Modelling short term probabilistic electricity demand in South Africa

Masters Dissertation

Author:
Mr. Molete Mokhele

Supervisor:
Dr. Caston Sigauke

Dissertation submitted for Masters of Science degree in Mathematical Statistics

in the

Faculty of Science,
School of Statistics and Actuarial Science,
University of the Witwatersrand
Johannesburg

May 2016

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily attributed to the NRF.

Declaration

By submitting this dissertation electronically for the qualification of Masters degree in Mathematical Statistics at the University of the Witwatersrand, I declare that the entirety of the work contained therein is my own, original work, that I am the authorship owner thereof (unless to the extent explicitly otherwise stated) and I have not previously in its entirety or in part submitted it for obtaining any qualification at another university. I further declare that all sources cited or quoted are indicated and acknowledged by means of a comprehensive list of references. I further cede copyright of the dissertation to the University of the Witwatersrand.

Signature:..

Date:..

Abstract

Electricity demand in South Africa exhibit some randomness and has some important implications on scheduling of generating capacity and maintenance plans. This work focuses on the development of a short term probabilistic forecasting model for the 19:00 hours daily demand. The model incorporates deterministic influences such as; temperature effects, maximum electricity demand, dummy variables which include the holiday effects, weekly and monthly seasonal effects. A benchmark model is developed and an out-of-sample comparison between the two models is undertaken. The study further assesses the residual demand analysis for risk uncertainty. This information is important to system operators and utility companies to determine the number of critical peak days as well as scheduling load flow analysis and dispatching of electricity in South Africa.

Keywords: Semi-parametric additive model, generalized Pareto distribution, extreme value mixture modelling, non stationary time series, electricity demand.

Dedication

To my father, Mr. Mac-Millan Clarke Konote

Acknowledgements

Its with great joy to express my sincere regards to every Tom, Dick and Harry who has contributed towards this dissertation:

- Heavenly Father, for His gift of life.
- All my colleagues for their different views, suggestions and assistance's. I cannot afford to forget to thank a friend from University of the Free State Mr. Gaonyalelwe Maribe for helping me with the simulation study in this dissertation.
- Dr Caston Sigauke for his professional work ethic. I really want to thank him for introducing me to the field of modelling and probabilistic forecasting, discrete time Markov chains, Bayesian statistics and extreme value mixture models.
- My parents and siblings for all the support they always show and without them the idea of Masters could have been nothing but only a dream. Most importantly, I am thankful to my father because of the spirit he has always instilled in me. He would always tell me, "*a man cannot discover new oceans unless he has courage to lose sight of his home shores*". I left my home shores two years ago to pursue my postgraduate studies, today, all the time I have been absent from home has finally paid off in spades. I cannot thank my father enough for his encouragement and for all that, I am desperately proud of him.

Contents

Declaration i

Abstract ii

Dedication iii

Acknowledgements iv

Contents v

List of Figures ix

List of Tables xi

Abbreviations xii

Symbols xiv

1 Introduction 1
- 1.1 Background 4
- 1.2 Dissertation Aims and Objectives 6
- 1.3 Scope of the Dissertation 7
- 1.4 Structure of the Dissertation 9

2 Literature Review 10
- 2.1 Introduction 10
- 2.2 Semi-Parametric Additive Models 11
- 2.3 Modelling Framework: Bayesian Analysis 15
- 2.4 Modelling Daily Peak Electricity Demand using Extreme Value Theory (EVT) 18

3 Methodology 21
- 3.1 Semi-Parametric Additive Model 21
  - 3.1.1 Setting up the Model 21
  - 3.1.2 Model Fitting 23
    - 3.1.2.1 Maximum Likelihood Estimation 23
    - 3.1.2.2 Bayesian parameter Estimation 23
  - 3.1.3 Evaluation Methods for Model Sufficiency and Prediction Accuracy 23
    - 3.1.3.1 The evaluation Methods 23

3.1.3.2 Measurement of prediction performance . . . 24
3.1.4 Model Selection . . . 24
3.1.5 Model Diagnostics . . . 25
3.1.5.1 Checking model assumptions . . . 25
3.2 Daily Peak Electricity Demand Modelling using Extreme Value Theory (EVT) . . . 25
3.3 Generalized Pareto Distribution (GPD) . . . 25
3.4 Non Stationary and Stationary Extremes . . . 26
3.5 Threshold Selection . . . 27
3.5.1 Mean Residual Plot . . . 28
3.5.2 Parameter Stability Plot . . . 28
3.5.3 Pareto Quantile Plot . . . 29
3.5.4 Extreme Value Mixture Models . . . 29
3.6 Parameter Estimation . . . 30
3.6.1 Maximum Likelihood Estimation . . . 30
3.7 Daily Peak Electricity Demand Forecasting using Bayesian Approach . . . 30

