{th} to get y_{th}^* and we then regress y_{th}^* on the covariates,

3. The residual autocorrelation in the new model is then checked. If the residuals are still autocorrelated the process is repeated until the desired results of uncorrelated errors are achieved.

For data which exhibits seasonality, [25] propose that the SARIMA(2,0,0)(1,0,0)[s] model be used as an initial model for modeling the residuals in step (1) above.

2.2 Data and variables

Let y_{th} denote electricity demand (measured in MegaWatts (MW)) on day t at hour h, where $t = 1, ..., n$ and $h = 1, ..., 24$ with the corresponding covariates $x_{th1}, x_{th2}, ..., x_{thp}$ as discussed in Section 2.1. In this study the following hours (hrs) are considered 18:00, 19:00 and 20:00, i.e. h =18:00, 19:00 and 20:00. Daily load profiles in South Africa show that it is during this period that peak demand is experienced [39, among others]. The variables used in this study are the load at 18:00, 19:00 and 20:00 hrs which is used as the response variable and the predictor variables are calendar effects, temperature and lagged demand. Lag1 and Lag2 simultaneously represent the lagged electricity demand for the same hour of the first and second previous days. Both the lags are differenced in order to make the series of the electricity demand stationary. According to [10] the inclusion of lagged demand effects reduce autocorrelation, however it does not completely remove it. DPED, AED were used to represent daily peak electricity demand and the average electricity demand for the past 24 hours.

The calendar variables considered in this study are denoted as Daytype, DBH, DH, DAH, representing the day of the week, the day before a holiday, day holiday, day after holiday respectively. Daytype is coded as 1 for Monday, 2 for Tuesday up to 7 for Sunday while the variable month is coded as 1 for January, 2 for February up to 12 for December. The variables DBH, DH and DAH were coded as follows, DBH takes value 1 if it is a day before a holiday and 0 otherwise, and similarly DH takes value 1 if the day is a holiday and 0 otherwise and DAH takes value 1 if its a day after a holiday and 0 elsewhere. Hourly temperature data used in this paper is from 28 South African weather stations. Initially all the weather stations in the country were considered and those which were selected for the study were in such a way that they represent the whole country. For the purposes of this study the selected stations were split into two main thermal regions, i.e. coastal and inland. This means that the aggregated coastal and inland temperature variables are considered in this paper including the overall aggregated temperature for the whole country which are are used for comparison sake.

The temperature variables are average daily coastal temperature (ADTC), average maximum and minimum coastal temperature (maxTC and minTC), average minimum, average maximum, average daily inland temperature (minTI, maxTI and ADTI), average minimum of coastal and inland temperatures (AminTCI), average of average daily coastal and inland temperatures (AADTCI), average maximum of coastal and inland temperatures (AmaxTCI), difference between average minimum of coastal and inland temperatures (DminTCI), difference between average maximum of coastal and inland temperatures (DmaxTCI), difference between average of average daily coastal and inland temperatures (DADTCI). A nonlinear trend variable (noltrend) is also used as a covariate. This trend variable is determined by fitting a penalized cubic smoothing spline to the response variable. The fitted values are then extracted and used for the nonlinear trend variable. The data is for the period January 2009 to June 2012. The data for the period January 2009 to December 2011 is used for training, while the remaining data i.e. January to June 2012 used for testing.

2.3 Variable selection, parameter estimation and forecast combination

2.3.1 Variable selection

Variable selection is done using the least absolute shrinkage and selection operator (Lasso) via hierarchical pairwise interactions [3, among others]. Introduced by [43], Lasso penalizes the absolute size of the regression coefficients. It is a useful variable selection method especially when dealing with highly correlated predictors.

Let y_{th} be electricity demand as defined in Eq. (1), with corresponding covariates $z_{th1}, ..., z_{thp}$ in which pairwise interactions are allowed between the predictors. In this study quadratic interactions are not allowed, i.e. $z_{thi} \times z_{thk}$ where $i = k$. The model is given as [3]:

$$y_{th} = \beta_{0h,\tau} + \sum_i \beta_{hi,\tau} z_{thi} + \sum_{i \neq k} \alpha_{ik} z_{thi} z_{thk} + \varepsilon_{th,\tau}. \tag{4}$$

The Lagrangean form of the strong hierarchical Lasso is ([3]):

$$\min_{\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p, \alpha \in \mathbb{R}^{p \times p}, \alpha = \alpha^T} h(\beta_0, \beta, \alpha) + \lambda \sum_i \max \left\{ |\beta_i|, \|\alpha_{ik}\|_1 \right\} + \frac{\lambda}{2} \|\alpha\|_1 \tag{5}$$

where $h(\beta_0, \beta, \alpha)$ is a quadratic loss function given as

$$h(\beta_0, \beta, \alpha) = \frac{1}{2} \left(\sum_{t=1}^{n} y_{th} - \beta_{0h} - z_{th}^T \beta_h - \frac{1}{2} z_{th}^T \alpha z_{th} \right)^2 \tag{6}$$

Whilst Eq. (4) shows interactions for linear variables, it should be noted that the interactions can be applied to nonlinear variables as well.

2.3.2 Parameter estimation

An adaptation of the Barrodale and Roberts algorithm for ℓ_1-regression will be used to estimate the PLAQR model parameters. The simplex approach to solving the general ℓ_1-regression problem [30,36] is given as follows:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(Y_i - Z_i^T \beta - s(X) \right) \tag{7}$$

This is then reformulated as a linear programming problem. Two artificial variables, $u_i, v_i, i = 1, ..., n$ are introduced to represent the positive and negative parts of the vector of residuals. This results in a new problem.

$$\min_{\beta, u, v \in \mathbb{R} \times \mathbb{R}_+^{2n}} \left\{ \mathbf{1}^T u + \mathbf{1}^T v | y = Z\beta + u - v, (u, v) \in \mathbb{R}_+^{2n} \right\}, \tag{8}$$

where $\mathbf{1}$ is a vector of ones. The dual formulation of Eq. (8) is

$$\max \left\{ y^T \alpha | \boldsymbol{Z} \alpha = 0, \alpha \in [-1, 1]^n \right\} \tag{9}$$

and setting $b = \alpha + \frac{1}{2}\mathbf{1}$ results in

$$\max \left\{ y^T b | \boldsymbol{Z} b = \frac{1}{2} Z^T \mathbf{1}, b \in [0, 1)^n \right\}. \tag{10}$$

2.3.3 Forecast combination

Combining forecasts improves the forecast accuracy [7,13, among others]. Let K denote the number of methods used to forecast the response variable, then the combined forecasts will be calculated as follows:

$$f_{th} = \sum_{k=1}^{K} \omega_{kth} y_{kth} \tag{11}$$

where ω_{kth} is the weight. The pinball loss function is used as an evaluation criterion. In this study, the r package 'opera' developed by [13] is used for combining the forecasts.

2.4 Error measures for probabilistic forecasting and evaluation of methods

Scoring rules are used to assess and compare the probabilistic models in this study. A scoring rule assigns a penalty score $S(y, F)$ where y is the observation used for forecast evaluation and F is the forecast distribution [15]. A smaller score corresponds to a better forecast. This study uses three of the commonly used error measures in probabilistic forecasting which are the continuous rank probability score (CRPS), logarithmic score (LogS) and the quantile loss function also known as the pinball loss function.

2.4.1 Continuous rank probability score

The CRPS measures the distance between the predicted and the observed cumulative density functions (CDFs) of scalar variables [15].

$$\mathrm{CRPS}(y, F) = \int_0^1 \mathrm{QS}_\tau \left(F^{-1}(\tau), y\right) d\tau \tag{12}$$

where F is the forecast distribution and QS_τ is the quantile score given as:

$$\mathrm{QS}_\tau \left(F^{-1}(\tau), y\right) = 2\left(\mathrm{I}\left[y \leq F^{-1}(\tau)\right] - \tau\right)\left(F^{-1}(\tau) - y\right) \tag{13}$$

2.4.2 Logarithmic score

The logarithmic score (LogS), is given by

$$\mathrm{LogS}(y, F) = -\mathrm{log} f(y) \tag{14}$$

where f is the density function for the forecast distribution.

2.4.3 Quantile loss function

The quantile loss function also known as the pinball loss function is given by:

$$L(\hat{Q}_{Y|X,Z}(\tau)) = \begin{cases} \tau(y - \hat{Q}_{Y|X,Z}(\tau)) & \text{if } y \geq \hat{Q}_{Y|X,Z}(\tau) \\ (1-\tau)(\hat{Q}_{Y|X,Z}(\tau) - y) & \text{if } y < \hat{Q}_{Y|X,Z}(\tau) \end{cases} \tag{15}$$

where $\hat{Q}_{Y|X,Z}(\tau)$ is the quantile forecast and y is the observation used for forecast evaluation.

2.4.4 Skill score (improvement rate)

The performance of the methods against the best methods for each of the peak hours is calculated using Eq. (16)

$$\mathrm{Improvement}(\%) = \left(1 - \frac{\mathrm{Pinball(best\ method)}}{\mathrm{Pinball(other\ method)}}\right) \times 100 \tag{16}$$

where Pinball(best method) denotes the pinball loss (quantile loss) of the best method and similarly Pinball(other method) is the pinball loss of the other method.

2.5 Unit commitment

In this study the bounded variable mixed integer linear programming (BVMILP) method is used in solving the unit commitment problem. The BVMILP is a special class of linear programming technique where some of the variables are restricted to integer values but the rest are ordinary continuous variables and all variables are bounded. The present study intends to demonstrate how the forecasts can be used in solving the UC problem.
Let P_{Gi}^{ht} be load from generating unit i, $i = 1, ..., m$ at hour h, h =18:00,19:00,20:00 on day t, $t = 1, ..., n$; P_{G}^{ht} national system load at hour h on day t; $P_{Gi(\min)}^{ht}$ the lower limit of the unit power output; $P_{Gi(\max)}^{ht}$ the upper limit of the unit power output; x_i^{ht} the 0-1 variable (In this study we will assume that during the peak period all units are up, i.e $x_i^{ht} = 1$ for all units); F_{si} the start up cost of unit i at hour h (in this study we will assume that the start up cost is zero); P_R^{ht} the power reserve at hour h on day t; F_i the average production cost of unit i (cost/MW). In this study we are going to use fuel cost to represent average production cost per megawatt produced.
Generating units are classified into 13 coal-fired, 1 nuclear, 4 gas, 2 pumped storage and 2 hydroelectric. These will be denoted as: $g_{1c}, g_{2c}, ..., g_{13c}, g_{14n}; g_{15g}, ..., g_{18g}, g_{19p}, g_{20p}, g_{21h}, g_{22h}$. As stated by [45], the main objective of the UC is to minimize cost of generating units over a particular period. Generally, we minimize:

$$\min \sum_{h=1}^{H} \sum_{i=1}^{m} \left[F_i \left(P_{Gi}^{ht} \right) x_i^{ht} + F_{si}(ht) x_i^{ht} \right] = F \left(P_{Gi}^{ht}, x_i^{ht} \right) \tag{17}$$

The constraints are:
Load balance equation

$$\sum_{i=1}^{m} P_{Gi}^{ht} x_i^{ht} = P_D^{ht}, h = 18, 19, 20, \ t = 1, ..., n \tag{18}$$

Generator power output limits

$$x_i^{ht} P_{Gi\min} \leq P_{Gi}^{ht} \leq x_i^{ht} P_{Gi\max}, h = 18, 19, 20, \ t = 1, ..., n, \ i = 1, ..., m \tag{19}$$

Power reserve constraints

$$\sum_{i=1}^{m} P_{Gi\max} x_i^{ht} \geq P_D^{ht} + P_R^{ht}, h = 18, 19, 20, \ t = 1, ..., n \tag{20}$$

The power reserve constraints ensure that system is reliable and stable.
Minimum up/downtime constraints

$$\left(U_{ht-1,i}^{up} - H_i^{up} \right) \left(x_i^{ht-1} - x_i^{ht} \right) \geq 0, h = 18, 19, 20, \ t = 1, ..., n, \ i = 1, ..., m \tag{21}$$

$$\left(U_{ht-1,i}^{down} - H_i^{down} \right) \left(x_i^{ht} - x_i^{ht-1} \right) \geq 0, h = 18, 19, 20, \ t = 1, ..., n, \ i = 1, ..., m \tag{22}$$

In this study $m = 22$.

3 Empirical results and discussion

This section presents the results of both forecasting electricity demand at hours 18:00, 19:00 and 20:00 including the optimal solution to the generating units to commit.

3.1 Exploratory data analysis

The summary statistics for hours considered as the peak period (i.e.18:00, 19:00 and 20:00) in this paper are shown in Table 1. Note that the sampling period considered is from January 2009 to June 2012. For each of the three hours the maximum demand is 36,125MW, 36,970MW and 36,073MW respectively. The skewness and kurtosis presented in Table 1 show that the distributions of the three hours are non-normal. A typical daily load profile over the sampling period is given in Fig. 1 which

Table 1: Summary statistics for electricity demand (MW) at hours 18:00, 19:00 and 20:00.

