University of Venda

FORECASTING HOURLY ELECTRICITY DEMAND IN SOUTH AFRICA USING MACHINE LEARNING MODELS

By

Maduvhahafani Thanyani
11631864

submitted in fulfilment of the requirements for the degree of Master of Science in Statistics

in the

Department of Statistics
School of Mathematical and Natural Sciences
University of Venda
Thohoyandou, Limpopo
South Africa

Supervisor : Dr. Caston Sigauke

Co-Supervisor : Dr. Alphonce Bere

Date submitted: 12 August 2020

Abstract

Short-term load forecasting in South Africa using machine learning and statistical models is discussed in this study. The research is focused on carrying out a comparative analysis in forecasting hourly electricity demand. This study was carried out using South Africa's aggregated hourly load data from Eskom. The comparison is carried out in this study using support vector regression (SVR), stochastic gradient boosting (SGB), artificial neural networks (NN) with generalized additive model (GAM) as a benchmark model in forecasting hourly electricity demand. In both modelling frameworks, variable selection is done using least absolute shrinkage and selection operator (Lasso). The SGB model yielded the least root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) on testing data. SGB model also yielded the least RMSE, MAE and MAPE on training data. Forecast combination of the models' forecasts is done using convex combination and quantile regression averaging (QRA). The QRA was found to be the best forecast combination model based on the RMSE, MAE and MAPE.

Keywords: *Electricity demand forecasting, Machine learning, Lasso, neural networks, support vector regression, Forecasting, Forecast combination.*

Declaration

I, Maduvhahafani Thanyani [student Number: 11631864], hereby declare that the dissertation titled: “Forecasting Hourly Electricity Demand in South Africa Using Machine Learning Models” for the Master of Science degree in Statistics at the University of Venda, hereby submitted by me, has not been submitted for any degree at this or any other university, that it is my own work in design and in execution, and that all reference material contained therein has been duly acknowledged.

Signature: Date: 12 August 2020

Dedication

To my dearest children, my son Muano Khano and daughter Orikonisaho...

Acknowledgment

I would like to thank Almighty God for the the gift He bestowed upon me, the gift of life, strength and wisdom.

To my supervisor, Dr Caston Sigauke, words fail me to express how grateful I should be for the patience, guidance, support and motivation throughout this project. May your territory keep increasing. You are one of a kind, Sir! To my co-supervisor, Dr Bere, I should be grateful for the input and support in everything.

I should also be grateful to the University of Venda for accepting me as their student, the DST-CSIR National e-Science Postgraduate Teaching and Training Platform (NEPTTP) for the financial support. What would I have done without my colleagues! Tendani, the phone calls, the chats, your patience, all highly appreciated. Masala, Lucky, you were my strongholds through it all. Masala, I cannot thank you enough for always being there!

To my kids, son Muano Khano and daughter Orikonisaho, soon you will understand why Mommy was sometimes so busy!!! You got me through the most difficult time of my life! I almost gave up on this project so many times, but your smiles kept me going my babies...This is for you, KhanOri. To my family, I am absolutely nothing without you. Thank you for everything and may God bless you all.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Electrical energy is one of the most widely used energy forms in the world. The amount of electricity generated balanced with the electricity drawn from the grid is the electricity load. During periods when there are no load-shedding, blackouts and the available electricity that is generated from sources such as renewable energy, the the electricity load and electricity demand are equivalent. In this study, the hourly electricity demand is defined as the amount of electricity (load) in MegaWatts (MW) sent out every hour by Eskom to meet the demand of the consumers (Mokilane et al., 2018).

Electricity has, over time, become a product that is bought and sold for various purposes in the market environment. Forecasting which is the first step of planning has become much more important and has been made mandatory for the market participants by energy market regulators. In this study, short term electricity load forecasting is done for 24 hours ahead.

Being a middle income country, South Africa falls under one of the most industrialized countries in the continent of Africa. Electricity has numerous utilities in the industries, as well as for commercial and household purposes. Stats (2011) reported that a 1996 census revealed that only 57.6 percent of the households in the country had access to electricity for lighting. Lehohla (2005) reported that in 2001 the census showed an increase in household electricity access by up to 70.2 percent.

The 2007 Community Survey indicated that 80.1 percent of the South African households had access to electricity for lighting (Stats, 2011). The 2011 census showed that this percentage went up to 84.7 percent (Stats, 2011).

From the year 1996 to 2007, it was shown by the evidence from censuses and surveys that the newly connected households in the electricity grid could result in increased demand of household electricity. There was stability in the percentage of newly connected households between the 2007 and 2011, indicating that household electricity demand would be expected to have been stable during this period. One of the contributing factors in the decline in electricity demand could have been the economy that shrinked between 2007 and 2015 (Lehohla, 2005).

A decline in electricity demand from Eskom could have been contributed to by the introduction of other sources of electricity such as renewable energies, such as solar and wind. Inglesi and Pouris (2010) report that some companies and households had to find other sources of electricity, since Eskom experienced a lack of capacity in the generation of electricity in 2007, causing a decline in electricity demand. To date, the actual capacity of electricity demand market is unclear because of the unavailability of certain types of data, such as the renewable energy and other forms of electricity generation (Mokilane et al., 2018).

The combined effect of all these changes in the demography, economy and usage patterns can be investigated using historical patterns, but contribute to uncertainties when trying to forecast future electricity demand (Mokilane et al., 2018).

1.1 Background

Accurate short term electrical load forecasting results in economic cost savings and increased security in operating conditions, allowing electrical utilities to commit their own production resources in order to optimize energy prices as well as exchanges with neighbouring utilities.

In forecasting, there are uncertainties. When developing forecasts for an unknown future, statisticians need to make provisions of quantities of such uncertainties to the decision makers. Sigauke (2014) reports that the growing usage of technologies that utilize electricity, growth in population, general randomness in individual usage of electricity, seasonal effects, prevailing economic patterns, change in weather conditions, escalating costs, use of power saving electrical appliances and the growing sources of renewable energies could result in the uncertainties in future electricity demand.

The inherent uncertainties in predictions imply that forecasts should ideally be probabilistic; in other words, they should take the form of probability distributions over future quantities or events (Gneiting and Katzfuss, 2014). Probabilistic forecasts could take the form of quantiles, prediction intervals or density forecasts to quantify uncertainties in predictions. They are an essential ingredient of optimal decision-making (Gneiting and Katzfuss, 2014). It is important to quantify the uncertainties around the demand forecasts for planning purposes, to avoid building unnecessary infrastructure and to ensure that future electricity demand is met. Tay and Wallis (2000) define density forecasts of the realisation of a random variable at some future time as estimates of the probability distribution of the possible future values of that variable.

1.2 Statement of the problem

Electricity demand forecasting is carried out from short term to long term forecasting. Short Term Load Forecasting (STLF) is an integral part of the energy planning industry. Designing a time-ahead power market involves scheduling of demands for different divisions of electricity, including generation, transmission and distribution. STLF supports power system operators with different power system decision-making, including supply planning, generation reserve, system security, scheduling of dispatch, management of demand, financial planning and so on. The problem considered in this study consists of developing accurate forecasting models for short term hourly electricity demand in South Africa.

1.3 Purpose of the study

1.3.1 Aim

The aim of this research is to carry out hourly electricity demand forecasting in South Africa using some machine learning models.

1.3.2 Objectives

The objectives of the study are to:

- develop machine learning models for forecasting hourly electricity demand,
- combine forecasts from the developed models using quantile regression averaging,
- evaluate the accuracy of the forecasts.

1.3.3 Scope of the dissertation

Machine learning (ML) and statistical techniques will be used to predict hourly electricity demand using South Africa's electricity data from Eskom. Eskom is South Africa's national public power utility company. The hourly historical load data is collected from January 2010 till November 2011.

The statistical models which will be used in this study are SARIMA and GAM models. The developed models will then be compared with ANN and SVR models. Both the statistical and machine learning techniques will be performed using the relevant R statistical packages.

1.3.4 Significance of the study

The economic growth of any country is dependent on the security of electricity. For a country to meet the electricity demand, there should be a reliable supply of electricity at any given time. Accurate load forecasts will enable effective load shifting between transmission substations, scheduling of start-up times of peak stations, load flow analysis and power system security studies (Sigauke, 2014). The forecasting accuracy and precision would be useful for planning and strategy in the 56 electricity generation and supply of energy resources.

1.3.5 Structure of the dissertation

The rest of the proposal is organized as follows: Chapter 2 reviews existing literature on machine learning and statistical techniques as well as the summary of studies that used methods similar to the proposed ones. Chapter 3 provides general theory of the methods that will be used in this study, i.e. machine learning as well as statistical techniques. Chapter 4 provides the of the final research report. A summary of research findings, contributions, suggested areas for further studies and concluding remarks are presented in chapter 5.

Chapter 2

Literature review

2.1 Introduction

Of late, Short Term Load Forecasting (STLF) has been given a significant amount of attention (Hyndman and Fan, 2010). There is an increasing number of models for STLF models that are developed, reviewed, and published every year. This chapter provides an overview of STLF, factors that affect STLF, as well as summarized studies previously done on the proposed methodology for electricity demand forecasting.

Short term load forecasting (STLF) has been receiving significant attention of late (Hyndman and Fan, 2010). Various models for short term load forecasting are developed, applied, reviewed and published every year. This chapter provides an overview of STLF, factors that affect STLF, as well as summarized studies previously done on the proposed methodology for electricity demand forecasting.

2.2 An overview of short term load forecasting (STLF)

One of the most important field of research for efficiency and reliability in operation of energy that has emerged in the previous decade is STLF. This is primarily for its significance in the field of scheduling, contingency analysis, load flow analysis, planning, as well as maintenance of power system (Baliyan et al., 2015).

The importance of short term load forecasting has been discussed by Lira et al. (2009) as the core of power system planning, scheduling and control of electricity. Lira et al. (2009) further expressed that factors such as multiple seasonalities, temperature and effect in the calendar have major effects on the forecasting of short term electricity demand. The accuracy and efficiency in forecasting are of importance to any energy supplying entity, since the forecasts come in handy in the prevention of some implications such as overloading which include load shedding, blackouts, or overloading, which includes the production of capacity that exceeds demand, which lead to suppliers accruing more costs (Feinberg and Genethliou, 2005; Pierrot and Goude, 2011).

Electric companies have been, for a long time, using models developed for forecasting. Such models have been perfected by adding onto them structures and applications. STLF, globally, has become more relevant as a consequence of market liberalisation.

In the literature that has been done to date, short term electricity demand forecasting has had significant attention since it contributes a lot in power system control, unit commitment and electricity markets.
Unlike short term load forecasting, medium- and long-term forecasting have not received as much attention so far, although they add value for system planning and budget allocation (Hyndman and Fan, 2010). Short term load forecasting still dominates the literature.