4 Modelling Daily Peak Electricity Demand in Different Hours of the day Using Discrete Time Markov Chain 33
4.1 Introduction . . . 33
4.2 Finite Discrete Time Markov Chains (DTMC) . . . 34
4.2.1 Description of the Data . . . 35
4.2.2 Markov Property . . . 37
4.2.3 DPED Changes Transition Probabilities Estimation . . . 38
4.2.4 The n-step Transition Probabilities . . . 38
4.2.5 Mean Return times (MRT) . . . 39
4.2.6 First Passage times (FPT) . . . 40
4.2.7 Two State Problem . . . 40
4.2.7.1 Transition matrix and mean return times . . . 40
4.2.7.2 First Passage Times (FPT) . . . 41
4.2.8 Two-State Problem Extreme Peaks Modelling . . . 43
4.2.9 Three-State Problem . . . 47
4.2.10 Electricity Demand Data . . . 50
4.3 Concluding Remarks . . . 51

5 Short Term Load Forecasting 53
5.1 Introduction . . . 53
5.2 Models . . . 54
5.2.1 The Load Forecasting Models . . . 54
5.2.2 Semi-Parametric Additive Model . . . 54
5.2.3 Model Parameters Estimation . . . 58
5.2.4 Final Model . . . 59
5.2.5 Forecasting results . . . 62
5.3 Benchmarking Model . . . 66
5.3.1 Benchmark Model . . . 67
5.3.2 Empirical Results on the Benchmark Model . . . 70
5.4 Comparative analysis of the models . . . 71

5.5 Concluding Remarks 71

6 Residual Demand Analysis **72**
6.1 Introduction 72
6.2 Generalized Extreme Value Distribution (GEVD) 73
6.2.1 Gumbel Class of Distributions 74
6.2.2 Frećhet Class of Distributions 74
6.2.3 Weibull Class of Distributions 74
6.2.4 Model Choice and Parameter Estimation 76
6.2.5 Parameter Estimation 76
6.2.5.1 Model 1 77
6.2.5.2 Model 2 77
6.2.6 Diagnostic Plots for GEV Models (Model 1 and Model 2) 79
6.2.6.1 Model 3 80
6.2.6.2 Model 4 81
6.2.7 Diagnostic Plots for GEVD Models (Models 3 and 4) 82
6.2.8 Return Level and Return Period Estimation 84
6.3 Generalized Pareto Distribution (GPD) 85
6.4 Threshold Selection 87
6.4.1 Introduction 87
6.4.2 Pareto Quantile and Mean Excess Plots 87
6.4.3 Extreme Value Mixture Models 89
6.5 Modelling Extreme Stationary Series 95
6.6 Modelling Extreme Non-Stationary Series 97
6.6.1 Return Levels and Confidence Limits 99
6.7 Bayesian Estimation and Simulation Study 100
6.7.1 Prior Elicitation 100
6.7.1.1 Jeffreys' Prior 101
6.7.2 Markov Chain Monte Carlo (MCMC) 101
6.7.2.1 Metropolis Algorithm 101
6.8 Simulation Study 106
6.9 Power Outage Risk Assessment and Management 110
6.9.1 Power Outage Risk Assessment 110
6.9.1.1 Outage risk (r) 111
6.10 Concluding Remarks 111

7 Conclusion **113**
7.1 Introduction 113
7.2 Summary and Concluding Remarks 113
7.3 Limitations of the Dissertation 116
7.4 Future Research 116

References **117**

Appendices **125**

A Discrete time Markov chain analysis R code **126**

B Extreme mixture models R code **133**

C Simulation study R code **135**

List of Figures

4.1 Daily peak electricity demand time series plot. The observation number is located in the horizontal axis ($x-axis$) and daily peak electricity demand along the vertical axis ($y-axis$) . . . 36
4.2 (a) Seasonally differenced daily peak electricity demand plot (top left panel), (b) DPED changes data density plot (top right panel), (c) DPED changes data QQ-plot (bottom left panel), (d) DPED changes data box-plot (bottom right panel) . . . 37
4.3 A two-state problem transition diagram given that the present system state is 1 (increase) . . . 42
4.4 Daily peak electricity demand changes Pareto quantile plot . . . 44
4.5 Plot of a transition diagram for a two-state Problem when the threshold is 2223 MW . . . 45
4.6 Bar chart of the daily electricity peak demand occurrences for the period 01 Jan 2000 to 10 Sept 2011 . . . 50
4.7 The daily electricity load consumption for the period 01 Jan 2000 to 10 Sept 2011 at 9:00 a.m. . . . 51
4.8 The daily electricity load consumption for the period 01 Jan 2000 to 10 Sept 2011, at 19:00 p.m. . . . 51