Hour	Mean	Median	Max	Min	St. Dev.	Skewness	Kurtosis
18:00	29,965	30,276	36,125	20,457	2,508	-0.2317	3.0012
19:00	31,063	31,121	36,970	20,372	2,692	-0.2920	2.8814
20:00	31,195	31,451	36,073	21,974	2,058	-0.5422	3.5533

shows that the peak load occurs between 18:00hrs and 20:00hrs. Forecasting electricity demand during this period is very very important to system operators as they have to balance the high demand with what can be supplied by power plants. The highest peak is around 19:00hrs. The South African daily electricity demand time series and density plots for the peak period considered in this paper are as shown in Fig. 2. The panels on the left of Fig. 2 show the time series plots while those on the right panels show the density plots of the demand for 18:00, 19:00 and 20:00 hours respectively. Note that the seasonality patterns of electricity demand in SA on the left panels of Fig. 2, i.e. the plots indicate higher electricity demand in winter and lower demand in summer yearly. The densities on the right panels of Fig. 2 also show that the distributions of the three hours do not follow normal distributions which is consistent with the skewness and kurtosis reported in Table 1. The distributions of electricity demand are further highlighted by the box and whisker plots in Fig. 3.

A plot of electricity demand at 19:00 hrs superimposed with nonlinear trend is given in Fig. 4. The other graphs for the other hours, 18:00 and 20:00 are done in the same way.

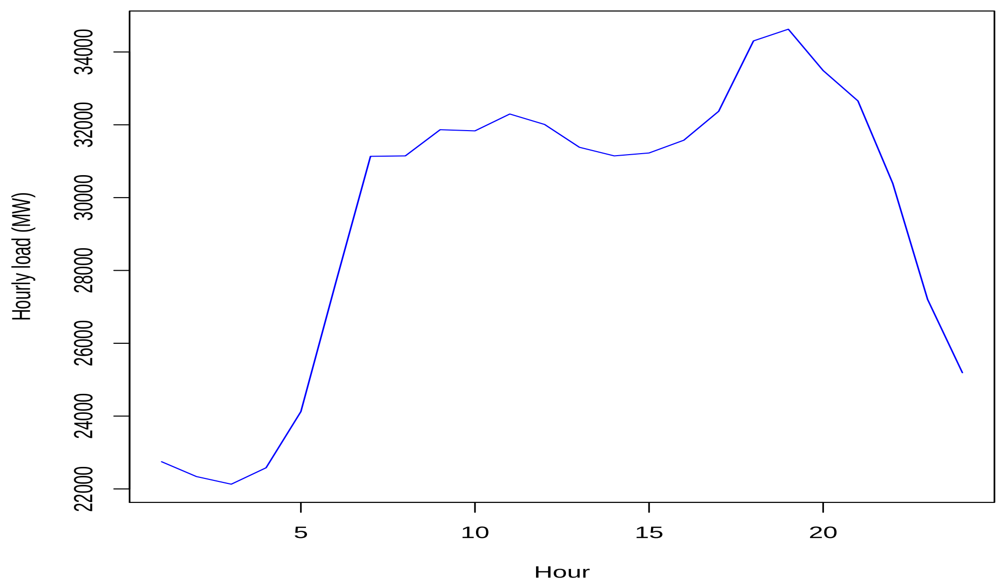


Figure 1: Typical daily load profile.

3.2 Forecasting results

The PLAQR models with interactions (cross effects) whose variables we got using Lasso via hierarchical interactions are:

$$y_{t(18)} = \beta_{0(18)} + \text{DH+bs(Daytype)+bs(DPED)+bs(AED)+bs(maxTI)+bs(DmaxTCI)+bs(noltrend)} \\ \text{+bs(Lag1)+bs(Lag2)+bs(month)+bs(AED,DmaxTCI)+bs(AED,month)} + \varepsilon_{t(18)} \quad (23)$$

$$y_{t(19)} = \beta_{0(19)} + \text{DH+bs(Daytype)+bs(DPED)+bs(AED)+bs(maxTI)+bs(minTI)+bs(noltrend)} \\ \text{+bs(Lag1)+bs(Lag2)+bs(month)+bs(maxTI,Lag1)+bs(noltrend,Lag1)+bs(noltrend,Lag2)} + \varepsilon_{t(19)} \quad (24)$$

$$y_{t(20)} = \beta_{0(20)} + \text{DH+ bs(DPED)+bs(minED)+bs(AED)+bs(maxTC)+bs(minTI)+ bs(noltrend)} \\ \text{+bs(Lag1)+bs(Lag2)+bs(month)+bs(DPED,month)+bs(AED,noltrend)+bs(maxTC,Lag2)} \\ \text{+bs(minTI,Lag2)+bs(noltrend,Lag1)} + \varepsilon_{t(20)} \quad (25)$$

where 'bs' is a B-spline. For each of the three hours the models were split into two, one with and the other without interactions. For both models the better model was selected based on Bayesian information criterion (BIC). For Model 1 (18:00) without interactions and Model 2 (18:00) with interactions the BIC values were found to be 1.51731 and 1.5110 respectively meaning that the model with interactions is a better model. After fitting the PLAQR models to electricity demand data for 18:00, 19:00 and 20:00, tests for autocorrelations in the residuals were then carried out. Residuals of PLAQR models were extracted so as to fit SARIMA models. The best fitting model for 18:00 without and with interactions were SARIMA(2,0,5)(2,0,3)[7] and SARIMA(5,0,3)(2,0,3)[7] respectively. Residuals of these SARIMA models were tested for autocorrelation and the p-values of the Ljung-Box test are found to be 0.8712 and 0.9050 respectively. Fitted values of the SARIMA model were extracted and

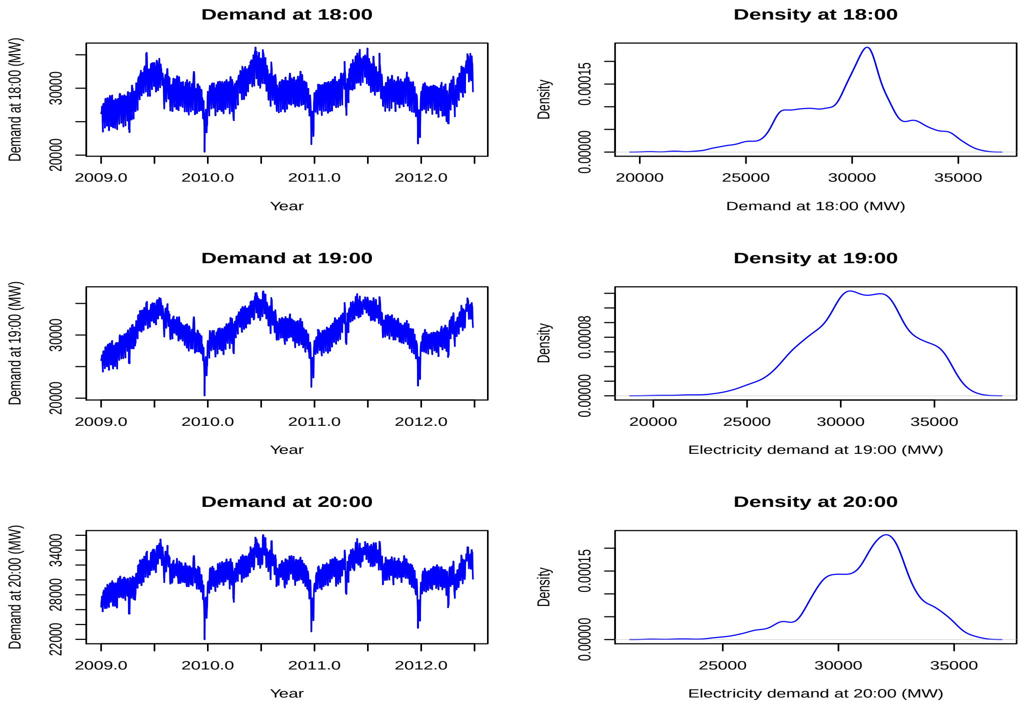


Figure 2: Plot of demand and densities at hours 18:00, 19:00 and 20:00.

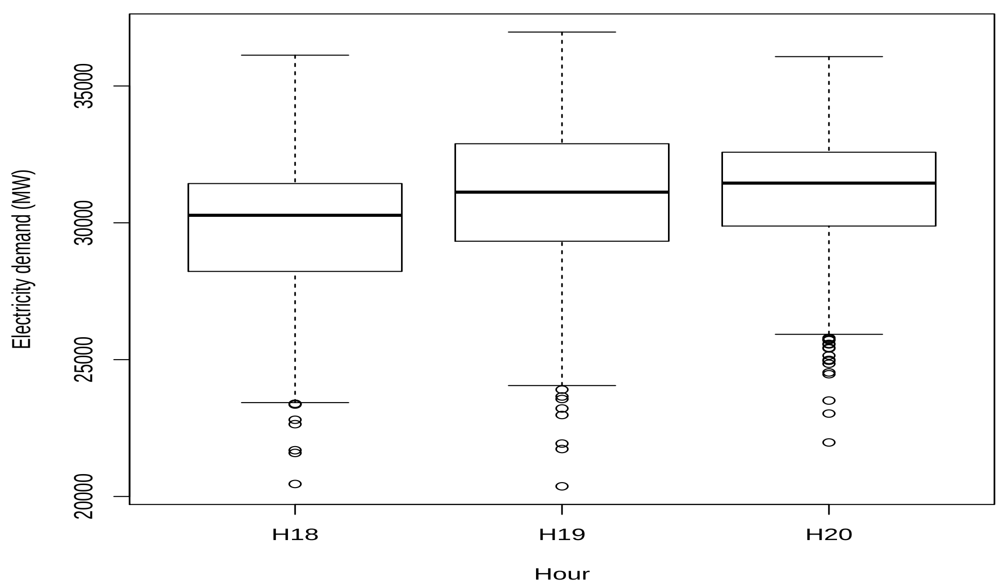


Figure 3: Box plots for hours 18:00, 19:00 and 20:00.

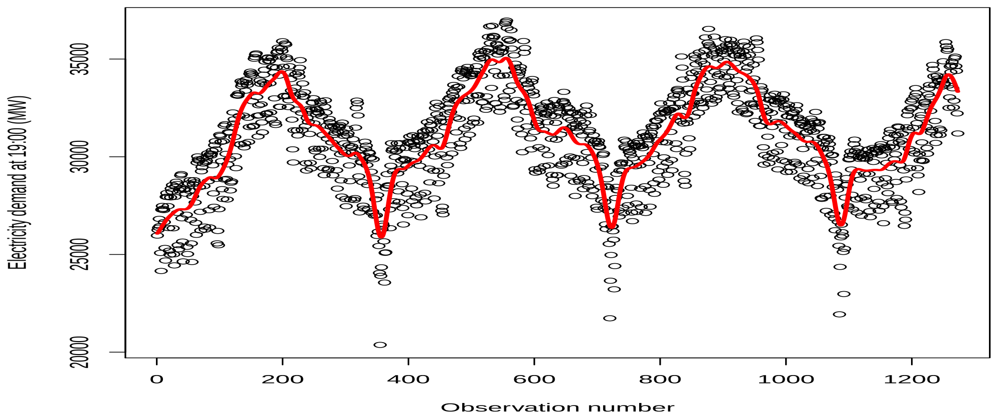


Figure 4: Plot of electricity demand at 19:00 with a nonlinear trend (solid curve).

subtracted from the data at 18:00hrs to get a new set of values for the response variable. The procedure discussed in Section 3 was then followed. The “opera” r package was used for combining the forecasts from Model 1 and Model 2. Based on the pinball loss function the weights assigned to the forecasts from these two methods were 0.68 and 0.32 for Models 1 and 2 respectively. A summary of the error measures for M1 (fH18 PLAQR model without interactions), M2 (fH18I PLAQR model with interactions) and M3 (combined) are presented in Table 2. Based on the error measures, M1 is found to be the best fitting model. From Table 2 the percentage improvements calculated using Eq. (16) of M1 over M2 and M3 are 0.51% and 0.077% respectively.

Table 2: Model comparisons with and without interactions and combined: Hour 18:00.

	M1	**M2**	**M3**
Pinball loss	146.5522	147.3104	146.6648
CRPS	1,343.867	1,344.592	1,344.169
LogS	9.199319	9.20058	9.199709

The BIC values for Models 4 (model without interactions) and 5 (model with interactions) for 19:00hrs were found to be 1.4578 and 1.4205 respectively. Model M5 is better than model M4. The best fitting model for the residuals of the PLAQR model for 19:00 with interactions was found to be SARIMA(1,0,4)(2,0,5)[7] and the corresponding p-value was 0.9889. Based on the pinball loss function the weights assigned to the forecasts from these two methods were 0 and 1 for Models 4 and 5 respectively. Based on the error measures given in Table 3, M5 was found to be the best fitting model. From Table 3 the percentage improvement calculated using Eq. (16) of M5 over M4 is 4.74%. The BIC values for Models M6 (model without interactions) and M7 (model with interactions) for 20:00hrs were found to be 1.303463 and 1.197581 respectively. Model M7 is better than model M6. The best fitting model for 20:00hrs without and with interactions are SARIMA(2,0,3)(2,0,5)[7] and SARIMA(3,1,4)(2,0,5)[7] respectively. Residuals of these SARIMA models are then tested for autocorrelation and the p-values of the Ljung-Box test are found to be 0.7844 and 0.8346 respectively.

Table 3: Model comparisons with and without interactions: Hour 19:00.