2.3 Factors influencing STLF

The complexity in nature of the electricity demand is due to the presence of several factors in this field of study. Such factors include metereological, environmental, time, and many other factors (Gross and Galiana, 1987). There has been a number of studies done to evaluate the influence that such factors have on the short term load forecasts. Meterelogical effects have, in most studies, been the main focus as they have significant influence on electricity demand. Hinman and Hickey (2009) suggest that an inclusion of weather variables in the forecast models could improve the accuracy in the forecast. Fahad and Arbab (2014) highlight that having knowledge about weather conditions help in the reduction of operational costs.

Various weather variables can be considered for load forecasting: temperature and humidity are the most commonly used, but wind and cloud cover are often taken into account (Apadula et al., 2012). Many researchers use temperature in their forecasting models as it is one of the major drivers of electricity demand. Chikobvu and Sigauke (2013), did a study on the impact of temperature on the daily peak electricity demand in South Africa and their study established that temperature is indeed an important factor in electricity demand. Chikobvu and Sigauke (2013) also found that in South Africa, cold weather conditions increases electricity demand.

Gupta and Kaur (2012) highlight that time factors in forecasting are essential. Time factors include different seasonal effects and cyclical behaviours such as the daily and weekly oscillations, and public holidays as well. Electricity demand over the weekend is different from the demand during the week (weekend or holiday load curve is lower than the weekday curve). Such differences are a result in variation of consumers' lifestyles and patterns such as working time, leisure time, sleeping time, and other lifestyle patterns.

2.4 Statistical technique for Short Term Load Forecasting

Various models for STLF have been developed by researchers from different countries. The main purpose for this study is to find the model that will produce the best or offer accuracy in load estimates.

Statistical techniques assume the load data follow a pattern and try to forecast the value of the future load by employing different time series analysis techniques. This study explores the GAM statistical technique.

Linear regression models, a class of additive models are normally used with Generalized Additive Models (GAMs) (Jones and Wrigley, 1995). GAMs are suitable for exploring the data set and visualizing the relationship between the dependent and independent variables (Liu, 2008). Nedellec et al. (2014) use GAMs in modelling electricity demand for the French distribution network at both short and medium term time scales for more than 2200 substations. The relationship between the load and the explanatory variables was estimated by their model. The proposed model is given in equations (2.1) and (2.2).

$$y_t = \sum_{i=1}^{p} f_i(x_{it}) + \varepsilon_t, \quad t = 1, \cdots, n, \tag{2.1}$$

$$y_t = f_1(x_{1t}) + f_2(x_{2t}) + \cdots + f_p(x_{pt}) + \varepsilon_t, \quad t = 1, \cdots, n, \tag{2.2}$$

where y_t is a univariate response variable, x_{pt} are the covariates that drive f_i, the smooth functions. Non linear functions are meant to be smoothed that they can be relatively well estimated by penalized regression in a spline basis (Nagy et al., 2016). Effects that drove the French hourly load consumption was modelled using GAM and compared with the operational one in (Pierrot and Goude, 2011). The effect of different variables was estimated with GAM given in equation (2.3).

$$L_t = f_1(L_{t-24}) + f_2(L_{t-168}) + f_3(T_t) + f_4((T_{t-24}, T_{t-48}) + f_5(cc) + f_6(posit) + C + \varepsilon_t, \tag{2.3}$$

where L_t, L_{t-24} and L_{t-168} are the load to be forecasted, one day lagged load and the one week lagged load respectively, T_{t-24}, T_{t-48} are the temperature, one day lagged temperature and the two day lagged temperature, cc represents the cloud cover, *posit* is the position of the day through the year, C is the intercept and ε_t is the residual error.

2.5 Machine Learning Techniques

2.5.1 Support Vector Regression

Shmilovici (2009) defines support vector machines (SVM) as a set of associated techniques for supervised learning, that can be applied in both classification and regression models.

Vapnik (1998) developed support vector machines. Milidiu et al. (n.d.) notes that while the traditional neural network models takes into implementation the empirical risk minimization principle, which seeks to minimize the misclassification error, the SVM implements the structural minimization principle that seeks to minimize an upper bound of generalization error.

Mohandes (2002) did a study that introduced support vector machines in the electrical load forecasting and compared their performance with that of the auto-regression model. The results showed that the support vector regression performed better than auto-regressive moel, for both the training and tesing data. The study used the root mean square error for metric evauation between the actual and predicted data. Support vector machines allow the training data set to be increased beyond what is possible using the auto-regressive model or other neural networks methods. Increasing the training data further improves the performance of support vector machines method.

Hong (2009) did a study on support vector regression that applied the structural risk minimization principle to minimize an upper bound of the generalization errors, rather than minimizing the training errors which are used by ANNs. The purpose of the study done by Hong (2009) was to present a SVR model with immune algorithm (IA) to forecast the electric loads. IA was applied to the parameter determine of SVR model. The empirical results indicated that the SVR model with IA (SVRIA) results in better forecasting performance than the other methods, namely SVMG, regression model, and ANN model.

(Yuancheng et al., 2002) presented a paper on least squares support vector machines

(LS-SVM) approach to short-term electric load forecasting (STLF). The proposed algorithm was found to be more robust and reliable as compared to the traditional approach when actual loads are forecasted and used as input variables. In order to provide the forecasted load, the LS-SVM interpolates among the load and temperature data in a training data set. Analysis of the experimental results proved that this approach can achieve greater forecasting accuracy than the traditional model.

2.5.2 Stochastic Gradient Boosting

(Touzani et al., 2018) assessed the performance of stochastic gradient boosting method dataset of 410 commercial buildings. The model training periods were varied and several prediction accuracy metrics were used to evaluate the model's performance. The results showed that using the gradient boosting machine model improved the R-squared prediction accuracy and the RMSE in more than 80 percent of the cases, when compared to piecewise linear regression, and to a random forest algorithm.

Another study by Papadopoulos and Karakatsanis (2015) compared four different methods, seasonal autoregressive moving average (SARIMA), seasonal autoregressive moving average with exogenous variable (SARIMAX), random forests (RF) and gradient boosting regression trees (GBRT) in forecasting day ahead load demand. The forecasting performance of each model was evaluated by the mean absolute percentage error (MAPE) and root mean square error (RMSE). The results of this study showed that the GBRT model is superior to the others for 24 hours ahead forecasts. This study claims that gradient boosting regression trees can be appropriate for load forecasting applications and yield accurate results.

(Mayrink and Hippert, 2016) presented a hybrid method for short-term load forecasting. The study combined Exponential Smoothing, a classical method for time series forecasting, with Gradient Boosting, a powerful machine learning algorithm. The proposed model was tested with real data and the results showed a considerable improvement in forecasting accuracy.

2.5.3 Neural Networks

Pao (2006) forecasted Taiwan energy consumption by neural networks and linear models. Neural network functioned better than the linear models. But depending on situation, accuracy of Artificial Neural Network (ANN) methods decreases because of several reasons. Forecasting accuracy of ANN depends on learning data set and their adequacy. Moreover ANN methods sometimes get stuck in local minimum, so choosing proper data set, is too critical in neural network models and these models get good results only when the number of data is high (Padmakumari et al., 1999).

Ho et al. (1992) designed an algorithm in which the momentum is automatically adapted in the training process. Lee et al. (1992*a*) proposed a nonlinear electricity forecasting model and several structures of ANNs were tested. Inputs to the ANN include past electricity values, and the output is the forecast for a given day. Lee et al. (1992*a*) also demonstrated that the ANN could be successfully used in STLF with accepted accuracy.

(Lee et al., 1992*b*) Applied artificial neural network (ANN) method to forecast the short-term load for a large power system. The load had two distinct patterns: weekday and weekend-day patterns. The sudy proposed a nonlinear load model is proposed and several structures of an ANN for short-term load forecasting were tested. Inputs to the ANN are past loads and the output of the ANN is the load forecast for a given day. The network with one or two hidden layers was tested with various combinations of neurons, and results are compared in terms of forecasting error. The results of the study showed that neural network, when grouped into different load patterns, gave good load forecasts.

A study was done by Din and Marnerides (2017) on the application of the Feed-forward Deep Neural Network (FF-DNN) and Recurrent Deep Neural Network (R-DNN) models on the basis of accuracy and computational performance in the context of time-wise short term forecast of electricity load. The outcome of the study showed that Feed-forward Deep Neural Network (FF-DNN) has higher accuracy in the forecasting of electricity demand.

2.6 Conclusions from Literature

The use of GAM, SVR, SGB and NN on electricity demand forecasting has been dicussed in detail in this chapter. This research project seeks to investigate the application of the above mentioned models in the forecasting of electricity in South Africa.

Chapter 3

Methodology

3.1 Introduction

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

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

3.2 Generalized Additive Model (GAM)

The generalized additive model (GAM) was first introduced by Hasties and Tibshirani (1990) and follows from the additive models. It has relation to a generalized linear model (GLM) but additionally the linear predictors include a sum of smooth functions of the independent variables and a consistent parametric component of the linear prediction. Henceforth, GAM incorporates both parametric and nonparametric model components and the model can indicate the dependence of the response on the independent variable in a flexible way, not relying upon the assumption that all relations can be modelled as linear. The general formula for GAM is given by as follows Hasties and Tibshirani (1990):

$$g(\mu_t) = \mathbb{A}\theta + \sum_{j=1}^{p} f_j(x_{jt}) + \varepsilon_t, \quad Y_t \sim EF(u_t, \phi)t = 1, \cdots, n, \tag{3.1}$$

where Y_t denotes the independent univariate response variable from an exponential family distribution having mean u_t, scale parameter ϕ, g represents the smooth monothonic link function, $\mathbb{A}$ is a design matrix of order n X k, θ represents an unknown parameter vector of order k X 1, f_j is an unknown smooth function of the predictor variable x_j that may have a vector value, and ε_t is an independent identical distribution random error k X 1 (**??**). The non-linear (non-parametric) functions are assumed to be smooth. It can be relatively well estimated by penalised regression in a spline basis. Each function is given by

$$f_j(x_{jt}) = \sum_{t=0}^{k_j} \beta_{j,t} b_t^j(x_{jt}) \tag{3.2}$$

where k_j denote the dimension of the spline basis to model the effect $f_t(x_{jt})$ and b_t^j is the corresponding spline functions, i.e. B-splines or cubic regression splines and $\beta_{j,t}$ are coefficients to be estimated.

3.3 Stochastic Gradient Boosting (SGB)

The Gradient Boosting Machine or Gradient TreeBoost were the terms previously used for Gradient Boosting when it was implemented newly by Friedman (Friedman, 2002). Gradient Boosting is a Machine learning model used for classification and regression problems (Mpfumali et al., 2019). It stage-wisely builds weak predictive models generalised by optimization of arbitrary differentiateable function. The Statistical framework of gradient boosting describes it as an optimization problem which minimizes the loss in a model by a stage-wise addition of weak learners to the models using a gradient descent procedure (Friedman, 2002; Mpfumali et al., 2019).