5.1 **Top panel:** Actual and predicted electricity demand in megawatts at hour 19:00 for February to August 2010. **Bottom panel:** (a) Kernel density of the actual electricity demand and (b) Kernel density of the predicted electricity demand at hour 19:00 for February to August 2010. . 62
5.2 The 95% prediction interval of the predicted electricity demand at 19:00 . 63
5.3 **Top Panel** (a) Probability density function for 1 February 2010, **Bottom panel** (b) Probability density function for 5 February 2010. . . . 64
5.4 Probability density functions of DPED for first seven days of February 2010 65
5.5 **Top panel:** (a) Average Daily Temperature series (2000-2010). (b) Daily Temperature series at 19:00 hours (2000-2010). **Bottom panel:** (c) Daily Peak Electricity Demand (2000-2010). (d) Daily electricity demand at 19:00 hours series (2000-2010). . . . 68
5.6 **Top panel:** (a) Daily peak electricity demand at 19:00 hours and peak electricity demand at 19:00 hours scatter plot. (b) Daily electricity demand at 19:00 hours versus the average daily temperature scatter plot. **Bottom panel:** (c) Daily electricity demand at 19:00 hours versus the average minimal temperature scatter plot. (d) Daily electricity demand at 19:00 hours versus the average maximal temperature scatter plot. . . . 69

6.1 (a) Positive and negative residuals scatter plot, (b) Positive and negative residuals box plot, (c) Positive and negative residual density plot and (d) qq-norm plot depicting the positive and negative residuals 75
6.2 Model diagnostics for Model 1 demonstrating both the positive and negative residuals applied to the GEVD, (a) Top left panel: Probability plot, (b) Top right panel: Quantile plot, (c) Bottom left panel: Return level plot and (d) Bottom right panel: Density plot 79
6.3 GEVD diagnostic plots for goodness of fit, allowing quadratic trend in μ. 80
6.4 GEVD diagnostic plots for model 3, allowing some linear trend in μ and σ 82
6.5 GEVD diagnostic plots for model 4, allowing some linear trend in μ . . . 83
6.6 (a) Positive residuals plot (b) Negative residuals plot 86
6.7 Positive residuals Pareto quantile plot 88
6.8 Positive residuals mean excess plot 89
6.9 Negative residuals mean excess plot 89
6.10 Density function of the extreme value mixture model (parametric) with bulk model based tail fraction approach with parameter vector and parameterised tail fraction approach with parameter, where the vertical dashed lines indicate the thresholds for the two different approaches. . . . 91
6.11 Bulk model based tail fraction approach diagnostics plots, (a) Top left panel: Return level plot, (b) Top right panel: Quantile plot, (c) Bottom left panel: Probability plot and (d) Bottom right panel: Density plot . . . 92
6.12 Parameterised tail fraction approach diagnostics plots, (a) Top left panel: Return level plot, (b) Top right panel: Quantile plot, (c) Bottom left panel: Probability plot and (d) Bottom right panel: Density plot 93
6.13 (a) Time series plot of residuals showing the upper and lower thresholds estimated using extreme value mixture model approach (Parameterised tail fraction approach), (b) Time series plot of residuals showing the upper and lower thresholds estimated using extreme value mixture model approach (bulk model based tail fraction approach) 94
6.14 Frequency analysis of the occurrences of (a) over estimation and (b) under estimation residuals demand above the thresholds. 94
6.15 Diagnostic plots used to assess GPD (only for positive residuals) goodness–of–fit, (a) **Top left panel**: Return level plot, (b) **Top right panel**: Quantile plot, (c) **Bottom left panel**: Probability plot and (d) **Bottom right panel**: Density plot 97
6.16 (a) Left panel: Residuals probability plot and (b) Right panel: Residuals Quantile plot 99
6.17 Generalized Extreme Value parameters MCMC realizations of the residuals data. 102
6.18 Generalized extreme value parameters MCMC realizations using different starting values than those used to produce the realizations of Figure 6.17 103
6.19 **Left Panels** (a) Trace plots of the three parameters and their marginal posterior densities, **Right panels** (b) Marginal posterior densities for the generalized extreme value parameters μ, σ and ξ respectively. 104
6.20 The Posterior distribution return level plot demonstrating the 95% Bayesian credible intervals in the form of the dashed lines 105

List of Tables

4.1 Probabilities of the first passage times (FPT) for the upcoming seven days, assuming that an increase (state 1) is the system current state . . . 41
4.2 First Passage time probabilities, assuming the system current state is state 2 (decrease) for the next seven days . . . 42
4.3 First passage time probabilities assuming the current system state is state 1 ($z_t > 2223$ MW) for next seven days . . . 46
4.4 Probabilities of the first passage time given that state 2 is the present state of the system for an upcoming week ($z_t \leq 2223$ MW) . . . 46
4.5 First passage time probabilities, assuming the current system state is "an extreme increase" ($z_t \geq 2223$) for the next seven days . . . 48
4.6 First passage time probabilities assuming the current system state is a "small increase" ($0 < z_t < 2223$) for the next seven days . . . 49
4.7 First passage time steady state probabilities assuming the current system state is a "decrease" ($z_t \leq 0$) . . . 49
4.8 Frequency occurrences of DPED in different hours of the Day . . . 50