	M4	**M5**
Pinball loss	122.1813	116.3901
CRPS	1,330.275	1,330.48
LogS	9.178344	9.178213

Based on the pinball loss function the weights assigned to the forecasts from these two methods were 0.168 and 0.832 for Models 6 and 7 respectively. A summary of the error measures for M6, M7 and M8 (combined) are given in Table 4. Based on the error measures, M8 is found to be the best model. From Table 4 the percentage improvements calculated using Eq. (16) of M8 over models M6 and M7 are respectively 10.49% and 0.86%.

Table 4: Model comparisons with and without interactions and combined: Hour 20:00.

	M6	**M7**	**M8**
Pinball loss	140.3187	125.6016	126.6951
CRPS	945.2762	944.7923	944.7907
LogS	8.85249	8.852136	8.851865

A visual inspection of Fig. 5 shows that there is very little autocorrelation left in the residuals for the model fitted to the hour 19:00hrs. The model is then used for forecasting. It is shown in Fig. 6 that the forecasts from model M5 follow the actual demand data remarkably well. As shown in Fig. 7, the density plot of the forecasts from model M5 is a better fit to the density of the actual demand compared to the density plot from model M4. The plots of actual demand with the forecasts including density plots for hours 18:00 and 20:00 are given in the appendix.

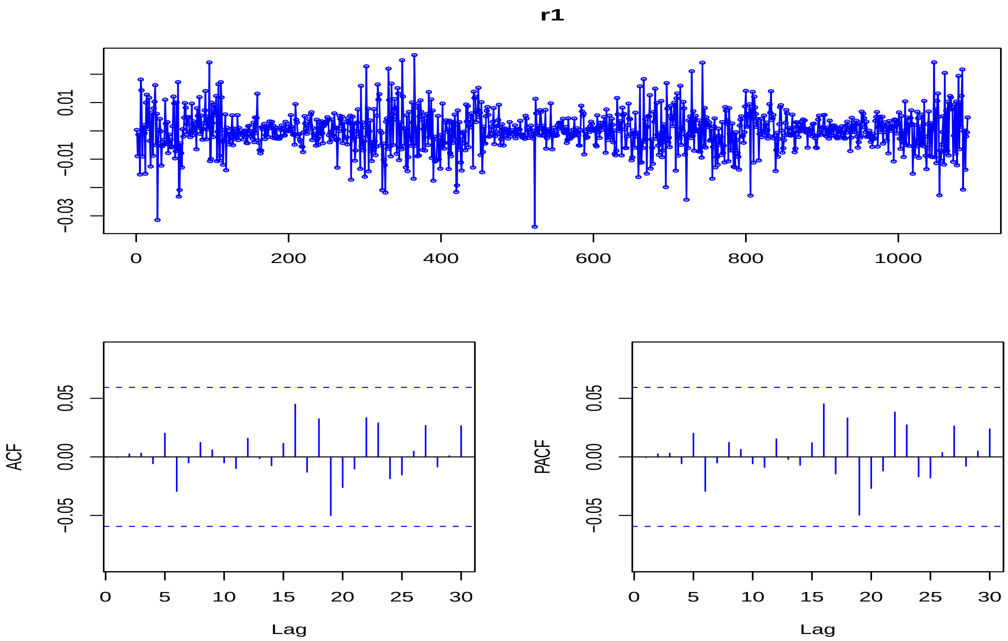


Figure 5: Time series display of the residuals H19I.

3.3 Unit commitment results

3.3.1 Solution using the bounded variable mixed integer linear programming model

The base load demand stations, i.e. coal fired and nuclear stations are given in Table 5. In this study fuel cost is used to represent the average production cost given in column 2 of Table 5. The data for the fuel costs is from [37]. The minimum and maximum production levels in megawatts are given in columns 3 and 4 respectively. Similarly Table 6 presents data for peaking stations.

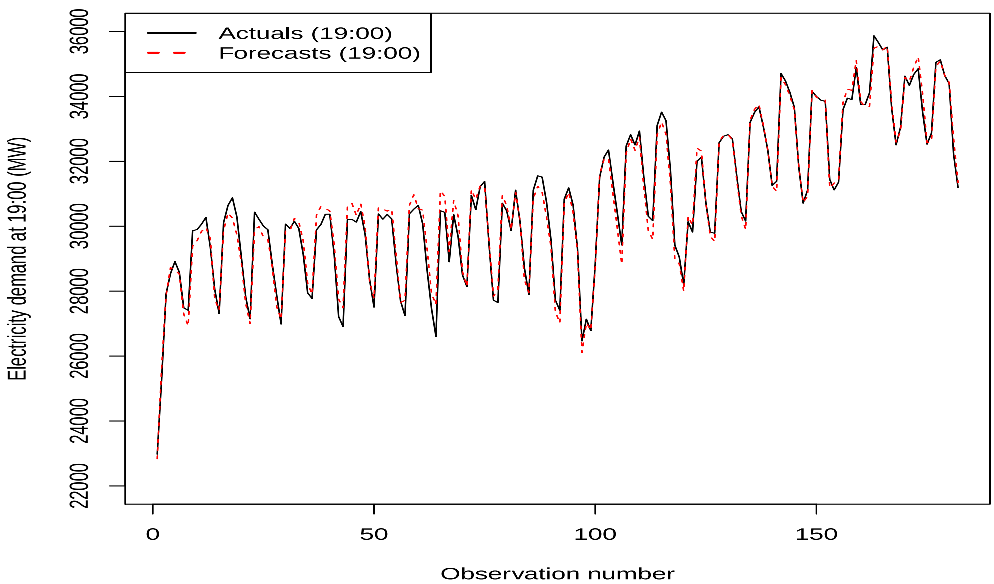


Figure 6: Plot of actual demand and forecasts.

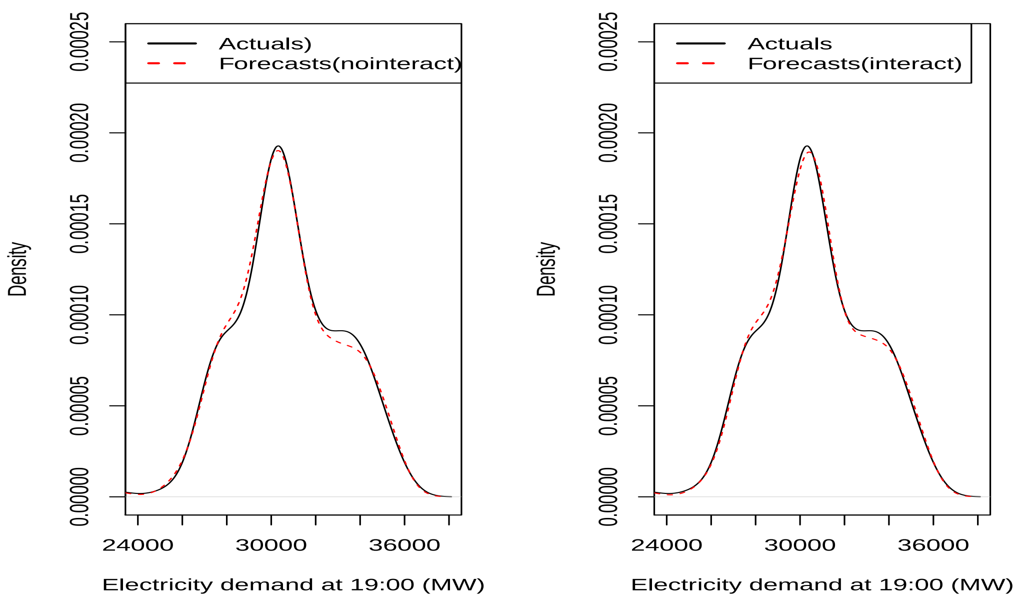


Figure 7: Density plots.

Table 5: Base load demand stations.

Unit	**Min Ave prod cost c_i(rands)**	**Min (MW)**	**Max (MW)**
g_{1c}	244.4	0	4,116
g_{2c}	247.2	0	4,110
g_{3c}	245.7	0	3,990
g_{4c}	270.6	0	3,708
g_{5c}	270.6	0	3,654
g_{6c}	245.7	0	3,600
g_{7c}	245.7	0	3,600
g_{8c}	247.2	0	3,000
g_{9c}	244.4	0	2,400
g_{10c}	247.2	0	2,000
g_{11c}	245.7	0	1,600
g_{12c}	247.2	0	1,200
g_{13c}	244.4	0	1,000
g_{14n}	231.8	0	1,931

Table 6: Peaking stations.

Unit	**Min Ave prod cost c_i(rands)**	**Min (MW)**	**Max (MW)**
g_{15g}	596.4	0	1,338
g_{16g}	490.6	0	740
g_{17g}	490.6	0	171
g_{18g}	460.8	0	171
g_{19p}	231.2	0	1,000
g_{20p}	231.2	0	400
g_{21h}	231.2	0	344
g_{22h}	231.2	0	240

The forecasts for the first seven days of January 2012 for 18:00hrs, 19:00hrs and 20:00hrs using the models M1, M5 and M8 respectively are presented in Table 7. The forecasted demand for Wednesday 4 January 2012 shown in bold is used to demonstrate the application of the unit commitment problem. It should be noted that the procedure can be applied to any other hours of the day. Lingo version 14 was used to solve the unit commitment problem.

Table 7: Out of sample forecasts ($Q_{0.5}$ quantile forecasts).

Date	18:00	19:00	20:00
Sun 01-Jan-12	22,581	22,835	24,255
Mon 02-Jan-12	25,306	25,657	26,872
Tue 03-Jan-12	28,192	27,860	28,802
Wed 04-Jan-12	**29,069**	**28,724**	**29775**
Thu 05-Jan-12	29,015	28,658	29,758
Fri 06-Jan-12	28,721	28,522	30,176
Sat 07-Jan-12	27,060	27,288	28,976

The bounded variable mixed integer linear programming model is.

$$\text{Min}\ \ Z \ = \sum_{i=1}^{13} c_i g_{ic} + c_{14} g_{14n} + \sum_{i=15}^{18} c_i g_{ig} + \sum_{i=19}^{20} c_i g_{ip} + \sum_{i=21}^{22} c_i g_{ih} \tag{26}$$

where Z denotes the total cost to be minimized.

Subject to

Load balance constraints

$$\sum_{i=1}^{13} g_{ic} + g_{14n} + \sum_{i=15}^{18} g_{ig} + \sum_{i=19}^{20} g_{ip} + \sum_{i=21}^{22} g_{ih} = P_D^{ht}, \quad h = 18, 19, 20; \quad t = 04/01/2012 \tag{27}$$

$$\text{with} \ \ P_D^{18t} = 29,069\text{MW}, \quad P_D^{19t} = 28,724\text{MW}, \quad P_D^{20t} = 29,775\text{MW}. \tag{28}$$

The generation reserve margin expressed as a percentage is the method usually used to evaluate the generation system adequacy which is normally in the range 15-25% ([4]). Using a 15% reserve margin we get

$$P_R^{18t} = 0.15 \times 29,069 = 4,360.35\text{MW}, \quad P_R^{19t} = 0.15 \times 28,724 = 4,308.6\text{MW},$$

$$P_R^{20t} = 0.15 \times 29,775 = 4,466.25\text{MW} \tag{29}$$

Power reserve constraints

$$1.15 \left(\sum_{i=1}^{13} g_{ic} + g_{14n} + \sum_{i=15}^{18} g_{ig} + \sum_{i=19}^{20} g_{ip} + \sum_{i=21}^{22} g_{ih} \right) \leq 44,003 \tag{30}$$

Generator power output limits

$$0 \leq g_{1c} \leq 4,116 \tag{31}$$

$$. \tag{32}$$

$$. \tag{33}$$

$$. \tag{34}$$

$$0 \leq g_{22h} \leq 240 \tag{35}$$

The optimal solution for each of the three hours on Wednesday 4 January 2012 is presented in Table 8. Column 2 of Table 8 shows the amount of electricity in megawatts each one of the generating units should produce to meet the predicted demand for that hour at the same time ensuring that the power reserve constraint is met. The minimum cost in rands of producing electricity at each of the three hours is given in parentheses.

Table 8: Optimal Solution using bounded variable MILP

Variable	18:00 (R7,082,499.00)		19:00 (R6,997,215.00)		20:00 (R7,257,022.00)	
	Soln[a]	RC[b]	Soln	RC	Soln	RC
g_{1c}	4116	0	4116	0	4116	0
g_{2c}	4110	0	4110	0	4110	0
g_{3c}	3990	0	3990	0	3990	0
g_{4c}	0	23.4	0	*23.4*	0	23.4
g_{5c}	0	23.4	0	23.4	0	23.4
g_{6c}	3600	0	3600	0	3600	0
g_{7c}	3600	0	3600	0	3600	0
$\mathbf{g_{8c}}$	738	0	393	0	1444	0
g_{9c}	2400	0	2400	0	2400	0
g_{10c}	0	0	0	0	0	0
g_{11c}	1600	0	1600	0	1600	0
g_{12c}	0	0	0	0	0	0
g_{13c}	1000	0	1000	0	1000	0
g_{14n}	1931	0	1931	0	1931	0
g_{15g}	0	349.2	0	349.2	0	349.2
g_{16g}	0	243.4	0	243.4	0	243.4
g_{17g}	0	243.4	0	243.4	0	243.4
g_{18g}	0	213.6	0	213.6	0	213.6
g_{19p}	1000	0	1000	0	1000	0
g_{20p}	400	0	400	0	400	0
g_{21h}	344	0	344	0	344	0
g_{22h}	240	0	240	0	240	0

[a]Soln-Solution
[b]RC-Reduced cost

The $Q_{0.99}$ (99^{th} percentile) quantile forecasts are given in Table 9. This means that it is highly unlikely that electricity demand will exceed 29,562MW, 29,232MW and 29,807MW on hours 18:00, 19:00 and 20:00 respectively on Wednesday 4 January 2012. The expected maximum costs associated with the $Q_{0.99}$ quantile forecasts are given in Table 10. This analysis may be useful to power utility companies such as Eskom with the unit commitment scheduling and economic dispatching of electricity. It is noted that the optimal solutions from Tables 8 and 10 are the same except for generating unit g_{8c}. This shows that for any increase in demand from the median forecasts ($Q_{0.5}$ quantile forecasts) it will be economical to increase the generation of electricity from generating unit g_{8c}.