Gradient descent traditionally minimizes set of parameters such as coefficient of regressors or ANN weights through loss or error calculation and weight update (Friedman, 2001). The weak learners are organised in substructures or decision trees that replaces the parameters. Parameterized tree is added to the model, thereby reducing the error and the residual losses using the parameters of the trees following the direction of the gradient (Friedman, 2001). The gradients spots the error in the weak learners. The major drawback to the gradient boosting is that it is a greedy algorithm that can easily over fit training data (Friedman, 2001; Hastie et al., 2005). One of the variants of gradient boosting is the stochastic gradient boosting (SGB) formed by taking a random sample of the training data set without replacement (Friedman, 2002, 2001). Its general formulation is given in equation (3.15).

$$f(x) = \sum_{m=1}^{M} \beta_m h(x; \gamma_m), \tag{3.3}$$

where $h(x;\gamma_m) \in \mathbb{R}$ are functions of x which are characterised by the expansion parameters γ_m and β_m. The parameters γ_m and β_m are fitted in a stage-wise way, a process which slows down over-fitting Friedman (2002).

The use of SGB in electricity demand forecasting is rare. Hence, this research investigates the use of SGB specifically as a variant of the gradient boosting models that have been mostly used in forecasting problems in the literature. This will enhance the provision of knowledge to the forecasting and machine learning community about the problems SGB is capable of solving.

3.4 Neural Networks (NN)

The non-linear aspects of the time series can be uncovered by the assistance of Artificial Neural Network (ANN) model. The structure of the ANN is similar to that of human brain in that the neurons are substituted by nodes which are arranged in various layers.

This study focuses on one Feed Forward Neural Network (FFNN) which is a type of Artificial Neural Network (ANN) in which connections between the nodes do not form a cycle or a loop. The study of this technique was first initiated by McCulloch and Pitts (1943) who created a computational model of the concept. Different researchers have further expanded the concept of ANN to cover many features

3.4.1 Single-layer perceptron

Single-layer perceptron is the simplest type of a neural network which has a single layer of output nodes. A single-layer neural network can be described mathematically as follows,

$$y_k = g\left(\sum_{i=0}^{D} \omega_i x_i\right), \tag{3.4}$$

where y_k is the output, $g(.)$ is an activation function, x_i is input and ω_i represent the corresponding weight for x_i. Single-layer neural networks are not usually used in practice, but help in understanding the basic concept of neural networks.

3.4.2 Multi-layer perceptron

Multi-layer perceptron consists of multiple layers of computational units, usually interconnected in a feed-forward way. A multi-layer neural network with one hidden layer can be written as,

$$y_k = h\left(\sum_{j=0}^{M} \omega_{kj}^{(2)} g(a_j)\right) \tag{3.5}$$

where,

$$a_j = g\Big(\sum_{i=0}^{D} \omega_{ij}^{(1)} x_i\Big). \tag{3.6}$$

The multi-layer neural network is very similar to single-layer neural network except that multi-layer neural network's output of the inner layer is again multiplied by a new weight vector and wrapped in an activation function.

3.5 Support Vector Regression

Support Vector Regression (SVR) is based on Support Vector Machine (SVM) which is a supervised machine learning technique which involves statistical learning theory and the principle of structural risk minimization. SVR was introduced by Drucker et al. (1997) extended from SVM model. There are different basic kernel functions that are used in SVM models, which can be classified as polynomial (Poly), Gaussian kernel, exponential radial basis function (ERBF), radial basis function (RBF), sigmoid and linear (Zendehboudi et al., 2018). The SVR works by mapping the input space into a high-dimensional feature space and constructs the linear regression in it which can be expressed as,

$$f(x) = \omega\phi(x) + b, \tag{3.7}$$

where ω is the weight vector, $\phi(x)$ maps inputs x into a high dimensional feature space that is nonlinearly mapped from the input space x, and b is the bias term.

To calculate the coefficients ω and b it is required to reduce the regularized risk function which can be expressed as:

$$\frac{1}{2}\|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\epsilon(y_i, f(x_i)), \tag{3.8}$$

where, $\|\omega\|^2$ is a regularized term which maintains the function capacity. C is a cost error. The empirical term from the second term in equation 3.8 can be defined as:

$$L_\epsilon(y_i, f(x_i)) = \{|y_i = f(x_i)| = \epsilon, |y_i = f(x_i)| \geq \epsilon\}. \tag{3.9}$$

The equation 3.8 expressed the transformation of the primal objective function in order to get the values of ω and b by introducing the positive slack variables $\xi_i(^*)$.

$$\text{minimize} \quad \frac{1}{2}\|\omega\|^2 + C\frac{1}{l}(\epsilon_i + \xi_i^*) \tag{3.10}$$

subject to

$$\alpha(x) = \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

The optimization problem expressed in equation 3.10 has to be transformed into its dual formulation by using the Lagrange multipliers to solve it in a more efficient way.

Data

The data used in this project is obtained from Eskom. The dataset will be split into training dataset (80%) and testing dataset (20%).

Features

All models developed in this project will use load to predict hourly electricity demand.

3.6 Variable Selection

Variable Selection involves selection of feature variables that explains a target variable thereby reducing number of feature variables. This process is beneficial in terms of avoiding overfitting, making a model easier to interpret, and reduces in computational time. There are many variable selection methods, however, in this study we use Lasso (least absolute shrinkage and selection operator) via hierarchical interactions (Bien et al., 2013). LASSO via hierarchical interactions considers pairwise hierarchical interactions only, however, it can be extended to higher order interactions. We assume a regression model of respond variable Y and predictors $X_1,...,X_p$ with pairwise interactions between these predictors. Our LASSO hierarchical model is given as

$$Y = \beta_0 + \sum_j \beta_j X_j + \frac{1}{2}\sum_{j \neq k} \Theta_{jk} X_j X_k + \varepsilon, \tag{3.11}$$

where $\varepsilon \sim N(0, \sigma^2)$, $\beta \in \mathbb{R}^p$, and $\Theta \in \mathbb{R}^{p \times p}$. There are two categories of hierarchical restrictions, which are strong and weak hierarchy.

$$\begin{aligned} \textit{Strong hierarchy}: \hat{\Theta}_{jk} &\neq 0 \Rightarrow \beta_j \neq 0 \textit{ and } \hat{\Theta} \neq 0 \\ \textit{Weak hierarchy}: \hat{\Theta}_{jk} &\neq 0 \Rightarrow \beta_j \neq 0 \textit{ or } \hat{\Theta} \neq 0 \end{aligned}$$

The major advantage of lasso via hierarchical interactions is that it leads to simpler and more interpretable models that involve only a subset of the predictors.

3.7 Forecast Combination

Forecast combination is a method used to combine forecasts from different fitted models with a purpose of improving forecast accuracy. There are many forecast combination methods, but this research will be based only on Quantile regression averaging (QRA) and convex combination method.

3.7.1 Quantile Regression Averaging

Quantile Regression Averaging was first initiated by Maciejowska et al. (2016). QRA treats forecasts from different models as independent variables and actual observations as a dependent variable. Let $\hat{y}_{t,\tau}$ be hourly electricity load, K be methods used to forecast the next m observations, i.e. m is the total number of forecasts. The forest combination, $\hat{y}^{QRA}_{t,\tau}$, is given by

$$\hat{y}^{QRA}_{t,\tau} = \beta_0 + \sum_{k=1}^{K} \beta_{t,k}\hat{y}_{t,k} + \varepsilon_{t,\tau}, \quad \tau \in (0,1), \quad t = 1, ..., m, \tag{3.12}$$

where $\hat{y}_{t,k}$ represents predictions from method k, $\hat{y}_{t,\tau}^{QRA}$ is the combined forecasts, and $\varepsilon_{t,\tau}$ is the error term. QRA aims to minimize

$$_{\beta} \sum_{t=1}^{n} \rho_{\tau}(\hat{y}_{t,\tau}^{QRA} - \beta_0 - \sum_{k=1}^{K} \beta_{t,k} \hat{y}_{t,k}). \tag{3.13}$$

In matrix form, we have

$$_{\beta \in IR} \sum_{t=1}^{n} \rho_{\tau}(\hat{y}_{t}^{QRA} - x_t^T \beta). \tag{3.14}$$

3.7.2 Convex Combination

$$\hat{y}_{t,\tau}^{c} = \sum_{m=1}^{M} \omega_{mt} \hat{y}_{mt,\tau}, \tag{3.15}$$

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

Convex combination method computes the sequence of instantaneous losses suffered by the predictions from the experts (models) using a loss function. The loss function can be based on square, absolute, percentage, or pinball loss. The combined forecasts will be compared with forecasts from each model using the equation given as

$$\hat{y}_{t,\tau}^{c} = \sum_{m=1}^{M} \omega_{mt} \hat{y}_{mt,\tau}, \tag{3.16}$$

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

3.8 Prediction Intervals

The prediction interval widths (PIWs) for every model, $M_j, j = 1, ..., k$, are denoted as $\text{PIW}_{ij}, i = 1, ..., n, j = 1, ..., k$.

PIW_{ij} is calculated as

$$\text{PIW}_{ij} = \text{UL}_{ij} - \text{LL}_{ij}, \tag{3.17}$$

where UL_{ij} and LL_{ij} are the upper and lower limits of the prediction interval, respectively. Probability density plots and box and whisker plots will be used in this study to find the model which which yields narrower PIWs.

3.9 Evaluation of Forecasts

Root mean square error

The Root mean square error (RMSE) is a measure of the differences between predicted values by the model and the actual observed values.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}, \tag{3.18}$$

where, $\hat{y}_t$ are predicted values by the model, y_t are the values actually observed, and n is the number of predictions.

Mean absolute error

The mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) are going to be used to evaluate the accuracy of our forecasts. The equations of the above measures are respectively given by: Mean Absolute Error (MAE) is a measure of how close forecasts are relative to the values actually observed. MAE is given by:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} y_t - \hat{y}_t, \tag{3.19}$$

where, $\hat{y}_t$ is the predicted value by the model, y_t is the value actually observed, and n is the number of fitted points. MAE uses the same scale as the data being measured. This is known as a scale-dependent accuracy measure and therefore cannot be used to make comparisons between series using different scales.

Mean absolute percentage error

The mean absolute percentage error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics. The accuracy of a fitted model is expressed as a percentage by the following equation:

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n} \frac{y_t - \hat{y}_t}{y_t}, \tag{3.20}$$

where, $\hat{y}_t$ are predicted values by the model, y_t are the values actually observed, and n is the number of predictions.

3.10 Implementation

The GAM and SVR algorithm will be implemented by using the relevant packages in R software.