5.1 Selection Model Summary . . . 59
5.2 Proposed final model parameter estimates . . . 61
5.3 Benchmark model parameter estimates . . . 70
5.4 Out-of-sample evaluation of the models (validation) . . . 71

6.1 Parameter estimates for stationary GEVD Model (Model 1) . . . 77
6.2 95% Confidence intervals of Model 1 parameter estimates . . . 77
6.3 95% confidence intervals, return levels for several return periods for stationary GEVD. . . . 84
6.4 Simulation results on the three methods . . . 106
6.5 For 200 samples, $\alpha = 0.1$, $n = 1000$, $k = 5\% * 1000, 15\% * 1000$ and $25\% * 1000$, EVI and MSE results of the three estimators are as follows . . . 107
6.6 For 200 samples, $\alpha = 0.1$, $n = 1000$, $k = 5\% * 1000, 15\% * 1000$, $25\% * 1000$ and $50\% * 1000$, EVI and MSE results of the three estimators are as follows . . . 108
6.7 For 200 samples, $\alpha = 0.1$, $n = 1000$, $k = 5\% * 1000, 15\% * 1000$, $25\% * 1000$ and $50\% * 1000$, EVI and MSE results of the three estimators are as follows . . . 108
6.8 Simulation Results Summary . . . 108

Abbreviations

ADD	**A**verage **D**aily **D**emand
AI	**A**rtificial **I**ntelligence
AIC	**A**kaike **I**nformation **C**riterion
ANN	**A**rtificial **N**eural **N**etworks
ARIMA	**A**uto-regressive **I**ntegrated **M**oving **A**verage
ARCH	**A**uto-regressive **C**onditional **H**eteroskedasticity
BIC	**B**ayesian **I**nformation **C**riterion
CDD	**C**ooling **D**egree **D**ays
CI	**C**onfidence **I**nterval
CV	**C**ross **V**alidation
DPED	**D**aily **P**eak **E**lectricity **D**emand
DTMC	**D**iscrete **T**ime **M**arkov **C**hain
EDPED	**Extreme D**aily **Peak E**lectricity **D**emand
EE	**E**nergy **E**fficiency
e.g.	**f**or **e**xample
EIEs	**E**xtensive **I**ntensives
ELD	**E**nthalpy **L**atent **D**ays
Eskom	**S**outh **A**frica's **p**ower **u**tility **c**ompany
EVI	**E**xtreme **V**alue **I**ndex
EVT	**E**xtreme **V**alue **T**heory
FPT	**F**irst **P**assage **T**ime
GARCH	**G**eneralized **A**uto-regressive **C**onditional **H**eteroskedasticity
GDP	**G**ross **D**omestic **P**roduct
GEVD	**G**eneralized **E**xtreme **V**alue **D**istribution
GPD	**G**eneralized **P**areto **D**istribution

GSP	**G**eneralized **S**ingle **P**areto
HDD	**H**eating **D**egree **D**ay
LMOM	**L**-moments **M**ethods
MAE	**M**ean **A**bsolute **E**rror
MAPE	**M**ean **A**bsolute **P**ercentage **E**rror
MCMC	**M**arkov **C**hain **M**onte **C**arlo
MDI	**M**aximal **D**ata **I**nformation
MLE	**M**aximum **L**ikelihood **E**stimation
MSE	**M**ean **S**quare **E**rror
NESO	**N**et **E**nergy **S**ent **O**ut
PLF	**P**robabilistic **L**oad **F**orecasting
PoE	**P**robability **o**f **E**xceedance
POT	**P**eaks **O**ver **T**hreshold
PP	**P**robability **P**robability
QQ	**Q**uantile **Q**uantile
RMSE	**R**oot **M**ean **S**quare **E**rror
RQ	**R**esidual **Q**uantile
SANN	**S**imulated **A**nnealing
SARIMA	**S**easonal **A**uto-regressive **I**ntegrated **M**oving **A**verage
STLF	**S**hort **T**erm **L**oad **F**orecasting
VAR	**Value a**t **R**isk

Symbols

x_p	quantile function
$\pi(\theta)$	prior distribution
ε_t	white noise process
log	natural logarithmic
τ	a sufficiently high threshold
x_t	daily peak electricity demand at t
$W_\xi(x)$	generalized Pareto distribution function
γ	EVI for the generalized extreme value distribution
z_p	extreme quantile (return level)
$z^*_{\alpha/2}$	the $(1-\alpha/2)$ quantile of the standard normal distribution
ξ	Extreme Value Index (EVI) for the generalized Pareto distribution

Chapter 1

Introduction

Since the first democratic government takeover in 1994, the increased electricity demand has led to sustainability issues. In addition to this, the electricity demand continues to rise everyday due to a variety of driving factors. Considering this, if the electricity generating capacity in South Africa does not show potential signs of meeting the country's electricity demand in the upcoming years, this may have a significant impact on the national grid causing it to operate in a risky and vulnerable state, leading to disturbances such as load shedding. More importantly, electricity is an important commodity used mainly as a source of energy in industrial, residential and commercial sectors. In order to have secure economic power system operations, grid companies find estimation of electricity consumption an important task to perform (Zhang et al., 2012). Notably, in various industrialised countries, Kou and Gao (2014) mention "the electrical energy consumption in energy intensive enterprises (EIE) constitutes a significant part of the country's total energy use with higher costs". EIEs in South Africa includes steel, alumina and petrochemical, among others. Kou and Gao (2014) further claim that, in many cases, EIEs have their own generating plants, thus forming a grid. Electricity consumption requires forecasting related to both production and maintenance of operating conditions in the EIE (Zhang et al., 2012).