Table 9: Out of sample forecasts ($Q_{0.99}$ quantile forecasts)

Date	18:00	19:00	20:00
Sun 01-Jan-12	23,127	23,879	24,718
Mon 02-Jan-12	25,722	26,417	27,085
Tue 03-Jan-12	29,047	28,331	28,862
Wed 04-Jan-12	**29,562**	**29,232**	**29,807**
Thu 05-Jan-12	29,399	29,300	29,912
Fri 06-Jan-12	28,908	29,362	30,300
Sat 07-Jan-12	27,488	28,187	29,019

Table 10: Optimal Solution using bounded variable MILP (for $Q_{0.99}$ quantile forecasts)

Variable	18:00 (R7,204,369.00)		19:00 (R7,122,793.00)		20:00 (R7,264,933.00)	
	Soln[a]	RC[b]	Soln	RC	Soln	RC
g_{1c}	4116	0	4116	0	4116	0
g_{2c}	4110	0	4110	0	4110	0
g_{3c}	3990	0	3990	0	3990	0
g_{4c}	0	23.4	0	23.4	0	23.4
g_{5c}	0	23.4	0	23.4	0	23.4
g_{6c}	3600	0	3600	0	3600	0
g_{7c}	3600	0	3600	0	3600	0
$\mathbf{g_{8c}}$	1231	0	901	0	1476	0
g_{9c}	2400	0	2400	0	2400	0
g_{10c}	0	0	0	0	0	0
g_{11c}	1600	0	1600	0	1600	0
g_{12c}	0	0	0	0	0	0
g_{13c}	1000	0	1.000	0	1000	0
g_{14n}	1931	0	1931	0	1931	0
g_{15g}	0	349.2	0	349.2	0	349.2
g_{16g}	0	243.4	0	243.4	0	243.4
g_{17g}	0	243.4	0	243.4	0	243.4
g_{18g}	0	213.6	0	213.6	0	213.6
g_{19p}	1000	0	1000	0	1000	0
g_{20p}	400	0	400	0	400	0
g_{21h}	344	0	344	0	344	0
g_{22h}	240	0	240	0	240	0

[a]Soln-Solution
[b]RC-Reduced cost

4 Conclusion

The paper presented an application of partially linear additive quantile regression models to short term electricity demand forecasting using South African data. The study focused on the peak hours of the day. Variable selection was done using Lasso via hierarchical pairwise interactions. Models for each hour were split into one with and one without pairwise interactions. Models of each hour were combined using an algorithm in which the average loss suffered was based on the pinball loss function. This resulted in three sets of forecasts for each hour. The best set of forecasts was selected based on probabilistic forecast error measures, pinball loss values, continuous ranked probability scores and the log scores. For instance, as shown in 2 the model without interactions has smaller pinball loss value as compared to the other two models considered for hour 18:00 and was found to be the best fitting model. while from hours 19:00 and 20:00 model with interactions and the combined model were the best fitting models respectively.

The forecasts from the three hours were then used as inputs in solving the unit commitment problem. A bounded variable mixed integer programming modelling approach was used and the developed optimization models were solved using Lingo version 14. Median forecasts for Wednesday 4 January 2012 were 29,069MW, 28,724MW and 29,775MW for the hours 18:00, 19:00 and 20:00respectively, with the corresponding optimal minimal costs of R7,082,499.00, R6,997,215.00 and R7,257,022.00. Results from the unit commitment problem show that it is very costly to use gas fired generating units. These were not selected as part of the optimal solution. Using the $Q_{0.99}$ quantile forecasts, it is noted that the optimal solutions from median forecasts ($Q_{0.5}$ quantile forecasts) are the same as those from the $Q_{0.99}$ quantile forecasts except for generating unit g_{8c}, which is a coal fired unit. This shows that for any increase in demand from the median forecasts it will be economical to increase the generation of electricity from generating unit g_{8c}. The $Q_{0.99}$ quantile forecasts for Wednesday 4 January 2012 for the hours 18:00, 19:00 and 20:00 were found to be 29,562MW, 29,232MW and 29,807MW respectively with optimal solutions of R7,204,369.00, R7,122,793.00 and R7,264,933.00

The main contribution of this study is in the use of nonlinear trend variables and the combining of forecasting with the unit commitment problem. The study could be useful to system operators in power utility companies in scheduling and dispatching of electricity at a minimal cost particularly during the peak period when the grid is constrained due to increased demand for electricity. Future work could look at peak load forecasting where the extreme peaks could be used in the assessment of power system reliability [4] including the use of stochastic programming techniques which are known to capture a lot of uncertainties in the unit commitment problem. The modelling framework discussed in this paper can also be used in solar power forecasting with similar input variables.

Acknowledgments

The authors acknowledge the National Research Foundation of South Africa (Grant No: 93613) for funding this research and to both Eskom, South Africa's power utility company and the South African Weather Services for providing the data.

Appendix

Appendix A1: Evaluating the prediction errors

Tables 11 - 13 show the accuracy measures, root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) for each of the models used for electricity demand at hours 18:00, 19:00 and 20:00 respectively. These error measures are generally used for predicting point forecasts.

Table 11: Model comparisons with and without interactions: Hour 18:00 ($Q_{0.5}$)

	M1	**M2**	**M3**
RMSE	353.05	354.54	352.63
MAE (MW)	293.10	294.62	293.30
MAPE (%)	0.9706	0.9710	0.9697

Table 12: Model comparisons with and without interactions: Hour 19:00 ($Q_{0.5}$)

	M4	**M5**
RMSE	306.69	290.49
MAE (MW)	244.36	232.78
MAPE (%)	0.8093	0.7736

Table 13: Model comparisons with and without interactions: Hour 20:00 ($Q_{0.5}$)

	M6	**M7**	**M8**
RMSE	350.78	327.85	326.89
MAE (MW)	280.64	251.20	253.40
MAPE (%)	0.9189	0.8242	0.8308

Appendix A2: Plots of forecasted demand for hours 18:00 and 20:00

Figs. 8 - 11 show the plots of actual demand with the forecasts including density plots for hours 18:00 and 20:00. In all cases both forecasts of electricity demand and forecasted density plots follow the actual demand and density plots remarkably well.

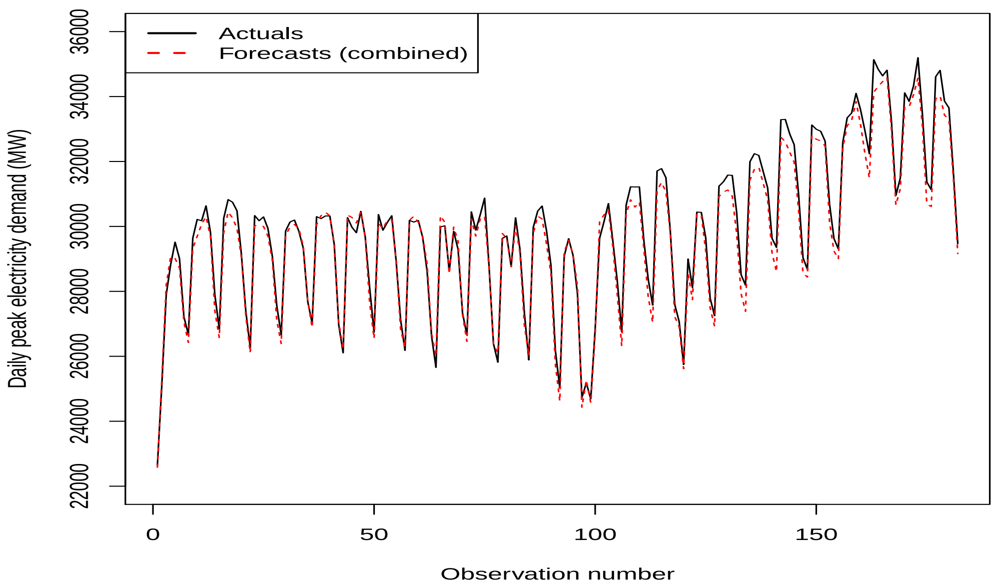


Figure 8: Plot of actual demand and combined forecasts (18:00)

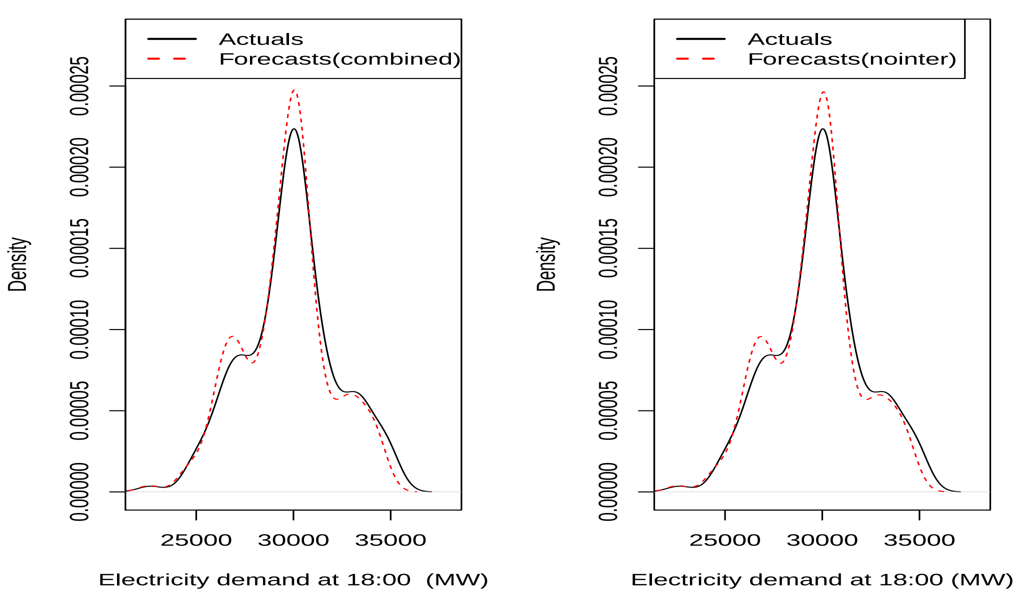


Figure 9: Density plots (18:00) combined forecasts

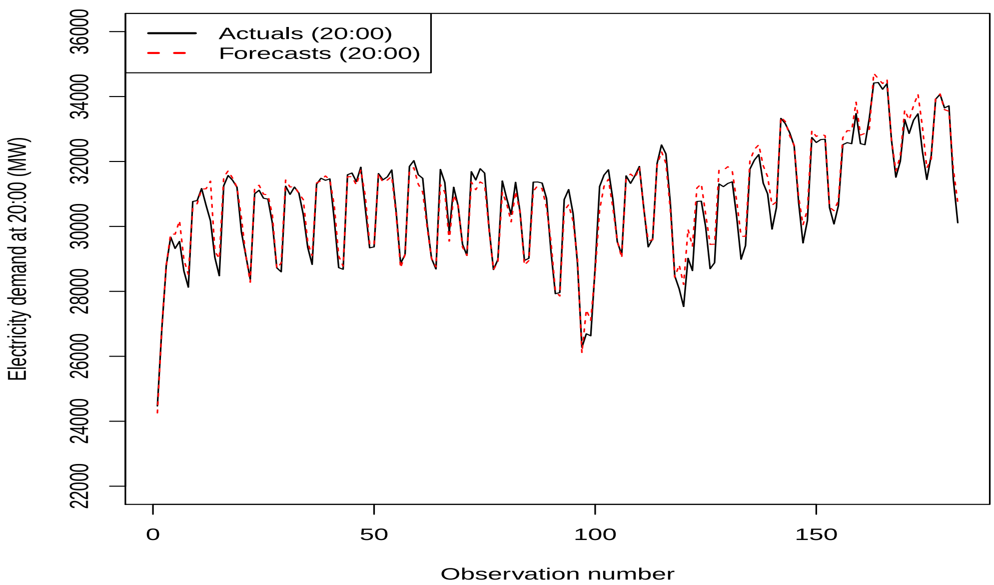


Figure 10: Plot of actual demand and forecasts (20:00)

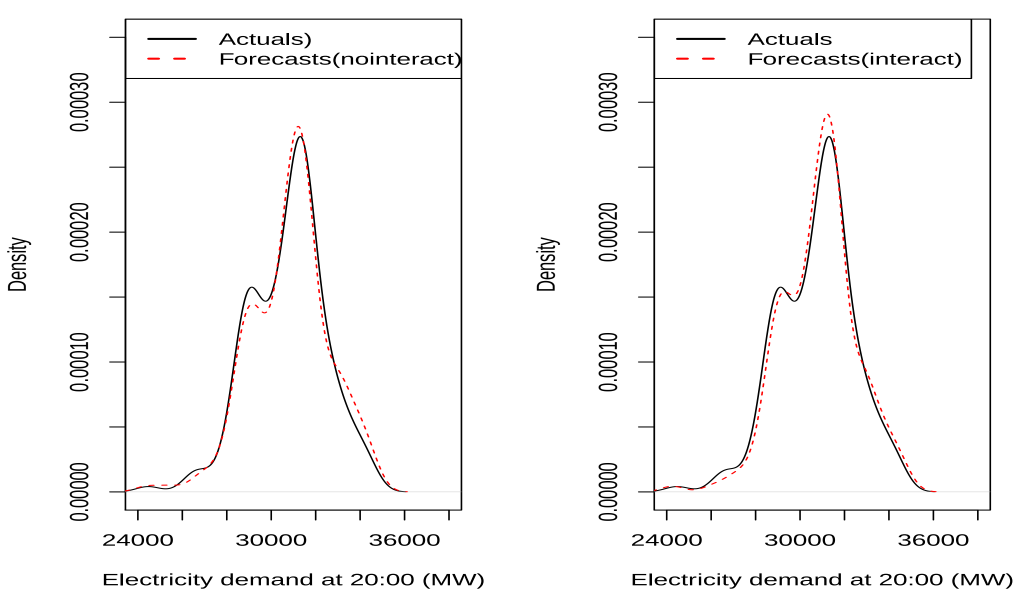


Figure 11: Density plots (20:00)

References

[1] Almeshaiei E., Soltan H. A methodology for electric power load forecasting. Alexandria Engineering Journal 2011; 50:137–144.