Chapter 4

Data Analysis

4.1 Introduction

This chapter presents detailed analysis of the data using the methods discussed in Chapter 3. Generalised additive models (GAMs) are discussed and compared to the machine learning models; support vector regression (SVR), stochastic gradient boosting (SGB), and feed forward neural network (NN), for the forecast of hourly electricity demand in South Africa.

4.2 Data

4.2.1 Data source

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. The data is divided into two subsets: a training set and a testing set to build the base forecasting methods and evaluate the performance.

4.2.2 Exploratory data analysis

The summary statistics of hourly electricity demand for the sampling period January 2010 to November 2011 is given in Table 4.1. The distribution of hourly load is not normal since it is skewed to the left and platykurtic as shown by the skewness value -0.242 and a kurtosis value of -0.989 given in Table 4.1 .

Table 4.1: Summary statistics for hourly electricity demand (MW).

Descr. Stats.	**Min**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Skewness**	**Kurtosis**
Load	19563	24963	28809	28114	30732	36664	-0.2423795	-0.9895812

It is considered important to first get the analysis of historical data before setting up the forecasting models. The time series, density, box and normal (Q-Q) plots for the dependent variable (Load (in MW)) are given in Figure 4.1

(a) Top left panel: Plot of DPED (b) Top right panel: Probability density function of HED. The distribution is left skewed. (c) Bottom left panel: Normal QQ plot of HED (d) Bottom right panel: Box plot of HED, which all show that the data is not normally distributed.

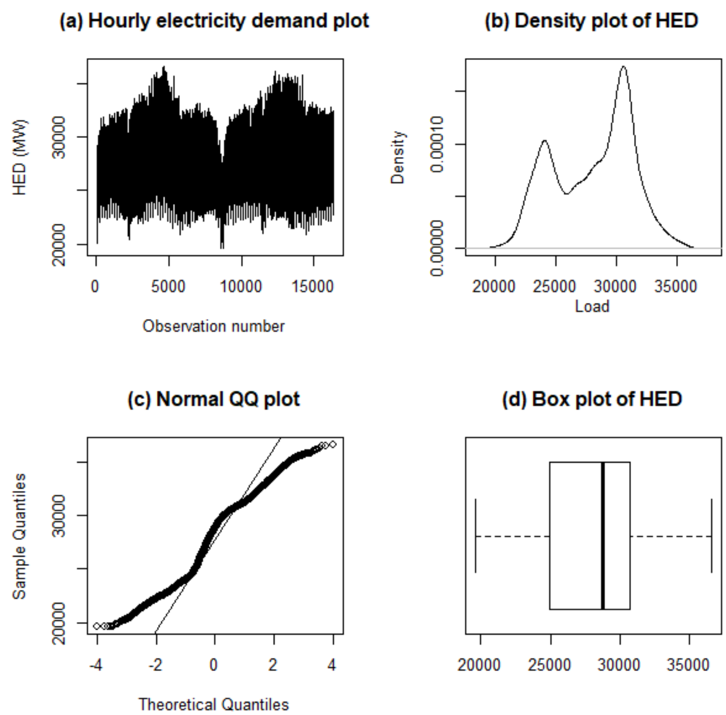

Figure 4.1: Diagnostic plots for hourly electricity demand in South Africa.

The non-linear trend values used in this project are obtained or extracted by tting a cubic smoothing spline function which is given by:

$$\pi(t) = \sum_{t=0}^{n} (y_t - f(t))^2 + \lambda \int f''^2(t) dt, \tag{4.1}$$

where *lambda* is a smoothing parameter which is estimated using the generalised cross validation (GCV) criterion. Figure 4.2 shows the non-linear trend, cubic smoothing spline fitted with an estimated lambda value. The fitted non-linear trend values are extracted and used to model hourly electricity demand.

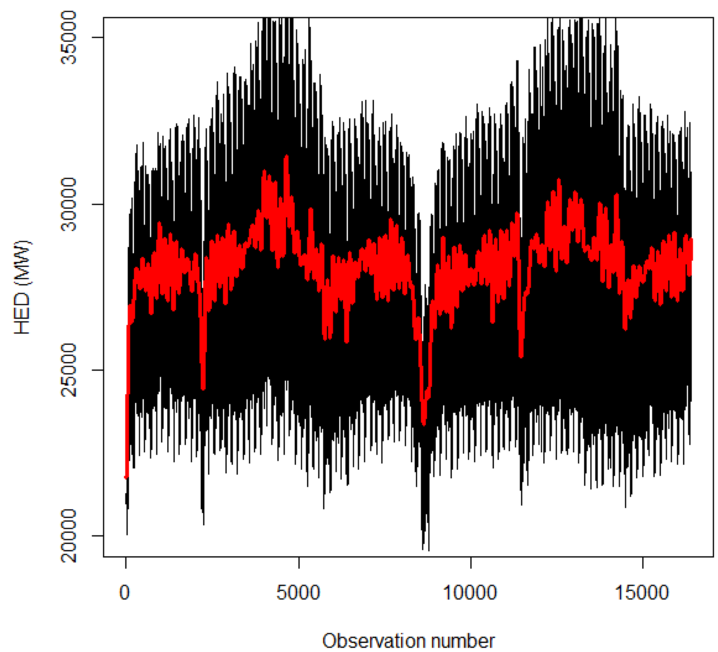

Figure 4.2: Plot of hourly electricity demand from 1 January 2010 to 15 November 2011 superimposed with a fitted cubic smoothing spline trend (non-linear trend).

Variable selection using Lasso

Variable selection using Lasso is done and found that all variables are significant as shown in Table 4.2.

Table 4.2: Variable selection using Lasso.

Variable	**Coefficient**
Intercept	$2.2328x10^4$
hour	$1.8705x10^2$
month	$3.5518x10^2$
daytype	$1.1975x10^2$
holiday	$-2.1699x10^3$
minT	$-1.6420x10^1$
maxT	$2.3024x10^2$
aveT	$-1.3994x10^2$
nolTrend	$1.4017x10^1$

4.3 Machine Learning Models

The models considered for this study are the M2 (SVR), M3 (SGB), and M4 (NN) and M1 (GAM) which is the benchmark model. RMSE, MAE and MAPE are normally used for forecast evaluation and this study considers them. Table 4.3 and Table 4.4 shows a comparative analysis of the fitted machine learning models and the benchmark model.

Table 4.3 summarises the error measures for evaluation of the models on the training set. From Table 4.3, model M3 (SGB) has the lowest RMSE (760.042), MAE (565.936) and MAPE (2.005). This indicates that M2 outperporms all other models and thus it is the best forecasting model on training data set.

The M3 (SGB) has the least RMSE (392.629), the least MAE (304.687) and also the least MAPE (1.079), showing that it is the best fitting model according to the summary of the error measures for evaluation of the models on the testing data set.

Table 4.3: Comparative analysis of the fitted models (Evaluation of the models on training data).

	M1 (GAM)	M2 (SVR)	M3 (SGB)	M4 (NN)
RMSE	3275.55	2572.83	760.042	3354.34
MAE	2783.628	2020.61	565.936	3626.45
MAPE	10.416	7.149	2.005	13.339

Table 4.4: Comparative analysis of the fitted models (Evaluation of the models on testing data).

	M1 (GAM)	M2 (SVR)	M3 (SGB)	M4 (NN)
RMSE	3107.431	1112.273	393.621	3539.911
MAE	2641.669	864.627	304.687	2649.795
MAPE	9.456	3.157	1.079	2.439

The graphical plot of the out of sample forecasts for the models M1, M2, M3 and M4 are given in the following figures:

Figure 4.3 : Graphical plot of model M1 (GAM) forecasts (dashed line) and actual hourly electricity demand (solid line).

Figure 4.4 : Graphical plot of model M2 (SVR) forecasts (dashed line) and actual hourly electricity demand (solid line).

Figure 4.5 : Graphical plot of model M3 (SGB) forecasts (dashed line) and actual hourly electricity demand (solid line).

Figure 4.6 : Graphical plot of model M4 (NN) forecasts (dashed line) and actual hourly electricity demand (solid line).

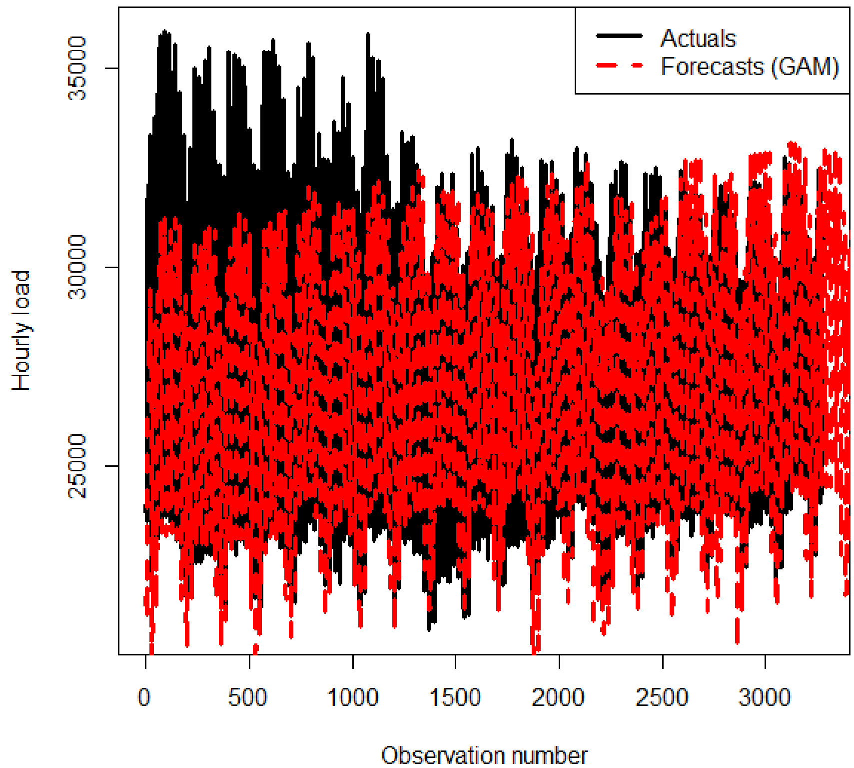

Figure 4.3: GAM forecasts.

4.4 Forecast Combination

This section gives results of forecast combination of machine learning models forecasts. Two forecast combination methods are used, which are convex combination and quantile regression averaging.