Notably, probabilistic electricity forecasting is an essential tool that is required to ensure proper planning in power systems, thus facilitating effective operation (Fan and Hyndman, 2012). Probabilistic electricity forecasting is categorised into three time horizons. To be specific they are short, medium and long term electricity forecasting.

Among the three, short term load forecasting (STLF) is a significant technique that is used to ensure sufficient electricity generation capacity (McSharry et al., 2005). But even so, not much has been done to address the generating power capacity challenges faced by South Africa using probabilistic short term forecasting, and it has always been a subject of major interest to the decision makers in the electricity sector. With this in mind, it is important to have accurate estimates and further maintain STLF accuracy as these has a huge significance in the electricity sector (Tong et al., 2014). Consequently, accurate electricity forecasts help system operators to effectively manage risks and costs associated with power system operations (McSharry et al., 2005). Among others who have studied load forecasting include Fan and Hyndman (2012), Hyndman and Fan (2010), Morita et al. (1996) and McSharry et al. (2005). Given this background, literature on Probabilistic Load Forecasting (PLF) when compared to probabilistic short term load forecasting is limited generally (Gneiting and Katzfuss, 2014).

The increasing electricity peak demand in South Africa is driven substantially by the rapidly growing population, extreme weather conditions, growing industrial developments, technological developments and a host of other socio-economic issues such as family size and the electricity price, just to name a few. As a result, the government is continually seeking additional electricity sources, while building new generating plants as a method to address the problem. Moreover, based on critical peak electricity demand, many countries throughout the world have found support for this challenge as an energy policy concern, symbolic to challenges such as power shortages (Strengers, 2012). Therefore, energy policy concern in South Africa suggests efficient usage of electricity is among strategic agendas of the country, and that prompts to guide development of the future economy. In addition to this, probabilistic load forecasting constitute an extremely important part of the national strategic energy policy agenda in South Africa, whose electricity consumption has been fast growing over the past twenty years since democracy in 1994 (Sigauke, 2014).

Furthermore, the transmission and distribution of electricity across South Africa continues to be a challenge and a great concern to system managers. We recognise that the relationship between the peak electricity demand and factors that influence the high demand since the 2008 financial recession has been a matter of debate. It is for these reasons that this dissertation focuses on the development of a STLF model in the semi-parametric additive regression framework, following the work of Fan and Hyndman (2012).

As mentioned earlier, one of the fundamental difficulties faced by power systems is unpredictability and variability (Bathurst et al., 2002), since electricity demand demonstrates some randomness and is subject to a selection of external hosts (Fan and Hyndman, 2012). The proposed topic concerns modelling the short term electricity demand in South Africa using a probabilistic approach similar to that of Fan and Hyndman (2012). The proposed model is a semi-parametric additive model, with the response variable being the hourly electricity demand in South Africa, while the predictor or explanatory variable(s) include variables such as temperature, calendar effects, "Cooling Degree Days (CDD_t)" and "Heating Degree Days (HDD_t)" among others. We estimate the "heating degree days" and the "cooling degree days" of the day t using the equations

$$\text{HDD}_t = \max(T_{ref} - T_t, 0) \tag{1.1}$$

$$\text{CDD}_t = \max(T_t - T_{ref}, 0) \tag{1.2}$$

where T_t is the average temperature for the day t and T_{ref} is a reference temperature that separates the cold and heat periods, exhibited by the demand and temperature relationship (Mirasgedis et al., 2006).

Furthermore, to model both the positive and negative residuals generated from the forecasting model (indicating under and over demand) we make use of extreme value theory (EVT). In order to model the data tail distribution we use EVT (Gencay and Selcuk, 2004). Data from the Electricity Supply Commission (Eskom), South Africa's utility company will be used. As mentioned earlier, in South Africa electricity demand exhibits some randomness and so the data is made stationary. Essentially, most statistical forecasting methods are based on the assumption that the time series data can be rendered approximately stationary. Therefore, we make the data stationary by taking the first differences, ensuring the removal of linear trend and seasonality that is exhibited in the load data. The first difference of the residuals will give an analysis and assessment of the model forecasting certainty. In addition to this, to model the tail behaviour of the residuals above some sufficiently high threshold we use Peaks Over Threshold (POT) distribution. However, in traditional extreme value modelling approaches, various diagnostics plots to evaluate the model fit are commonly used for choosing the threshold. Coles et al. (2001) outline three diagnostics which include general model fit diagnostics, mean residual and parameter stability plots. Further, It is essential to analyze the distribution of both residuals (positive and negative) so that we can assess the efficacy of the probabilistic electricity forecasting technique on the power system from an operational point of view (Tong et al., 2014).