[2] Amina M, Kodogiannis VS, Petrounias I, Tomtsis D. A hybrid intelligent approach for the prediction of electricity consumption. Electrical Power and Energy Systems 2012; 43:99–108.

[3] Bien J, Taylor J, Tibshirani R. A lasso for hierarchical interactions. The Annals of Statistics 2013; 41(3): 1111–1141.

[4] Čepin M. Assessment of power system reliability: Methods and applications. London: Springer, 2011.

[5] Chikobvu D, Sigauke C. Modelling influence of temperature on daily peak electricity demand in South Africa. Journal of Energy in Southern Africa 2013; 24(4):63–70.

[6] Dai H, Zhang N, Su W. A literature review of stochastic programming and unit commitment. Journal of Power and Energy Engineering 2015; 3:206–214.

[7] Devaine M, Gaillard P, Goude Y, Stoltz G. Forecasting the electricity consumption by aggregating specialized experts; a review of sequential aggregation of specialized experts, with an application to Slovakian and French country-wide one-day-ahead (half-) hourly predictions, Machine Learning 2012; 90: 231-260.

[8] Engle R F, Granger C W J, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. Journal of the American Statistical Association 1986; 81: 310–320.

[9] Fahad MU, Arbab N. Factor affecting short term load forecasting. Journal of Clean Energy Technologies 2014; 2(4):305–309.

[10] Fan S, Hyndman RJ. Short term load forecasting based on a semiparametric additive model. IEEE Transactions on Power Systems 2012; 27(1):134–141.

[11] Fasiolo M, Goude Y, Nedellec R, Wood S N. Fast calibrated additive quantile regression 2017. Available at `https://arxiv.org/abs/1707.03307`

[12] Feng Y, Ryan SM. Day-ahead hourly electricity load modeling by functional regression. Applied Energy 2016; 170: 455–465.

[13] Gaillard P, Goude Y, Nedellec R. Semi-parametric models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. International Journal of Forecasting 2016; 32(3): 1038–1050.

[14] Gibbons CE, Faruqui A. Quantile Regression for Peak Demand Forecasting. Available at SSRN 2014; `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2485657`. (accessed on 8 December 2016).

[15] Gneiting T, Katzfuss M. Probabilistic forecasting, Annual Review of Statistics and Its Application 2014; 1, 125–151.

[16] Hahn H, Meyer-Nieberg S, Pickl S. Electric load forecasting methods:Tools for decision making. European Journal of Operational Research 2009; 99:902–907.

[17] Hastie TJ, Tibshirani RJ. Generalized Additive Models. London:Chapman and Hall, 1990.

[18] He Y, Liu R, Li H, Wang S, Lu X. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and copula theory. Applied energy 2017; 185: 254–266.

[19] Hinman J, Hickey E. Modelling and forecasting short-term electricity load using regression analysis. Institute of Regulatory Policy Studies 2009; 1–51.

[20] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting 2016; 32(3): 914–938.

[21] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting 2016; 32(3): 896–913.

[22] Hong T, Wang P, White L. Weather station selection for electric load forecasting. International Journal of Forecasting 2015; 31:286 295.

[23] Hoshino T. Quantile regression estimation of partially linear additive models. Journal of Nonparametric Statistics 2014; 26(3): 509–536.

[24] Hu Y, Zho K, Lian H. Bayesian quantile regression for partially linear additive models. Statistics and Computing 2015; 25(3):651–668.

[25] Hyndman RJ. and Athanasopoulos G. Forecasting: principles and practice. OTexts, 2013. [Online]. Available: `https://www.otexts.org/fpp`

[26] Janicki M. Methods of weather variables introduction into short-term electric load forecasting models - a review. 2017 [Online]. Available: `http://DOI:10.15199/48.2017.04.18`

[27] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. Springer 2015.

[28] Jiang YL. Double-Penalized Quantile Regression in Partially Linear Models. Open Journal of Statistics 2015; 5: 158-164.

[29] Koenker R, Bassett GW. Regression quantiles. Econometrica 1978; 46:33–50.

[30] Koenker R. Quantile Regression. Cambridge University Press, 2005.

[31] Kurban M., Filik U.B. Unit commitment scheduling by using the autoregressive and artificial neural network models based short-term load forecasting. In: Probabilistic methods applied to power systems 2008; IEEE: 1–5.

[32] Lim M, Hastie T. Learning interactions via hierarchical group-lasso regularization. J Comput Graph Stat. 2015, 24(3): 627–654.

[33] Odhiambo NM. Energy consumption, prices and economic growth in three SSA countries: A comparative study. Energy Policy 2010; 38:2463–2469.

[34] Pierrot A, Goude Y. Short-term electricity load forecasting with generalized additive models. Proceedings of ISAP power 2011; 593–600.

[35] Olusegun AM, Dikko HG, and Gulumbe SU. Identifying the limitation of stepwise selection for variable selection in regression analysis. American Journal of Theoretical and Applied Statistics 2015, 4(5):414 419.

[36] Portnoy S, Koenker R. The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators. Statistical Science 1997; 12(4): 279–296.

[37] Power Generation Technology Data for Integrated Resource Plan of South Africa. EPRI, Palo Alto, CA: 2015. (Accessed on 21 October 2017). http://www.energy.gov.za/IRP/2016/IRP-AnnexureA-EPRI-Report-Power-Generation-Technology-Data-for-IRP-of-SA.pdf

[38] Saravanan B, Das S, Sikri S, Kothari DP. A solution to the unit commitment problem-a review. Front Energy 2013; 7(2):223–236.

[39] Sigauke C. Modelling electricity demand in South Africa, PhD Thesis, University of the Free-State 2014.

[40] Short M, Crosbie T, Dawood M, Dawood N. Load forecasting and dispatch optimisation for decentralised co-generation plant with dual energy storage. Applied Energy 2017; 186: 304–320.

[41] Taieb SB, Huser R, Hyndman RJ., Genton MG. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. IEEE Transactions on Smart Grid 2016; 7(5): 2448–2455.

[42] Taylor JW, McSharry PE. Short term load forecasting methods: An evaluation based on european data. IEEE Transactions on Power Systems 2008; 22:2213–2219.

[43] Tibshirani R. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B 1996; 58(1): 267–288.

[44] Wood SN. Generalized Additive Models: An introduction with R. London: Chapman and Hall, 2006.

[45] Wright B. A review of unit commitment 2013; http://blog.narotama.ac.id/wp-content/uploads/2014/12/A-Review-of-Unit-Commitment.pdf (accessed 26 June 2017).

[46] Wu L, Shahidehpour M, Li T. Stochastic Security-Constrained Unit Commitment. IEEE Transactions on Power Systems 2007, 22(2):800–811.

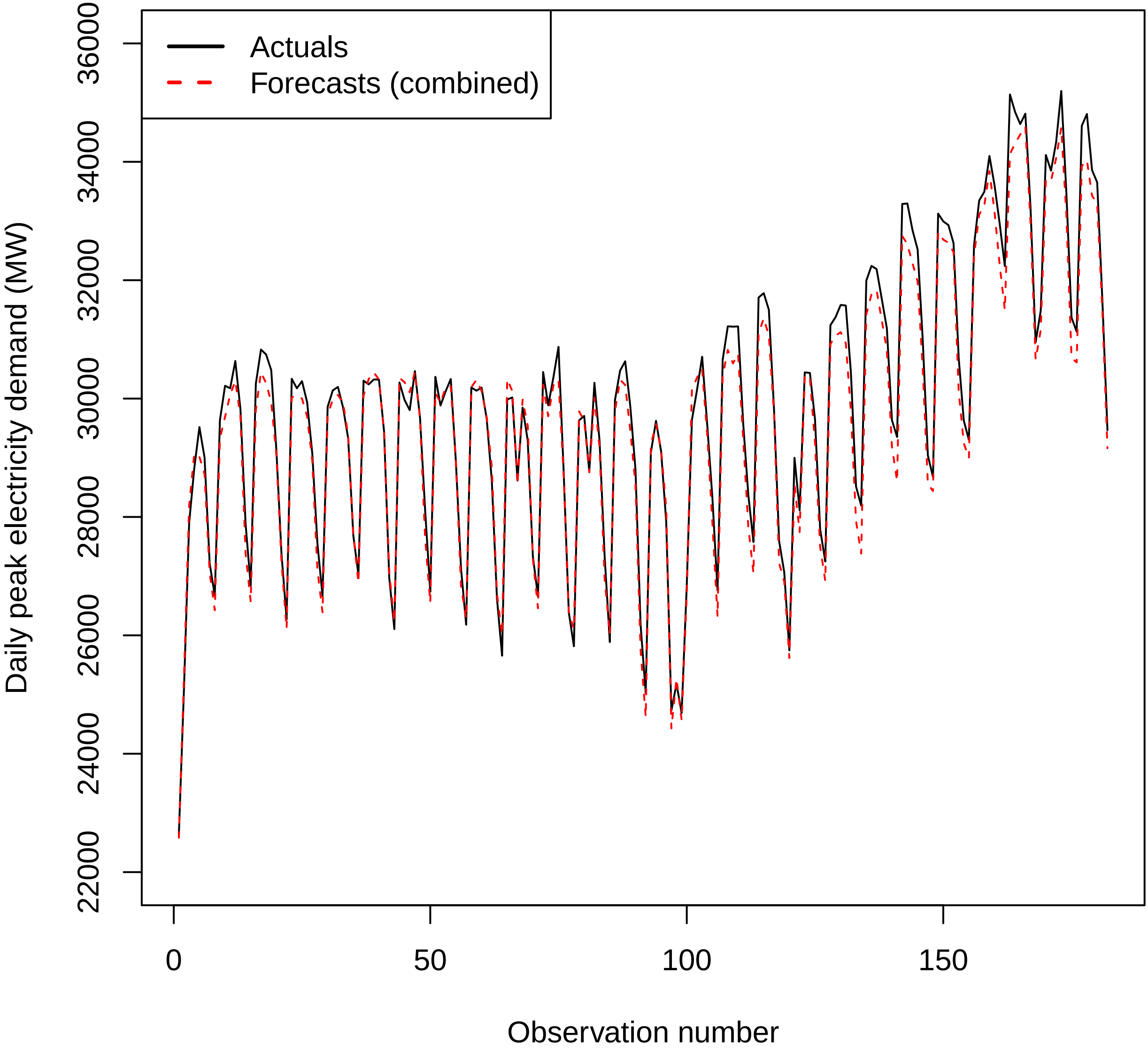
[47] Xia C., Wan J., McMenemy K. Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. Electrical Power and Energy Systems 2010; 32:743–750.

[48] Yang Y, Li S, Li W, Qu M. Power load probability density forecasting using Gaussian process quantile regression. Applied Energy 2018; 213: 499–509.