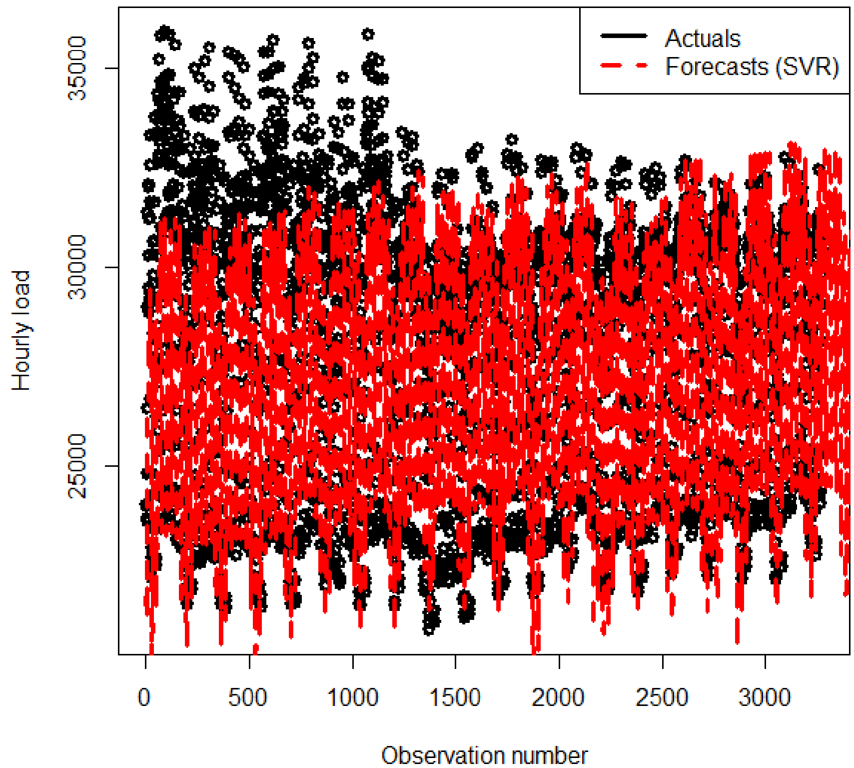

Figure 4.4: SVR forecasts.

4.4.1 Convex Combination

It is known that combining forecasts often leads to better forecast accuracy. The forecasts from different regression based time-series which are fitted above may be improved by combining them using R package called opera developed by Devaine et al. (2013). The models developed are also referred to as experts. To combine fore-

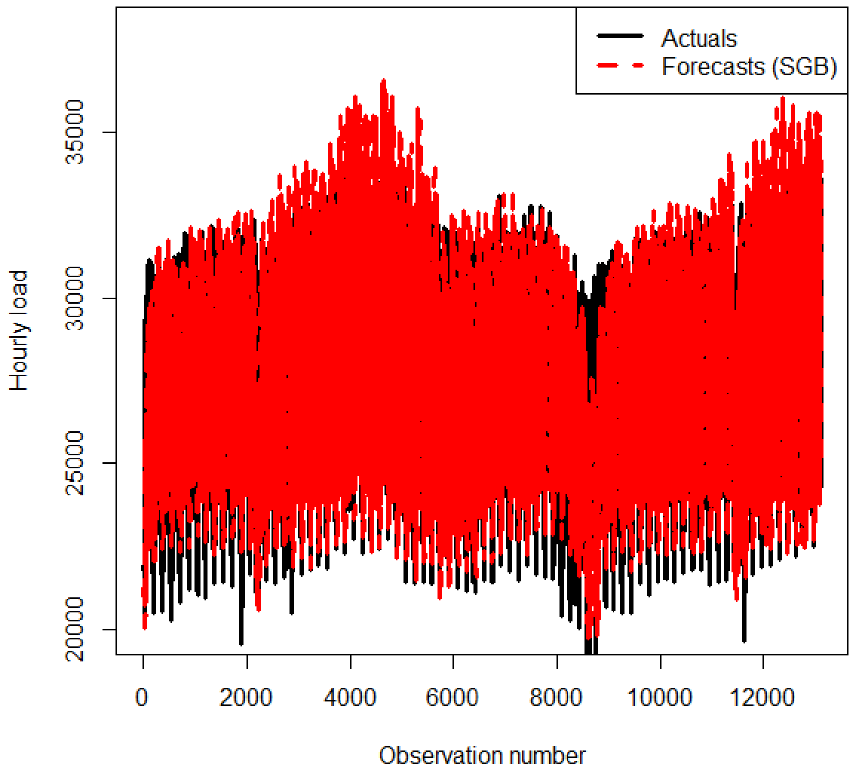

Figure 4.5: SGB forecasts.

casts we find aggregation of experts and analyse them by looking at the oracles. The opera package computes weights when combining the forecasts. Convex combination method works by computing the sequence of instantaneous losses suffered by the predictions from the experts (models) using loss function.

The loss function can be based on square, percentage, or pinball loss. Table 4.5 sum-

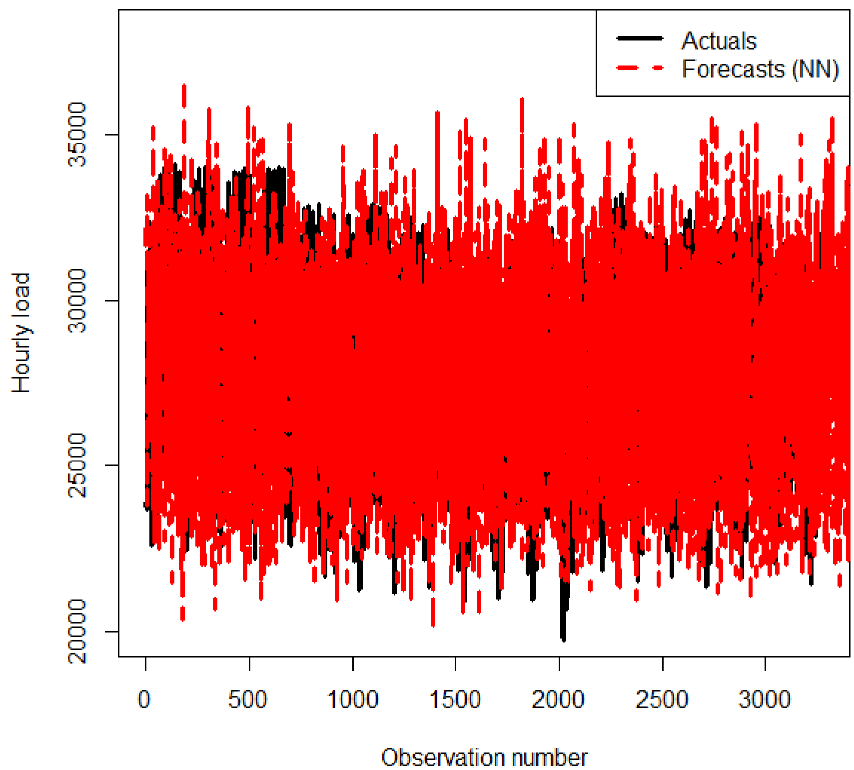

Figure 4.6: NN forecasts.

marises the results for forecast combination models. It is giving weight 0.264 to model M2, 0.063 to model M3, and 0.721 to model M4 when the loss function is specified as square. The pinball and absolute loss function are giving weight 0.992 to model M2, 0 to model M3, and 0.0083 to model M4. The percentage loss function is giving weight 0.879 to model M2, 0 to model M3, and 0.0065 to model M4. From the accuracy measures Mix 3 has the least RMSE (1164.774), MAE (935.214) and MAPE (3.402) compared to other mix models.

Table 4.5: Models (experts).

	M2	M3	M4
Square (Mix 1)	0.264	0.063	0.721
Pinball (Mix 2)	0.992	0	0.0083
Percentage (Mix 3)	0.879	0	0.0065

Table 4.6: Accuracy measures.

	Mix 1	Mix 2	Mix 3
RMSE	1588.83	1479.104	1164.774
MAE	1372.697	1252.565	935.215
MAPE	3.402	4.607	4.95

Figure 4.7 shows that the M3 (SGB) model is the best forecasting model followed by the convex combination model, M2 (SVR), uniform combination (Uniform), and model M4 (NN) respectively.

4.4.2 Quantile Regression Averaging

Quantile regression averaging (QRA) is another technique normally used to combine forecasts by using forecasts from each model as independent variables. The three models M2 (SVR), M3 (SGB) and M4 (NN) are combined based on QRA, resulting in model M3. Model M6 (QRA) is given by:

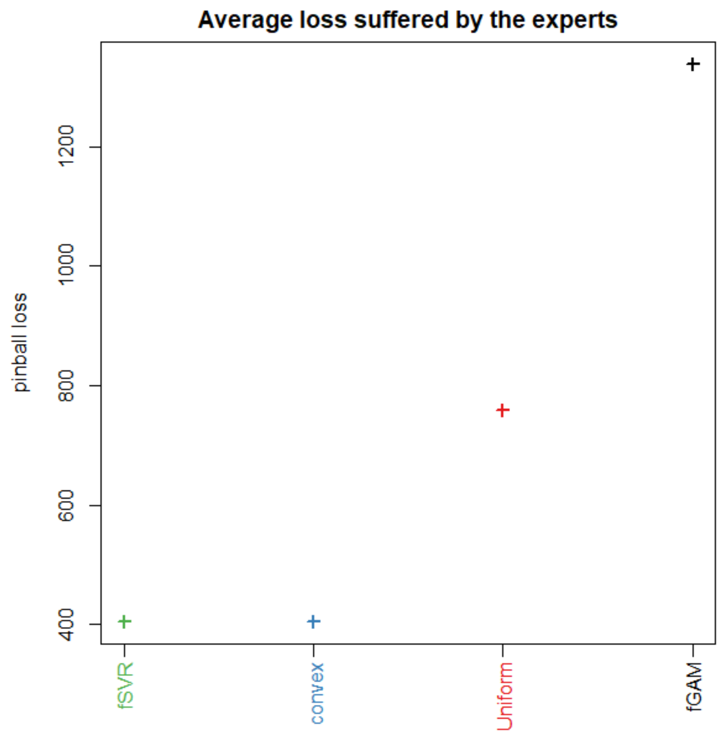

Figure 4.7: Average loss suffered by the models.

$$y_{t,\tau}(QRA) = \beta_0 + \beta_1 fM2 + \beta_2 fM3 + \beta_3 fM3 + \epsilon_t. \tag{4.2}$$

where fM2, fM3 and fM4 represent the forecasts from models M2, M3 and M4 respectively.

Table 4.7 gives a summary of the accuracy measures for the models, including M5 (Convex) and the M6 (QRA) model. Based on MAE, model M2 is the best forecasting model compared with other models. MAE shows a slight improvement after forecast

Table 4.7: Comprative analysis of the machine learning models, Convex model and QRA model.

Accuracy measure	M2	M3	M4	M5 (Convex)	M6 (QRA)
RMSE	1112.273	392.629	3539.911	1164.774	1030.524
MAE	2641.669	304.687	2649.795	786.940	786.940
MAPE	9.456	1.079	2.349	3.402	2.826

averaging.