1.1 Background

In this dissertation we study the relationship that possibly exist between the electricity demand and the exogenous variables as proposed by Fan and Hyndman (2012). The main objective is to make some predictions on the probability distribution of the electricity demand in South Africa. It is essential to take notice that South Africa's forever growing electricity demand is driven by growing population, urbanization, growing industrial development, meteorological influences and a host of other factors. Consequently, system operators are continually looking for alternative methods and tools to help in the planning and conservation of electricity by considering probabilistic load forecasting. Accordingly, the purpose of the analyses that will be performed in this research is to identify the most appropriate forecasting model to be applied to the South African hourly demand data, make predictions and evaluate the peak event risk of the model.

The proposed topic studies the origins of load forecasting and how it has been a conventional and an important process in electric utilities since the early 20^{th} Century (Wang et al., 2015). According to Wang et al. (2015) load forecasting techniques are also typically classified into two groups, namely statistical techniques and Artificial Intelligence (AI) techniques. However one should know that these two groups cannot be mutually exclusive. Furthermore, the different probabilistic forecasting versions can be defined as short, medium and long term forecasting.

A typical empirical study that examines probabilistic load forecasting is that of Tong et al. (2014). Additionally, in a study by Gross and Galiana (1987) on probabilistic forecasting, there is an indication of the fact that there is ongoing large number of models proposed in the literature around short term probabilistic forecasting. Notably, a significant number of studies have focused on long term spatial load forecasting, which essentially looks at the occurrence and variation of load growth at different times (Willis and Northcote-Green, 1983).

From a practical viewpoint, STLF models are very important as they provide a platform for making load predictions for at most a week in advance (Hong, 2010). In addition to this, the forecasting models provide frameworks with embedded information, while taking into account the electricity demand and weather parameters such as temperature, rainfall, humidity, and sunshine, random influences that are triggered by the different ways consumers choose to use electricity and a host of other variables. Above all, the importance of the models is emphasized in providing forecast distributions and assessing risk uncertainty. In an economic environment, Singh et al. (2014) highlight that STLF is significant and further helps in quantifying security operation of the power system.

Many STLF models have been proposed and include a mixture of multiple regression models and Artificial Neural Networks (ANN). Abu-El-Magd (1983) review several STLF techniques, such as multiple linear regression, spectral decomposition, exponential smoothing, Box-Jenkins approach, state space models and some multivariate models. Furthermore, In this study we find the usefulness of the multiple regression in electricity modelling. To carry out the multiple regression, we investigate some relationship that exist between various predictor or explanatory variables such as weather related variables and the electric load among others.

Past studies show that accurate STLF has become a challenging issue in the ongoing deregulation of the electricity sector. The accuracy basis of the model highly depends on variables incorporated in the model, however there are methods used to better the precision and accuracy in probabilistic forecasting, such as bootstrapping and Monte Carlo simulation studies (Bracale et al., 2013). Notably, there are models based on bootstrapping, Monte-Carlo, regression and ANN already in place used by Eskom for the improvement of performance related to scheduling and planning in South Africa.

The overall research aim and objective of this dissertation is underpinned by the need to develop a better performing semi-parametric additive model and improve the precision and accuracy on the performance of the underlying model. Probabilistic load forecasting is mainly essential for the electricity industry, and constitutes a vital part of energy policy in South Africa and many other countries in the world (Sigauke, 2014).

The data used in this dissertation is the electricity demand consumption data for the period 2000 to 2010 from Eskom. Essentially, Net Energy Sent Out (NESO) is preferred in this dissertation as an alternative to electricity demand. NESO (in megawatts) can be defined as the rate at which electricity is supplied to its consumers. As already mentioned, we use NESO as an alternative to electricity demand. However, NESO has its shortcomings as it excludes the demand from consumers who are determined (or undetermined) and can (or cannot) afford to cover their electricity bills, but at the present moment do not have electrical homes. Although there are some limitations to NESO, it remains an essential alternative for electricity demand. In particular, the NESO considered in this project is maximal hourly electricity demand data. Aggregated maximal hourly demand data is used, which means it comes from several sectors of the South African economy (which includes residential, commercial and industrial sectors). Residential sector comprises of the households. Industrial sector comprises of manufacturing business establishments and commercial sector comprises of non manufacturing business establishments which include retail stores, health, social and educational institutions and the street and traffic lighting.

Finally, historical hourly data on temperature (in °C) from 36 meteorological stations from all the provinces of South Africa are also used in the development of the short term forecasting model.