LaTeX Source Files
Click here to download LaTeX Source Files: revised manuscriptR2.tex

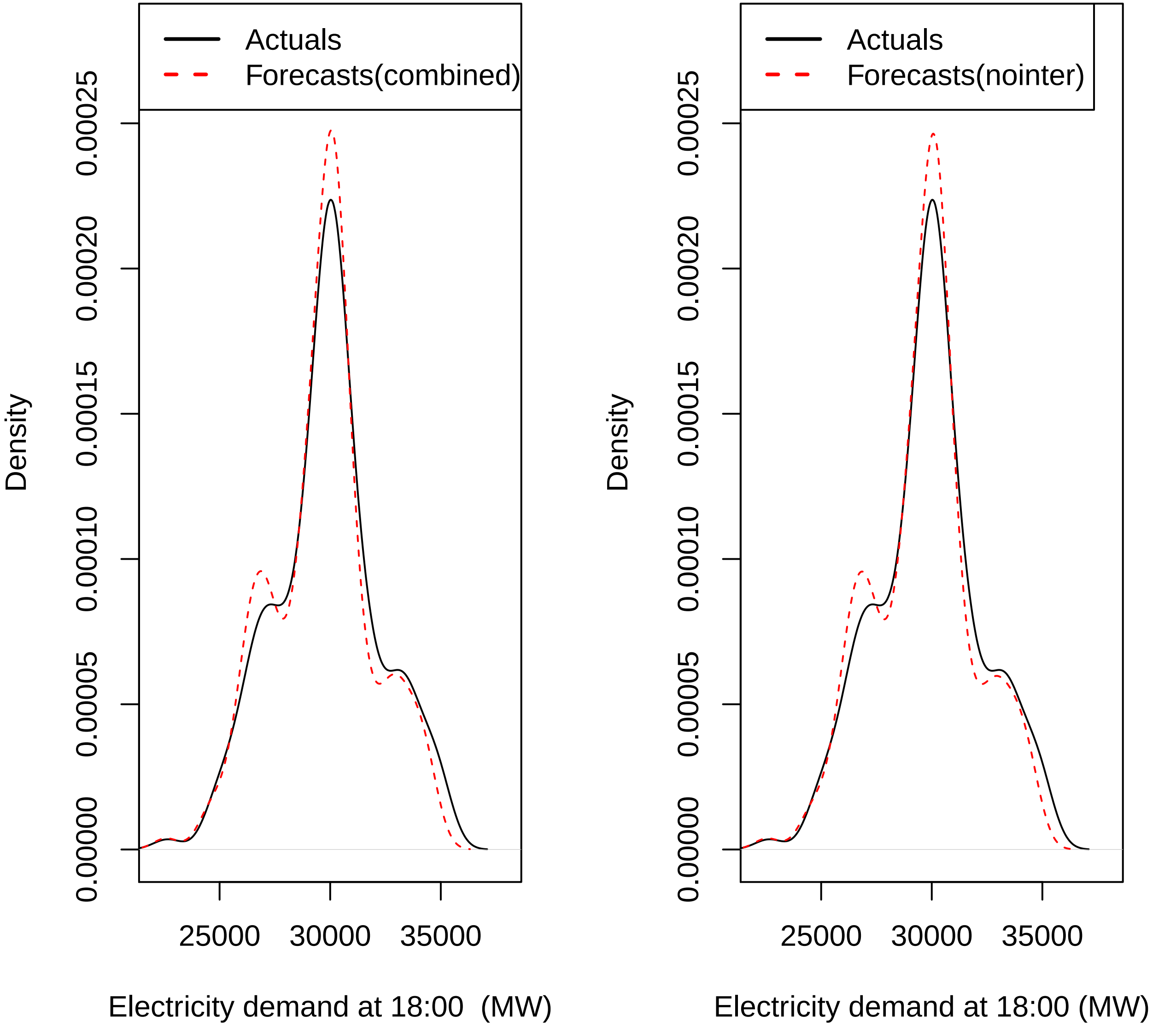
Data in Brief
Click here to download Data in Brief: Datainbrief Sigauke et al 2017.zip

Figure

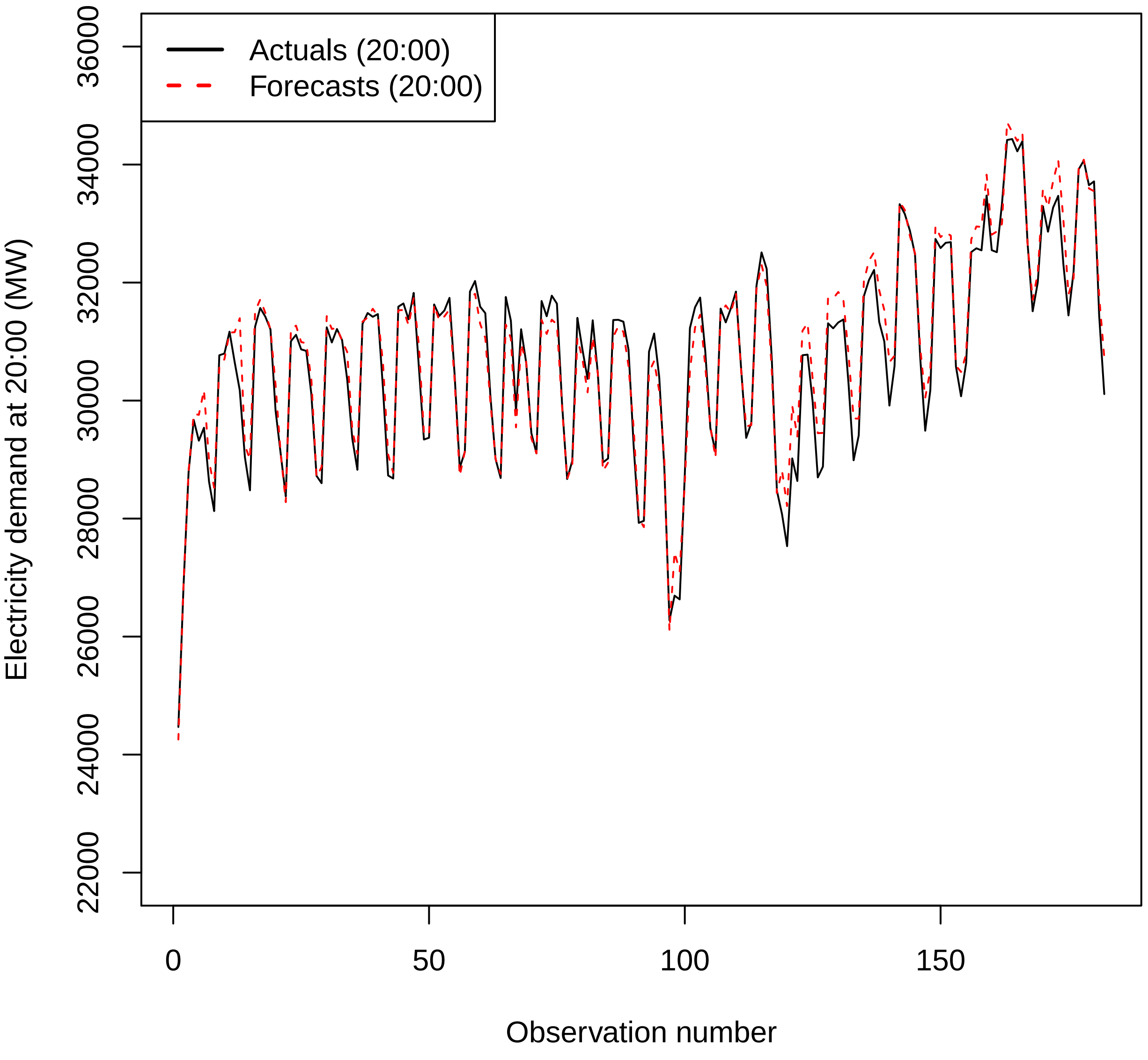


Actuals
Forecasts (combined)
Daily peak electricity demand (MW)
22000
24000
26000
28000
30000
32000
34000
36000
0
50
100
150
Observation number

Figure

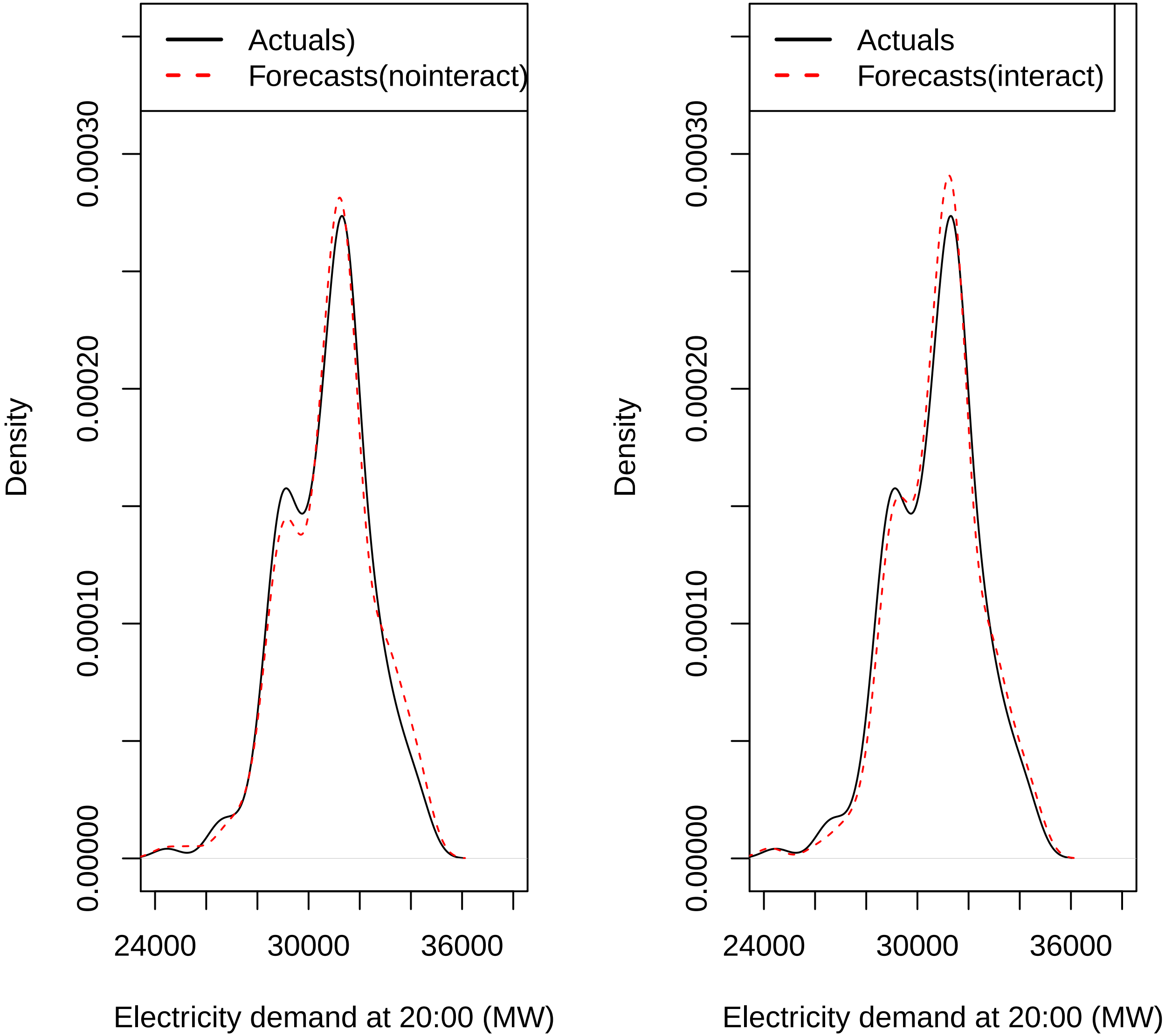

Actuals
Forecasts(combined)
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
Density
25000
30000
35000
Electricity demand at 18:00 (MW)
Actuals
Forecasts(nointer)
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
Density
25000
30000
35000
Electricity demand at 18:00 (MW)

Figure

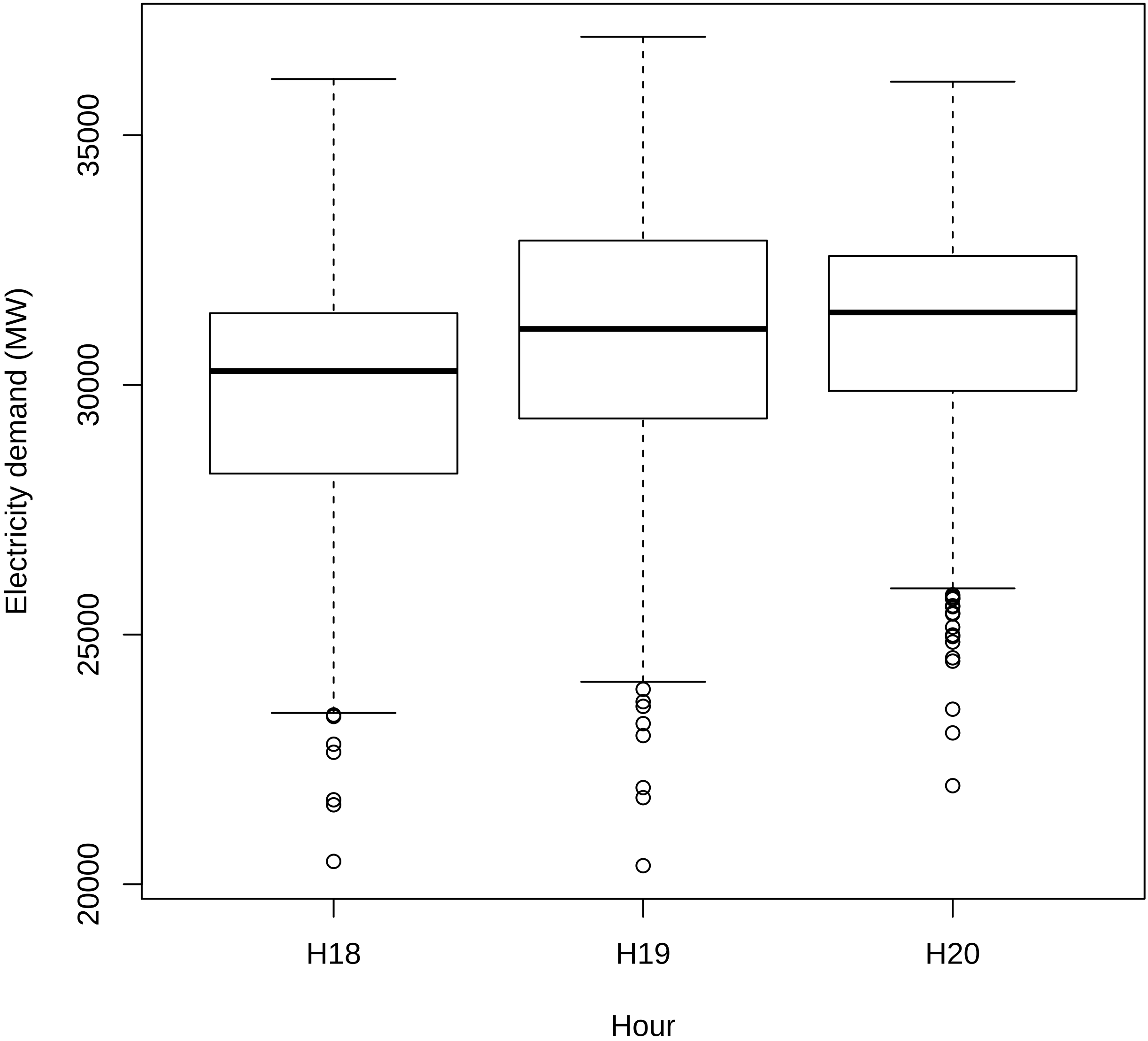


Actuals (20:00)
Forecasts (20:00)
Electricity demand at 20:00 (MW)
22000
24000
26000
28000
30000
32000
34000
36000
0
50
100
150
Observation number