4.5 Comparative Analysis of the Models

This section presents the evaluation of the fitted models based on the empirical prediction intervals (PIs) and forecast error distributions of each model forecasts.

4.5.1 Evaluation of Prediction Imtervals

gives summary statistics of the PIWs for the models M2 to M6 for PINC value of 95% level of confidence. Model M4 has the narrowest standard deviation which indicates that it has narrower PIWs compared to models M2, M3, M5 and M6. All the PIWs distributions are skewed to the left. The values for kurtosis are all less than 3, showing that the distributions are all platykurtic since for a normally distributed (mesokurtic) data kurtosis value should be equal to three.

Figure 4.9 shows box plots of widths of the PIs for all the

fitted models. From the figure, M4 has the narrowest PI compared to models M2,

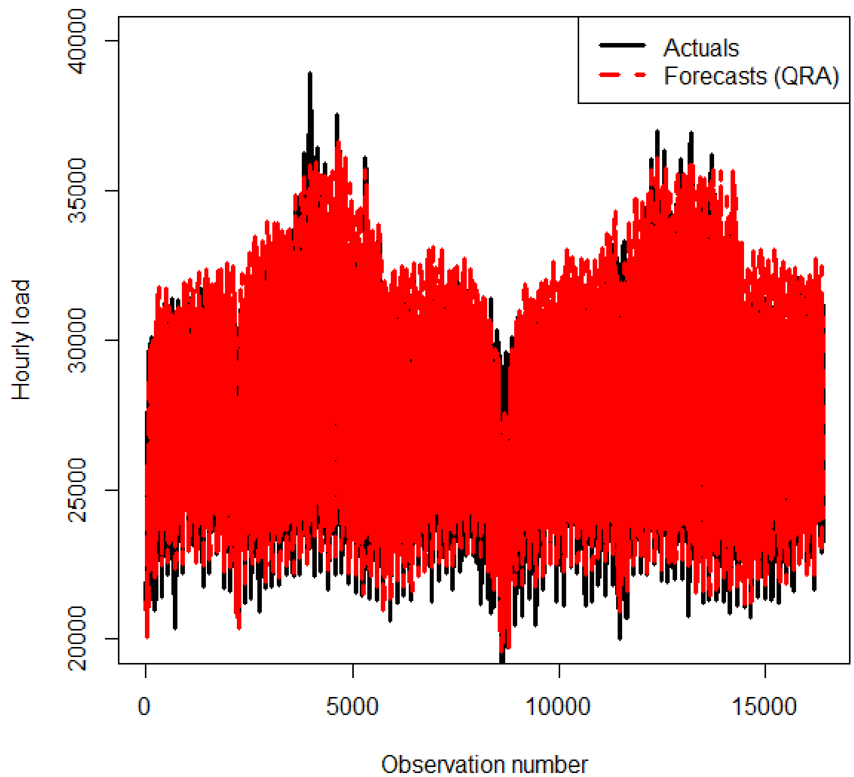

Figure 4.8: Model M6 (QRA) forecasts.

M3, M4, M5 and model M6. Figure 4.10 shows density plots of widths of the PIs for the forecasting models M2 to M6. The density plots are all similar except for the PI from model M4 which has the widest PI.

4.5.2 Residual Analysis

Summary statistics of the residuals of the models are given in Table 4.8. It can be seen that model M4 has the smallest standard deviation which indicates that it has the

Table 4.8: Model PIWs comparisons.

	Mean	Median	Min	Max	St.Dev.	Skewness	Kurtosis
M2	22960	23009	19120	24978	2937.579	-0.241	-0.741
M3	27218	27940	18233	35991	3257.451	-0.257	-1.012
M4	28397	28407	23209	32699	1531.893	-0.171	-0.536
M5	26436	26879	17467	36439	2958.601	-0.187	-0.658
M6	28123	28603	18421	38950	3202.281	-0.187	-0.658

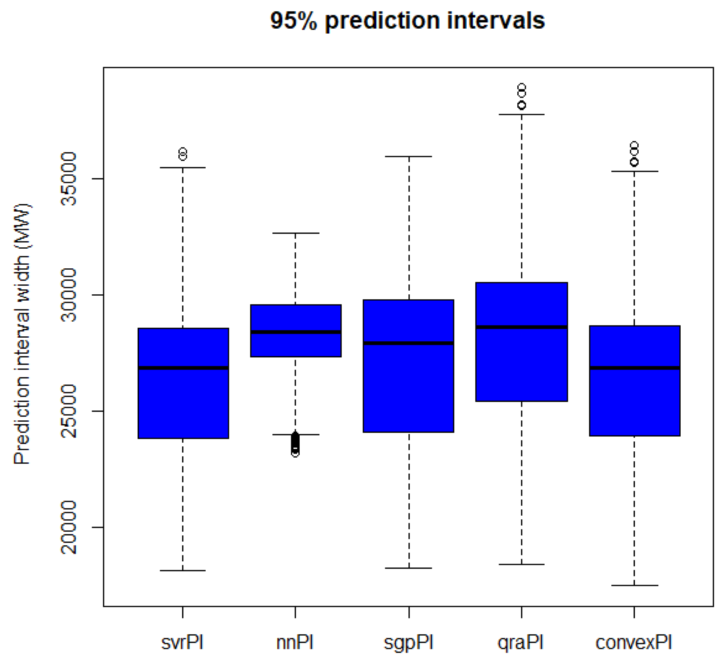

Figure 4.9: Prediction interval widths for models M2 (svrPI), M3 (sgbPI), M4 (nnPI), M5 (convexPI), and M6 (qraPI).

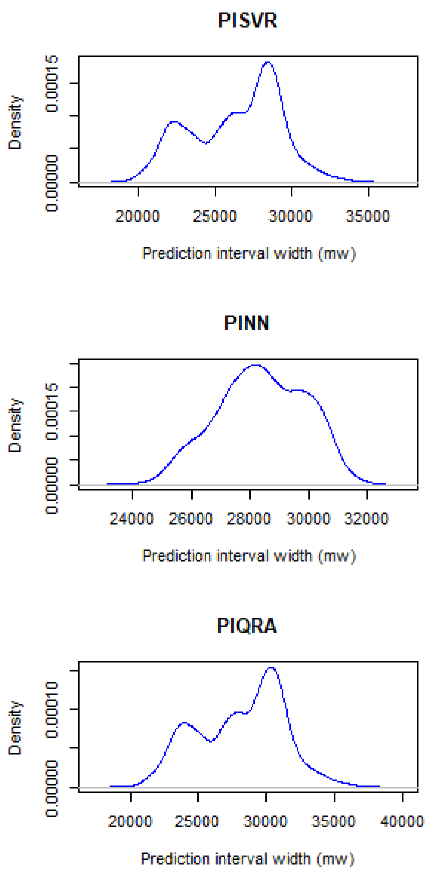

Figure 4.10: Density plots of the Prediction interval widths for models M2 (svrPI), M3 (sgbPI), M4 (nnPI), M5 (convexPI), and M6 (qraPI).

narrowest error distribution compared to models M1,M2 and M4, this implies that M4 is the best compared to other models, followed by model M1. All the error distributions for the models M1, M2 and M4 are skewed to the left since the values of their skewness are all negative. The values for kurtosis are all smaller than 3 for all the models, showing that the distributions are all platykurtic.

Table 4.9: Models residuals comparisons.

	Mean	**Median**	**Min**	**Max**	**St.Dev.**	**Skewness**	**Kurtosis**
M2	28123	28671	19455	38483	3090.884	-0.241	-0.743
M3	28114	28833	19172	36846	3242.529	-0.257	-1.011
M4	28116	28351	22684	32433	1697.357	-0.240	-0.950
M5	2109	2127	1702	2433	127.317	-0.240	-0.950
M6	-8.542	0	-5138	4568	1030.52	-0.126	1.134

Figure 4.11 shows box plots of the forecast errors for all the fitted models. From the figure, M2 has the narrowest error distribution compared to models M1, M2 and M4, implying that M2 is the best model compared to other models. Figure 4.11 shows density plots of the forecast errors for the forecasting models M2 to M6. The density plots are similar for the forecast errors from models M2 and M6, the rest of the boxplots are all different.

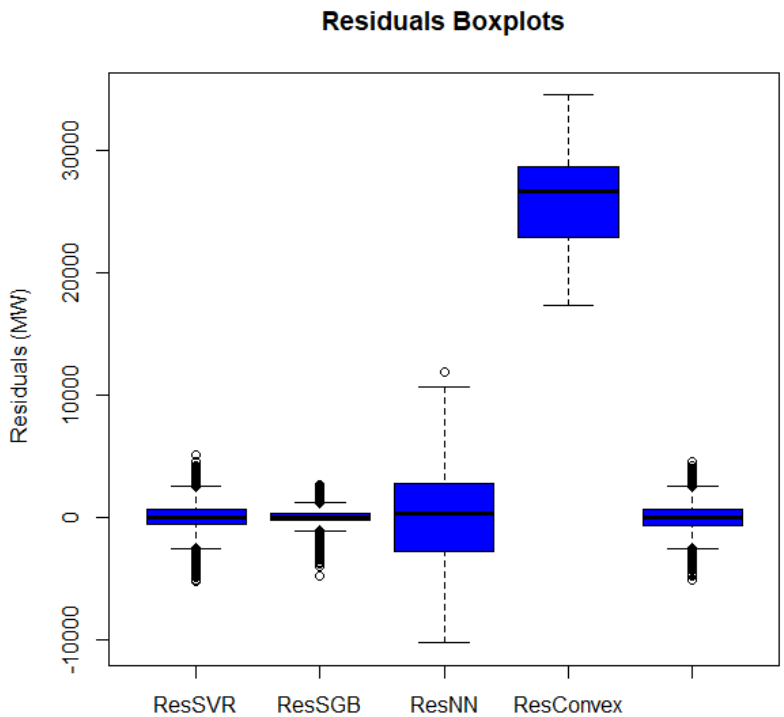

Figure 4.11: Box plots of residuals from models M2 (ResSVR), M3 (ResSGB), M4 (ResNN), M5 (ResConvex), and M6 (qraPI).