1.2 Dissertation Aims and Objectives

The aim of this dissertation is to model short term probabilistic electricity demand in South Africa. These objectives are the main points that the dissertation seeks to address:

1. develop a short term electricity forecasting in the regression framework and compare it with the benchmark model,
2. assess residual demand analysis,
3. model daily peak electricity demand using the Bayesian approach and make comparisons with the frequentist approach,
4. evaluate the hourly peak event risk,
5. suggest areas for further study.

1.3 Scope of the Dissertation

The research will be based on a recent paper by Fan and Hyndman (2012) in which they develop a short term forecasting model in a semi-parametric framework. We develop a semi-parametric additive model similar to the one discussed in Fan and Hyndman (2012) with a few modifications. Firstly, modifications in the relationships between peak electricity demand in different hours of the day, then the driver variables, including calendar effects, temperature (in °C) and the covariates such as "Cooling Degree Days (CDD_t)", "Heating Degree Days (HDD_t)". Furthermore, to estimate the model parameters we use both the Bayesian inference and the Maximum Likelihood Estimation (MLE) methods, after which a comparative analysis will be conducted. In addition to this, the forecasting model is developed to better forecast and capture the whole probability distribution of the hourly demand in South Africa, in contrast to all previous models which focused mainly on the average peak demands. The model will be used to obtain the hourly electricity demand predictions, and also short term electricity forecasts which are important for power system planning (Hyndman and Fan, 2014). However, the model's forecasts ($\hat{y}$) will not accurately match the actual demand value (y), but based on South African historical data, rough ad hoc forecasts can be obtained. The developed model is more prone to some uncertainty and risk, due to large residuals from the forecasted model (over and underestimated demand). The next section provides a detailed discussion on how large residuals are dealt with.

The forecasting methodological framework of the model is summarised in the following stages: model development as part of the first stage, the second stage involves forecasting and a simulation study and the last stage will involve evaluating the general model performance. Furthermore, we forecast the probability distribution in order to examine the prediction interval based on a series of residuals observed. A further subsidiary purpose of the research is to model a series of residuals resulting from the forecast model using Extreme Value Theory (EVT), that is over and under demand predictions. In an effort to model distribution of rare occurring events we use the extreme value theory and we use the Generalized Pareto distribution (GPD) to model the upper tail distributions. We will therefore use the GPD to model residuals which will indicate over and under demand estimations. Furthermore, the over and under estimation estimates can be represented as, $r_t = y_t - \hat{y}_t > 0$ and $r_t = y_t - \hat{y}_t < 0$ respectively, where y_t is the observed demand value, $\hat{y}_t$ is the predicted demand value and r_t is the residual term.

We will further consider two cases: The first case will be for a scenario where we assume stationary GPD models while in the second case non-stationary GPD models are assumed. Moreover, stationary models are made up of the following parameters; a scale parameter σ and a shape parameter ξ which do not change overtime, as they are assumed to be constant. Similarly, a non-stationary model is made up of the scale parameter (σ) and shape parameter (ξ), which vary in time. At the very least, to understand the performance and accuracy of the forecast model, we will consider simulation studies. Monte Carlo simulation will be the first technique to consider and secondly the Bootstrapping technique. These are the techniques used to fully comprehend the performance of the developed probabilistic prediction model.

Monte Carlo Simulation

This is the technique that is used to understand the impact and performance of the forecasting models. Therefore, we develop a short term probabilistic forecasting model under certain assumptions. The model will give short term forecasts, and estimate the expected demand value. Then based on the range of over and underestimated expected values generated by the developed semi-parametric model, one begins to understand the risk and uncertainty in the model. Also, to understand how likely the resulting outcomes of the model, to forecast some unknown future value we use the Monte Carlo simulation.

Bootstrapping

It is a statistical tool used for assessing uncertainty, and which is comprised of different methods. They are mostly important when inference is to be made, based on a complex procedure which produces theoretically unavailable results (Davison and Kuonen, 2002).

Bootstrap based Confidence Intervals

There are several ways in which the bootstrap based confidence intervals of residuals can be computed in the statistical inference spectrum. However, in this research we will consider only one method. The method is based on normal approximation, which can be justified by the results of Oberhofer and Haupt (2005).

1.4 Structure of the Dissertation

The dissertation structure is as follows; Chapter 2 reviews literature on semi-parametric additive models, modelling daily optimum electricity demand using extreme value theory (EVT) and Bayesian statistics. Chapter 3 reviews the relevant methodology. Chapter 4 deals with modelling daily peak electricity demand using discrete time Markov chain analysis (DTMC). The dissertation is divided into two broad sections. Chapter 5 deals with fitting of the short term electricity predicting model (semi-parametric additive model) and a benchmark model, from which an out-of-sample evaluation will be undertaken to determine the better model. Modelling of tail behaviour of residual demand series through the EVT approach is then covered in Chapter 6. A summary of research findings, contributions, suggested areas for further study and concluding remarks are presented in Chapter 7.

The following statistical packages were used in this dissertation: R, R-studio and Eviews. Some of the R codes are given in an appendix section.