Figure

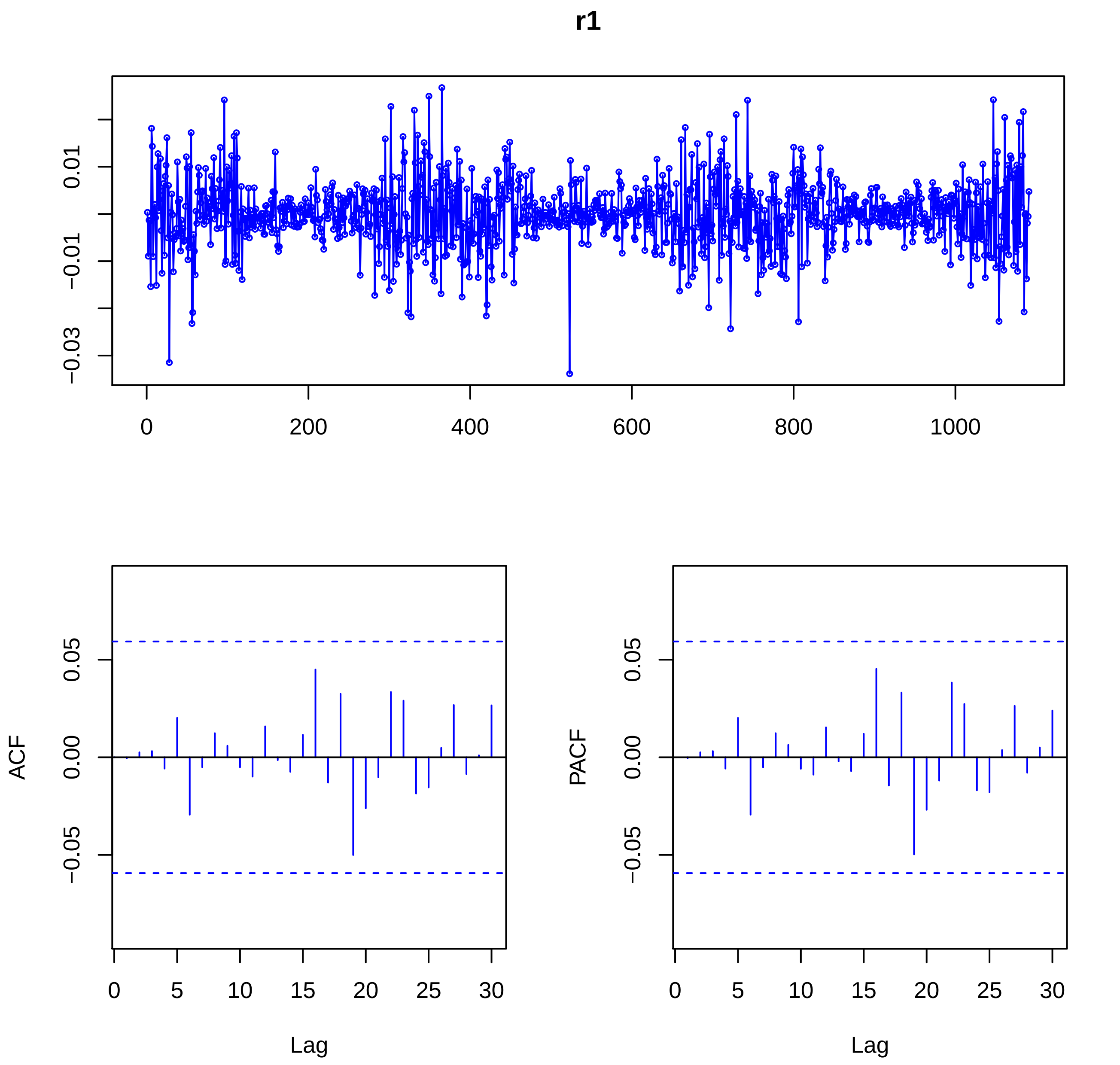

Actuals)
Forecasts(nointeract)
Density
0.00000
0.00010
0.00020
0.00030
24000
30000
36000
Electricity demand at 20:00 (MW)
Actuals
Forecasts(interact)
Density
0.00000
0.00010
0.00020
0.00030
24000
30000
36000
Electricity demand at 20:00 (MW)

Figure



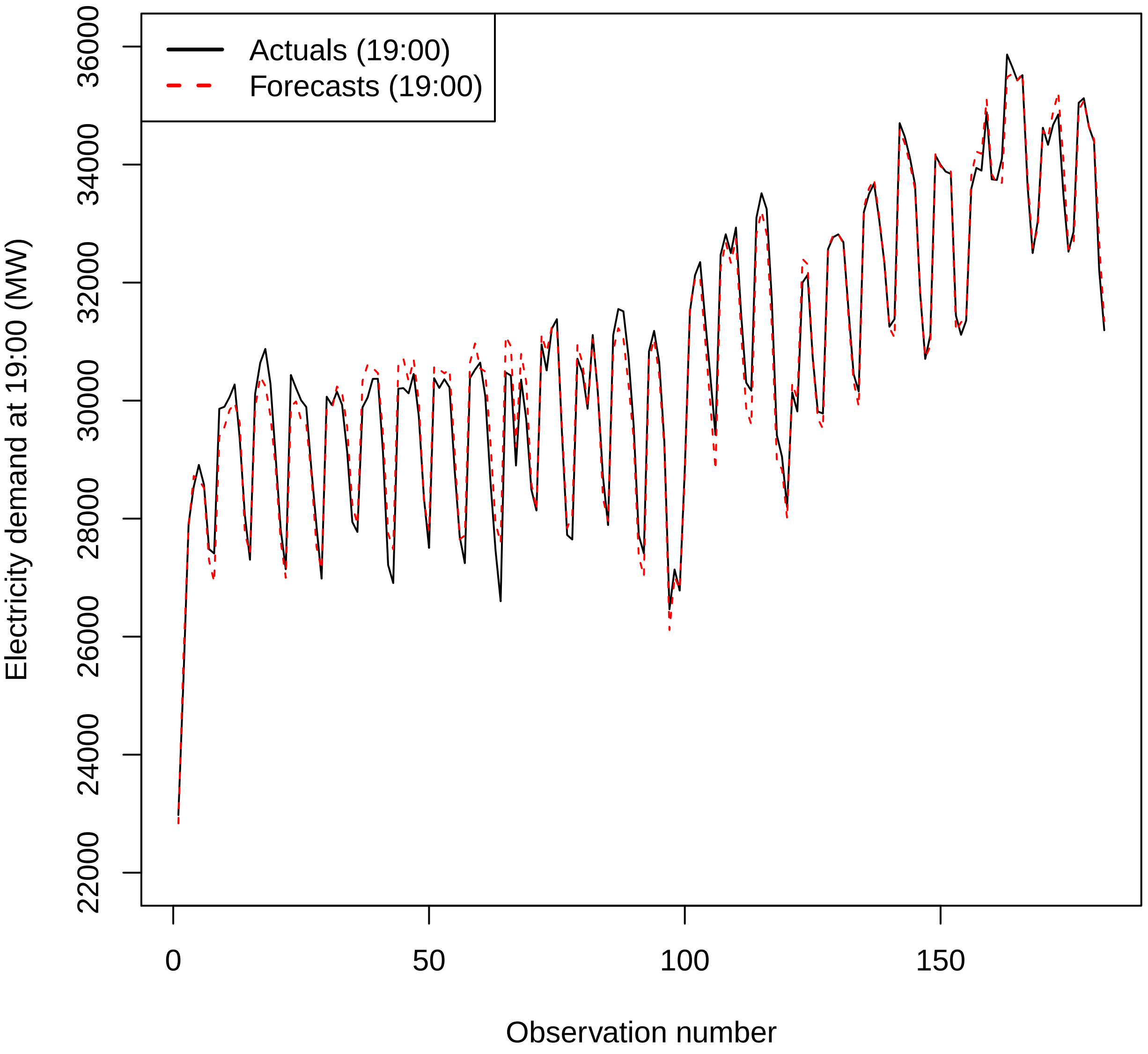
Electricity demand (MW)
35000
30000
25000
20000
H18
H19
H20
Hour

Figure



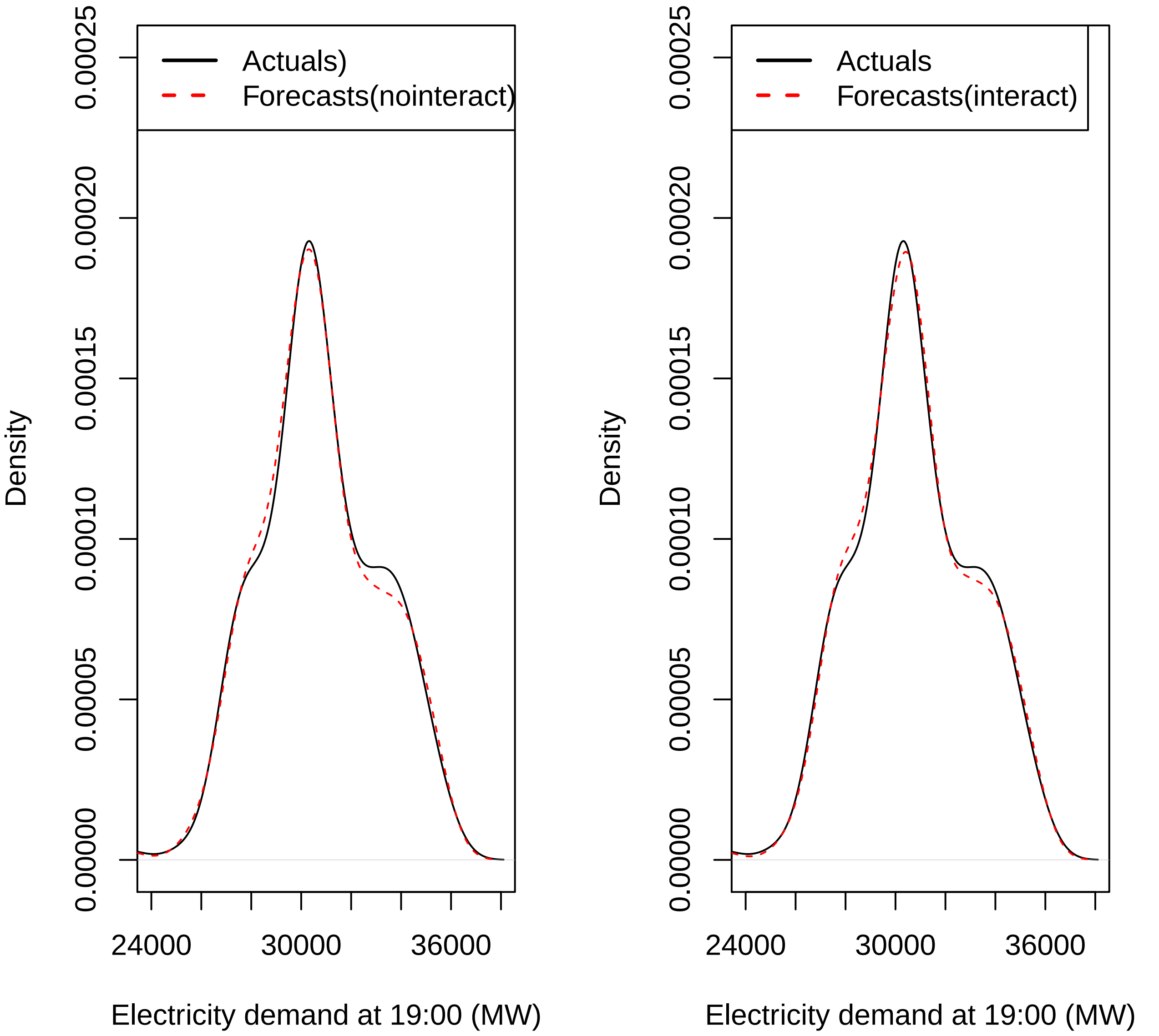
r1
0.01
−0.01
−0.03
0
200
400
600
800
1000
ACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag
PACF
0.05
0.00
−0.05
0
5
10
15
20
25
30
Lag