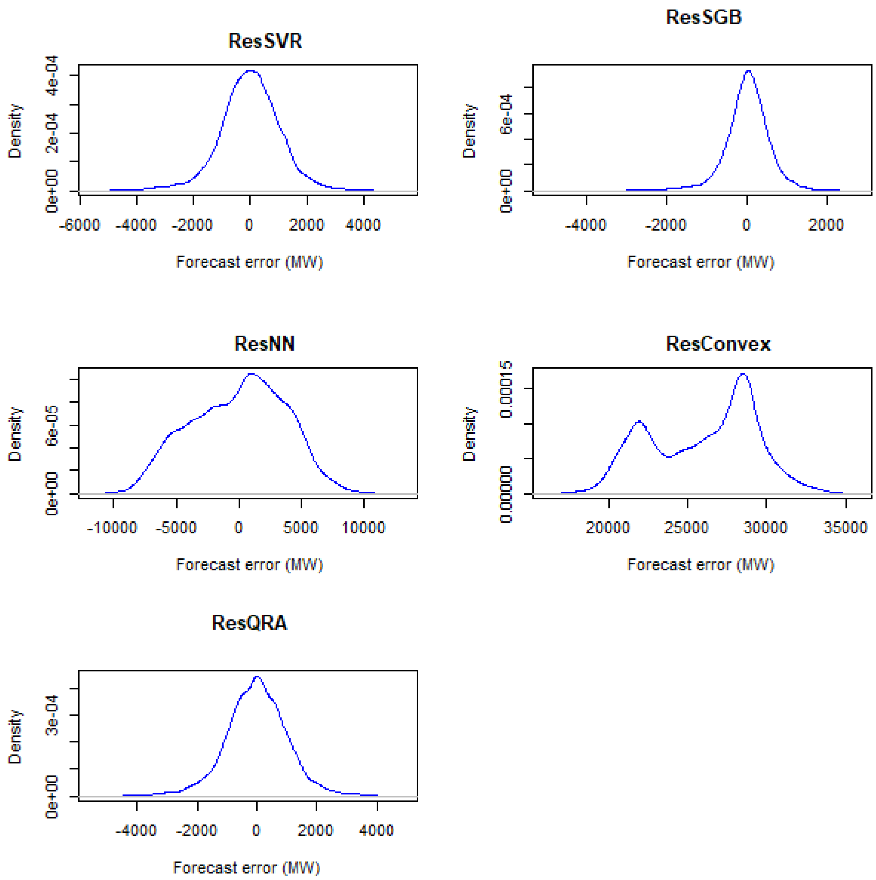

Figure 4.12: Density plots of residuals from models M2 (ResSVR), M3 (ResSGB), M4 (ResNN), M5 (ResConvex), and M6 (qraPI).

4.6 Chapter Summary

The chapter served a purpose of reporting on different analysis and the discussion of the results. Exploratory data analysis and variable selection using lasso under was discussed. The benchmark model, GAM, was presented and then the machne learning models, SVR, SGB, NN were presented and using tables and plots. Forecast accuracy measures in which point forecasts, interval and its combined forecast accuracies were presented giving a further discussion of each of the findings. Finally, the residual analysis on the models was presented. The summary of the findings will be presented in the next chapter.

Chapter 5

Conclusion

5.1 Introduction

This chapter summarises the research findings that were discussed in Chapter 5 and presents recommendations, limitations of the study and suggests areas for future research.

5.2 Research Findings

The economy of every country is heavily dependent on energy. Electricity is used for a number of purposes that include industrial, commercial and household purposes. Thus, forecasting electricity demand is of importance for the purposes of planning by the responsible institution that supplies electricity. This project was based on the forecasting of hourly electricity demand in South Africa for the period 2010 to 2011. The data was obtained from Eskom.

Modelling hourly electricity demand using support vector regression (SVR), stochastic gradient boosting (SGB) and feed forward neural networks (FFNN) was discussed in Chapter 4. The least absolute shrinkage and selection operator (Lasso) was used for variable selection.

The findings in Chapter 4, Section 4.3 showed that the SGB model produced the best forecast accuracy based on the accuracy measured MAE, MAPE and RMSE, among the machine learning models.

Later, the forecasts from the machine learning models were combined using the convex combination method and quantile regression averaging (QRA). Based on MAE, MAPE and RMSE, the QRA model was found to be the best forecast combination method, and also the best forecasting model compared with the machine learning models.

From the prediction interval widths (PIWs) analysis in Section 4.5, at 95% level of confidence, M4 (NN) has the narrowest PI compared to models M2, M3,M5 and M6, implying that M4 is the best compared to other models.

5.3 Recommendations

The main contribution of this dissertation is the inclusion of nonlinear trend variables and the extension of combining forecasting models using QRA. This study could be useful to system operators, including decisionmakers in power utility companies such as Eskom.

5.4 Limitations of the Study

There are many factors that affect the demand of electricity and this study did not include all of them. As the data is averaged, perhaps considering the electricity demand for households and various industrial sectors could have been considered.

5.5 Future Research

Future work will focus on modelling electricity demand for each of the nine provinces of South Africa separately including analysis by sector (i.e. residential, commercial, agricultural and industrial sectors). More forecasting techniques will also be used.

Appendix A

Some selected R codes

```
# The following packages in R are used :
# forecast , ggplot2 , qgam , mgcv , tseries , e1071 , glmnet , hierNet .
# ######################################################################
library ( forecast )
library ( ggplot2 )
library ( tseries )
library ( e1071 )
 library ( glmnet )
library ( hierNet )
# #####################################################################
# time series , qqnorn , density and box plot for electricity demand
# #####################################################################
attach ( analyticDATA_1_ )
head ( analyticDATA_1_ )
win.graph()
```

```
par ( mfrow =c(2 ,2))
A <- ts(load )
plot (A, xlab =" Observation number ",ylab =" Hourly electricity demand (MW) "
      ,main ="(a) Plot of HED ",col = " blue ")
plot ( density (A),xlab =" Hourlr electricity demand (MW)
       ",main ="(b) Density plot ",col = " blue ")
qqnorm (A, col = " blue ",main ="(c) Normal QQ plot ")
qqline (A)
boxplot (A, main ="(d) Box plot ",varwidth =TRUE ,
         xlab = " HED (MW) ", col = " blue ",horizontal = TRUE )
# ###################################################################
## calculating summary statistics , skewness and kurtosis of load
# ####################################################################
summary (A)
library ( e1071 )
sd(A)
skewness (A)
kurtosis (A)
# ##################################################
## Fitting and extracting Non - linear Trend values
# ####################################################
win.graph()
e = ts(load)
plot(e)
z = (smooth.spline(time(load), load))
z
```

```
lines(smooth.spline(time(load), load, spar=0.014408),col="red",lwd=3,)
dpdfits = fitted((smooth.spline(time(load), load, spar=0.014408)))
plot(dpdfits)
write.table(dpdfits,"~/nolfits.txt",sep="\t")

z <- ts( load )
win.graph ()
plot (z, xlab =" Observation number ",ylim =c (20000,35000) ,type ="l",
      ylab =" HED (MW) ")
length ( load )
r= smooth.spline ( time (z), z)
r # 0.1112481
lines ( smooth.spline ( time (z), z, spar = 0.1112481) , lwd =3, col="red ")
dpdfits = fitted (( smooth.spline ( time (z), z, spar = 0.1112481) ))
write.table ( dpdfits ,"~/ loadfittedspline . txt ",sep ="\t")

########################################
##GAM model###

library(mgcv)
#library(qgam)
#library(caret)

fit1 <-gam(load~s(hour, bs="cc", k=24)+s(month, bs="ps",k=12)+s(daytype, bs="ps",k=7)
           +s(holiday,bs="ps")+s(minT,bs="cc")+s(maxT,bs="cc")+s(aveT,bs="cc"),
           family=gaussian,data = data_train)# insample
```

```
summary(fit1)

par(mfrow = c(2,2))
gam.check(fit1)

fit1.forecast <- predict(fit1, newdata = data_test)

fit1.forecast <- round(fit1.forecast,0)
#write.table(fit1.forecast,"~/fSGB.txt",sep="\t")
f <- ts(fit1.forecast)
x <- ts(data_test$load)

library(forecast)
accuracy(f,x)

win.graph()
plot(x,xlab="Observation number",lwd=3,ylab="Hourly load")
lines(f,col="red", lty=2,lwd=3)
legend("topright",col=c("black","red"), lty=1:2,lwd=3,
       legend=c("Actuals", "Forecasts (GAM)"))

# ######################################
## FORECAST COMBINATION OPERA
# #####################################
```

```
library (forecast)
attach ( GAMSVRforecasts_1_ ) # Forecasts from all ML models
head ( GAMSVRforecasts_1_ )
win.graph()
accuracy (GAMf , load )
accuracy (SVRf , load )
#accuracy (Fffnn , load )
# ###############################################
Y <-load
X <- cbind (GAMf , SVRf )
matplot ( cbind (Y,X), type ="l", col =1:4)
fGAM <- GAMf
fSVR <- SVRf
#fFFNN <- Fffnn
X <- cbind (fGAM , fSVR )
# How good are the expert ? Look at the oracles

library ( opera )
oracle.convex <- oracle(Y = Y, experts = X, loss.type = "percentage",model ="convex")
oracle.convex
summary(oracle.convex)
plot( oracle.convex )
print ( oracle.convex ) # print is same as plot in results


# Accuracy measures
```

```
accuracy (mix1 , load )
accuracy (mix2 , load )
accuracy (mix3 , load )

mix1
write.table (mix1 ,"~/ Fconvex . txt ",sep="\t")
# ################################################
## QUANTILE REGRESSION AVERAGING
# ################################################
attach ( GAMSVRforecasts_1_ )
head ( GAMSVRforecasts_1_ )
win.graph ()
y <- ts( load )
plot (y, xlab =" Observation number ", ylab =" Hourly irradiance ")

library ( quantreg )
qr.load = rq( load ~ GAMf + SVRf , data = GAMSVRforecasts_1_ , tau =0.5) # tau = 0.025

summary.rq(qr.load ,se="boot") # can use se = " nid" or se =" ker"
lines (qr.load $fit , col="red")
fQRA = fitted (qr.load )
write.table (fQRA,"~/ QRA05 . txt ",sep ="\t") #LL0025 , QRA05 , UL0975
accuracy (fQRA , load)


# ####################################################################
```

```
### Prediction Interval Width for all models
# ######################################################################

##   Fsvr ,Fsgb, Fnn , Fconvex , Fqra
qr.load = rq( load ~ Fconvex , data = Fconvex , tau =0.05) #LL = 0.05,0.025 ,
0.005
#UP = 0.950 , 0.975 , 0.995
summary.rq(qr.load ,se="boot") # can use se = " nid" or se =" ker"
# lines (qr. load $fit , col =" red ")

fConvexx = fitted (qr.load )
write.table (fConvexx ,"~/ PIcon .txt",sep ="\t") # LL0025 , QRA05 , UL0975
accuracy (Fconvex , load)

# #######################################################################
# ### Model PIWs comparisons at 95%
# #######################################################################
attach ( PIs )
head ( PIs )

PIW = c(" PIgam "," PIsvr "," PIqra "," PIconvex ")
win.graph ()
boxplot ( PIgam , PIsvr , PIqra , PIconvex , names = PIW ,
          horizontal = FALSE , main =" 95% prediction intervals ",
          ylab =" Prediction interval width (MW)", col = " blue ")
```

```
win.graph ()

par ( mfrow =c(3 ,2))
plot ( density ( PIgam ),xlab =" Prediction interval width (mw)", col =" blue ",
       main =" PIGAM ")
plot ( density ( PIsvr ),xlab =" Prediction interval width (mw) ", col =" blue ",
       main =" PISVR ")
#plot ( density ( PInn95 ),xlab =" Prediction interval width (w/m ^2)", col =" blue ",
      # main =" PINN ")
plot ( density ( PIconvex ),xlab =" Prediction interval width (mw) ", col=" blue",

       main =" PIConvex ")
plot ( density ( PIqra ),xlab =" Prediction interval width (mw) ", col =" blue ",
       main =" PIQRA ")

## Summary statistics for PIWs

library ( e1071 )
summary ( PIgam )
sd( PIgam )
skewness ( PIgam )
kurtosis ( PIconvex )
PIsvr
# ################################################################
# Residual Error Analysis
# ################################################################
```

```
attach ( GAM )
head ( GAMSVRforecasts_1_ )
ResGAM = GAMSVRforecasts_1_ $load - GAMSVRforecasts_1_ $ GAMf
ResSVR = GAMSVRforecasts_1_ $ load - GAMSVRforecasts_1_ $ Fsvr
ResFFNN = GAMSVRforecasts_1_ load - GAMSVRforecasts_1_ $ Fnn
ResConvex = GAMSVRforecasts_1_ $ load - GAMSVRforecasts_1_ $ Fconvex
ResQRA = GAMSVRforecasts_1_ $ load - GAMSVRforecasts_1_ $ Fqra

## Summary Statistics for forecast errors
##   ResSVR , ResSGB. ResNN , ResConvex , ResQRA
library ( e1071 )
summary ( ResQRA )
sd( ResQRA )
skewness ( ResQRA )
kurtosis ( ResQRA )

# Residual box - pot and density plot

RESID = c(" ResSGB "," ResSVR "," ResFFNN "," ResConvex ", " ResQRA ")
win.graph ()
boxplot ( ResSGB , ResSVR , ResNN , ResConvex , ResQRA , names = RESID , horizontal
          = FALSE , main ="",
          ylab =" Residuals (w/m ^2)", col = " blue ")

win.graph ()
par ( mfrow =c(3 ,2))
```

```
plot ( density ( ResSGB ),xlab =" Forecast error (w/m^2)", col =" blue ", main ="
       ResSGB ")
plot ( density ( ResSVR ),xlab =" Forecast error (w/m ^2) ", col =" blue ", main ="
ResSVR")
plot ( density ( ResNN ),xlab =" Forecast error (w/m^2)", col =" blue ", main ="
       ResNN ")
plot ( density ( ResConvex ),xlab =" Forecast error (w/m ^2) ", col=" blue ", main ="
 ResConvex ")
plot ( density ( ResQRA ),xlab =" Forecast error (w/m ^2) ", col =" blue ", main ="
ResQRA")
```