Figure

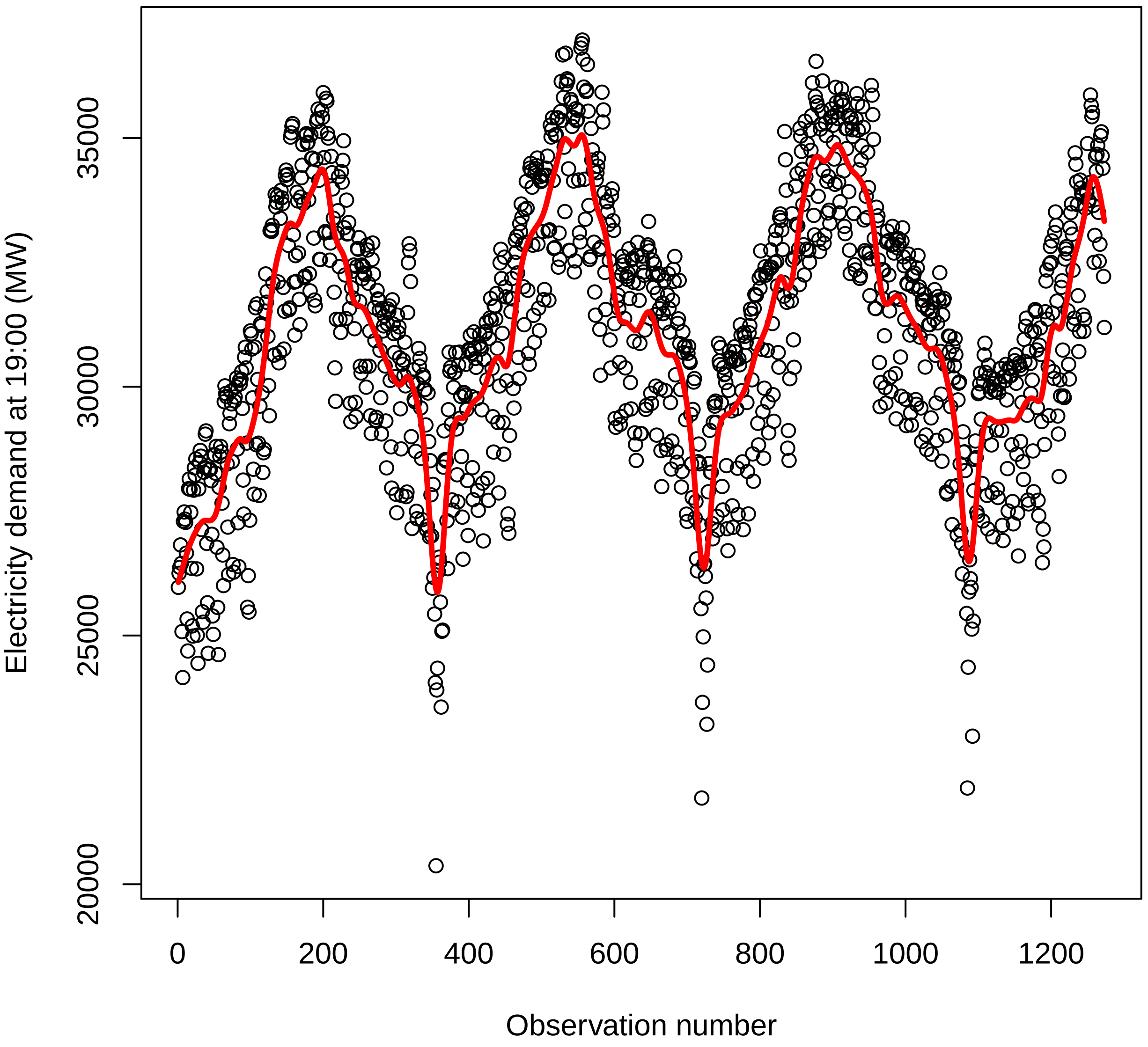


Actuals (19:00)
Forecasts (19:00)
Electricity demand at 19:00 (MW)
22000
24000
26000
28000
30000
32000
34000
36000
0
50
100
150
Observation number

Figure

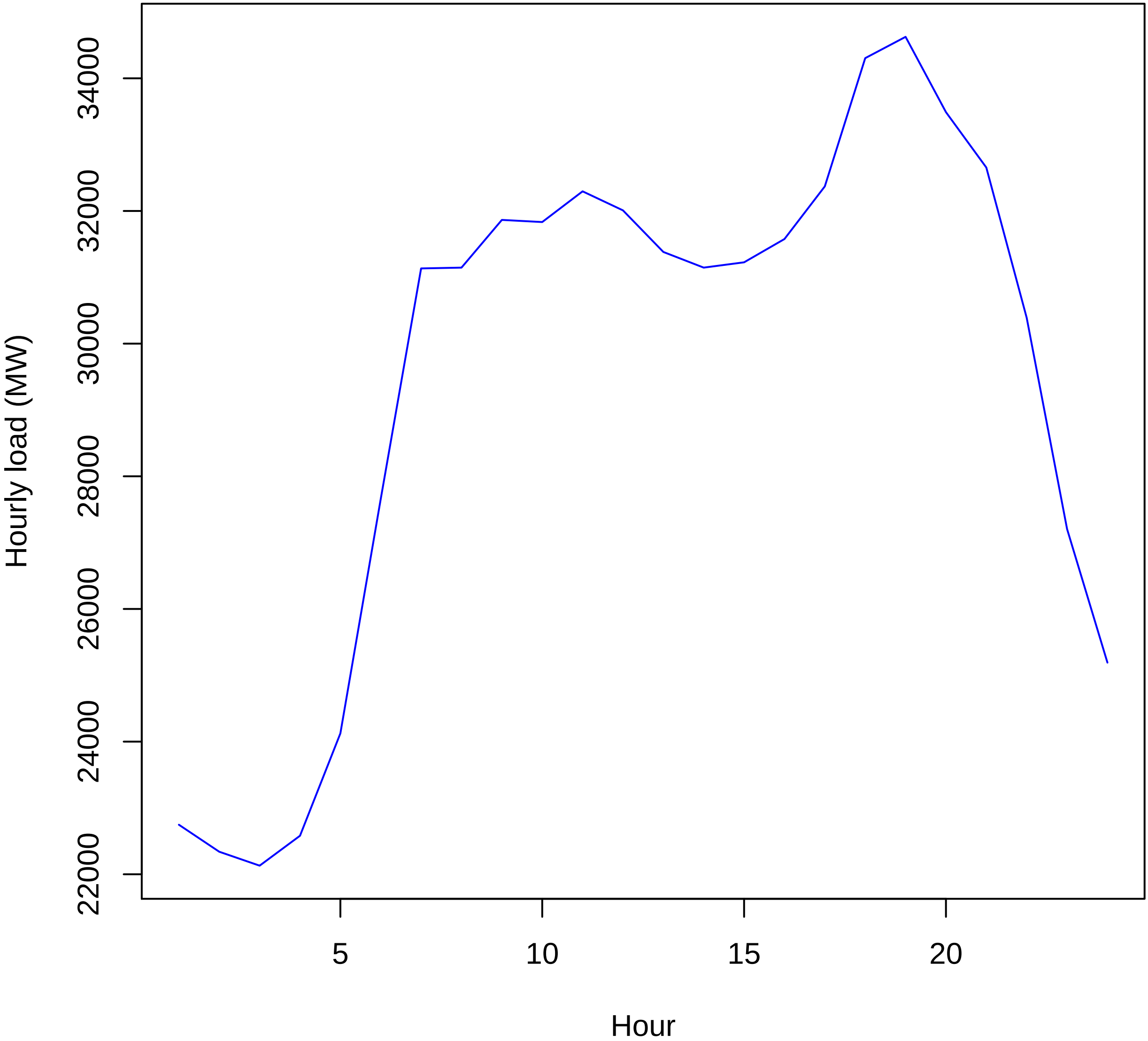

Actuals)
Forecasts(nointeract)
Density
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
24000
30000
36000
Electricity demand at 19:00 (MW)
Actuals
Forecasts(interact)
Density
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
24000
30000
36000
Electricity demand at 19:00 (MW)

Figure



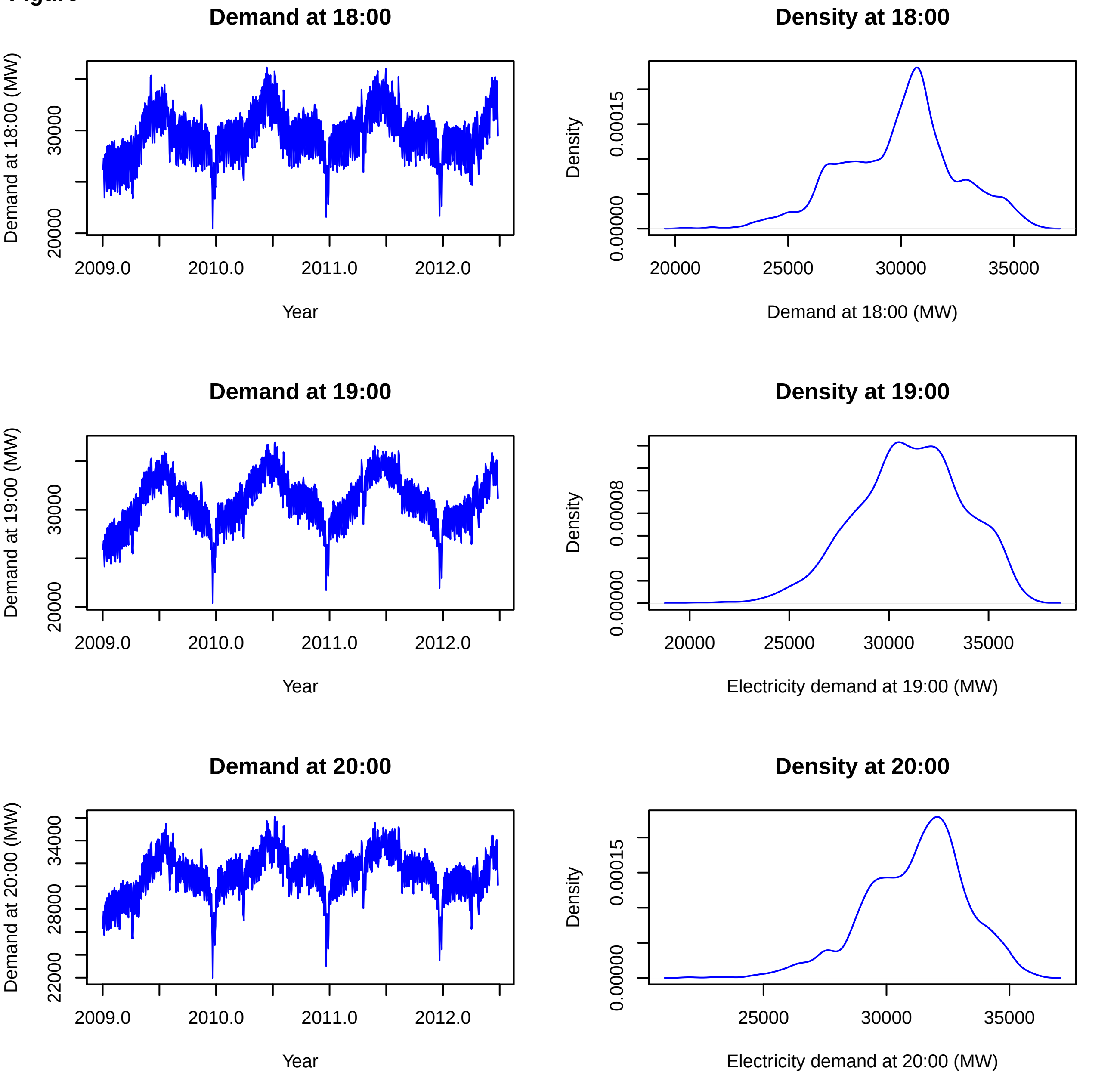
20000
25000
30000
35000
Electricity demand at 19:00 (MW)
0
200
400
600
800
1000
1200
Observation number

Figure



Hourly load (MW)
22000
24000
26000
28000
30000
32000
34000
5
10
15
20
Hour

Figure


Demand at 18:00
Demand at 18:00 (MW)
20000
30000
2009.0
2010.0
2011.0
2012.0
Year
Density at 18:00
Density
0.00000
0.00015
20000
25000
30000
35000
Demand at 18:00 (MW)
Demand at 19:00
Demand at 19:00 (MW)
20000
30000
2009.0
2010.0
2011.0
2012.0
Year
Density at 19:00
Density
0.00000
0.00008
20000
25000
30000
35000
Electricity demand at 19:00 (MW)
Demand at 20:00
Demand at 20:00 (MW)
22000
28000
34000
2009.0
2010.0
2011.0
2012.0
Year
Density at 20:00
Density
0.00000
0.00015
25000
30000
35000
Electricity demand at 20:00 (MW)