References

Apadula, F., Bassini, A., Elli, A. and Scapin, S. (2012), 'Relationships between meteorological variables and monthly electricity demand', *Applied Energy* **98**, 346–356.

Baliyan, A., Gaurav, K. and Mishra, S. K. (2015), 'A review of short term load forecasting using artificial neural network models', *Procedia Computer Science* **48**, 121–125.

Bien, J., Taylor, J. and Tibshirani, R. (2013), 'A lasso for hierarchical interactions', *Annals of statistics* **41**(3), 1111.

Chikobvu, D. and Sigauke, C. (2013), 'Modelling influence of temperature on daily peak electricity demand in south africa', *Journal of Energy in Southern Africa* **24**(4), 63–70.

Devaine, M., Gaillard, P., Goude, Y. and Stoltz, G. (2013), 'Forecasting electricity consumption by aggregating specialized experts', *Machine Learning* **90**(2), 231–260.

Din, G. M. U. and Marnerides, A. K. (2017), Short term power load forecasting using deep neural networks, *in* '2017 International Conference on Computing, Networking and Communications (ICNC)', pp. 594–598.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J. and Vapnik, V. (1997), Support vector regression machines, *in* 'Advances in neural information processing systems', pp. 155–161.

Fahad, M. U. and Arbab, N. (2014), 'Factor affecting short term load forecasting', *Journal of Clean Energy Technologies* **2**(4), 305–309.

Feinberg, E. A. and Genethliou, D. (2005), Load forecasting, *in* 'Applied mathematics for restructured electric power systems', Springer, pp. 269–285.

Friedman, J. H. (2002), 'Stochastic gradient boosting', *Computational statistics & data analysis* **38**(4), 367–378.

Gneiting, T. and Katzfuss, M. (2014), 'Probabilistic forecasting', *Annual Review of Statistics and Its Application* **1**, 125–151.

Gross, G. and Galiana, F. D. (1987), 'Short-term load forecasting', *Proceedings of the IEEE* **75**(12), 1558–1573.

Gupta, M. and Kaur, M. G. (2012), Weather Sensitive Short Term Load Forecasting using Non-fully connected Feed Forward Neural Network, PhD thesis.

Hasties, T. and Tibshirani, R. (1990), 'Generalized additive models', *London: Chapm ana n dHal1* pp. 137–173.

Hinman, J. and Hickey, E. (2009), 'Modeling and forecasting short-term electricity load using regression analysis', *Journal of IInstitute for Regulatory Policy Studies* .

Ho, K.-L., Hsu, Y.-Y. and Yang, C.-C. (1992), 'Short term load forecasting using a multilayer neural network with an adaptive learning algorithm', *IEEE Transactions on Power Systems* **7**(1), 141–149.

Hong, W.-C. (2009), 'Electric load forecasting by support vector model', *Applied Mathematical Modelling* **33**(5), 2444–2454.

Hyndman, R. J. and Fan, S. (2010), 'Density forecasting for long-term peak electricity demand', *IEEE Transactions on Power Systems* **25**(2), 1142–1153.

Inglesi, R. and Pouris, A. (2010), 'Forecasting electricity demand in south africa: A critique of eskom projections', *South African Journal of Science* **106**(1-2), 50–53.

Jones, K. and Wrigley, N. (1995), 'Generalized additive models, graphical diagnostics, and logistic regression', *Geographical Analysis* **27**(1), 1–18.

Lee, K., Cha, Y. and Park, J. (1992*a*), 'Short-term load forecasting using an artificial neural network', *IEEE Transactions on Power Systems* **7**(1), 124–132.

Lee, K. Y., Cha, Y. T. and Park, J. H. (1992*b*), 'Short-term load forecasting using an artificial neural network', *IEEE Transactions on Power Systems* **7**(1), 124–132.

Lehohla, P. (2005), 'Statistics south africa 2005', *Statistics* **39**(3).

Lira, F., Muñoz, C., Nunez, F. and Cipriano, A. (2009), 'Short-term forecasting of electricity prices in the colombian electricity market', *IET generation, transmission & distribution* **3**(11), 980–986.

Liu, G. (2008), 'A generalized gradient smoothing technique and the smoothed bilinear form for galerkin formulation of a wide class of computational methods', *International Journal of Computational Methods* **5**(02), 199–236.

Maciejowska, K., Nowotarski, J. and Weron, R. (2016), 'Probabilistic forecasting of electricity spot prices using factor quantile regression averaging', *International Journal of Forecasting* **32**(3), 957–965.

Mayrink, V. and Hippert, H. S. (2016), A hybrid method using exponential smoothing and gradient boosting for electrical short-term load forecasting, *in* '2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)', pp. 1–6.

McCulloch, W. S. and Pitts, W. (1943), 'A logical calculus of the ideas immanent in nervous activity', *The bulletin of mathematical biophysics* **5**(4), 115–133.

Milidiu, R. L., Machado, R. J. and Renter journal=Neurocomputing, volume=28, n. p. y. p. (n.d.), 'Time-series forecasting through wavelets transformation and a mixture of expert models'.

Mohandes, M. (2002), 'Support vector machines for short-term electrical load forecasting', *International Journal of Energy Research* **26**(4), 335–345.

Mokilane, P., Galpin, J., Sarma Yadavalli, V., Debba, P., Koen, R. and Sibiya, S. (2018), 'Density forecasting for long-term electricity demand in south africa using quantile regression', *South African Journal of Economic and Management Sciences* **21**(1), 1–14.

Nagy, G. I., Barta, G., Kazi, S., Borbély, G. and Simon, G. (2016), 'Gefcom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach', *International Journal of Forecasting* **32**(3), 1087–1093.

Nedellec, R., Cugliari, J. and Goude, Y. (2014), 'Gefcom2012: Electric load forecasting and backcasting with semi-parametric models', *International Journal of forecasting* **30**(2), 375–381.

Padmakumari, K., Mohandas, K. and Thiruvengadam, S. (1999), 'Long term distribution demand forecasting using neuro fuzzy computations', *International Journal of Electrical Power & Energy Systems* **21**(5), 315–322.

Pao, H.-T. (2006), 'Comparing linear and nonlinear forecasts for taiwan electricity consumption', *Energy* **31**(12), 2129–2141.

Papadopoulos, S. and Karakatsanis, I. (2015), Short-term electricity load forecasting using time series and ensemble learning methods, *in* '2015 IEEE Power and Energy Conference at Illinois (PECI)', IEEE, pp. 1–6.

Pierrot, A. and Goude, Y. (2011), 'Short-term electricity load forecasting with generalized additive models', *Proceedings of ISAP power* **2011**.

Shmilovici, A. (2009), Support vector machines, *in* 'Data mining and knowledge discovery handbook', Springer, pp. 231–247.

Sigauke, C. (2014), Modelling electricity demand in South Africa, PhD thesis, University of the Free State.

Stats, S. (2011), 'Statistics south africa', *Formal census* .

Tay, A. S. and Wallis, K. F. (2000), 'Density forecasting: a survey', *Journal of forecasting* **19**(4), 235–254.

Touzani, S., Granderson, J. and Fernandes, S. (2018), 'Gradient boosting machine for modeling the energy consumption of commercial buildings', *Energy and Buildings* **158**, 1533–1543.

Vapnik, V. (1998), *Statistical learning theory. 1998*, Wiley, New York.

Weron, R., Misiorek, A. et al. (2004), Modeling and forecasting electricity loads: A comparison, *in* 'International Conference EEM-04 proceedings', Citeseer, pp. 135–142.

Yuancheng, L., Tingjian, F. and Erkeng, Y. (2002), Short-term electrical load forecasting using least squares support vector machines, *in* 'Proceedings. International Conference on Power System Technology', Vol. 1, IEEE, pp. 230–233.

Zendehboudi, A., Baseer, M. and Saidur, R. (2018), 'Application of support vector machine models for forecasting solar and wind energy resources: A review', *Journal of cleaner production* **199**, 272–285.