Chapter 2

Literature Review

2.1 Introduction

Short term probabilistic load forecasts are important in dispatching and scheduling of electricity, in load flow analysis, reliability analysis including unit commitment and maintenance planning. In this section we review some literature on short term probabilistic forecasting. Furthermore, the literature review discusses the semi-parametric additive regression model of Fan and Hyndman (2012), which relates "forecasts of models half hourly electricity demand for up to a week ahead", in terms of calendar, including work days, and temperature effects, the latter being at two locations. The models of Liu et al. (2006) are stated, as are those of Hor et al. (2005). It is also noted that Goude et al. (2014) suggest that 'a semi-parametric analysis based on generalised additive models' could be used, and further suggest some improvements on the model which they have used. The literature review also outlines Bayesian inference issues as well as those relating to extreme value theory, particularly related to electricity forecasting. Hor et al. (2005) highlight that the proposed models are more numerically intense in prediction than their counterparts, and that the models are not more constraining. In this dissertation, such shortcomings will also be investigated.

2.2 Semi-Parametric Additive Models

Semi-parametric models deal with some very general nonlinear functional forms in regression analysis. Semi-parametric regression modelling is further discussed in detail by (Ruppert et al., 2003).

Semi-Parametric Regression Models

Fan and Hyndman (2012) introduce a short term electricity forecasting model which is used to investigate some effects that exogenous variables have on the electricity demand based on a semi-parametric additive model. It is difficult to find an entire model that incorporates the right exogenous variables; Fan and Hyndman (2012), consider the calendar effects, lagged demand observations and historical temperature data into the nonlinear model proposed. Also, out of sample forecasts on the electricity demand data of 30 minutes up to a week ahead are researched on and the model performance is validated using the real data.

The model for each 30 minute period developed by Fan and Hyndman (2012) is given as:

$$'\log(y_{t_p}) = h_p(t) + f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}) + a_p(\mathbf{y}_{t-,p}) + n_t, \tag{2.1}$$

where y_{t_p} denotes the demand at time t (measured in 30 minute period intervals) during period p ($p = 1, ..., 48$), $h_p(t)$ model all calendar effects, $f_p(\mathbf{w}_{1,t}, \mathbf{w}_{2,t})$ models the temperature effects where $\mathbf{w}_{1,t}$ is a vector of recent temperatures at one area and $\mathbf{w}_{2,t}$ is a vector of recent temperatures at a second area, $a_p(\mathbf{y}_{t-,p})$ models the effects of recent demands, n_t denotes the model error at time t and $h_p(t)$ includes daily, weekly and annual seasonal patterns including the public holidays:

$$h_p(t) = \alpha_{t,p} + \beta_{t,p} + l_p(t) \tag{2.2}$$

where $\alpha_{t,p}$ takes a different value for each day of the week, $\beta_{t,p}$ takes values zero on a non-work day, some nonzero value on the day before a non-work day, and a different value on the day after a non-work day and $l_p(t)$ is a smooth function that repeats each year. Empirical results show that the model performs well on forecasting the Australian half hourly electricity demand (Fan and Hyndman, 2012).

In an effort to model serially correlated data, Liu et al. (2006) consider a semi-parametric two component modelling approach which is made of both the parametric component (e.g "Auto-regressive Integrated Moving Average (ARIMA)") and a non-parametric component. The daily temperature effects are taken into account in a non-parametric model using hourly electricity data. They define their proposed model with two dimensional versions as:

$$Y_t = f_{w_t}(X_t, T_t) + \varepsilon_t \tag{2.3}$$

and

$$(1 - \phi_1 B)(1 - \phi_2 B^{24})(1 - B^{24})\varepsilon_t = (1 - \theta_1 B)(1 - \theta_2 B^{24})a_t \tag{2.4}$$

where $w_t \in \{1, 2\}$ indicates whether observation t belongs to work day or a non-workday and $T_t = t\text{mod}24$ represents the time of the day for observation t, and f is a two dimensional smooth function with T_t being a circular variable.

Empirical results show that the model is flexible and can be used to model highly nonlinear relationships of unknown functional form. However, the authors mention that the theoretical properties of the model and the estimation procedure still require more careful and rigorous study in the future (Liu et al., 2006).

Hor et al. (2005) present a parametric model to investigate the impact that weather related predictor or explanatory variables have on the monthly electricity demand in England and Wales. Hor et al. (2005) use data for the period 1983 to 1995. Since the relationship demonstrated by the temperature and electricity demand is complex, and as a result they use the degree days. The degree days are more descriptive than the temperature versus load relationship. Thus, the predicted electricity demand, $\hat{E}_A$ is expressed as:

$$\hat{Y}_{E_A} = \alpha_0 + \alpha_1 CDD + \alpha_2 HDD + \alpha_3 ELD + \alpha_4 V_w + \alpha_5 M_s + \alpha_6 M_r \tag{2.5}$$

where α_i are constants for $i = 1, 2, ..., n$,

- "CDD- Cooling Degree Days, a measure of the severity and duration of the hot weather used to quantify the cooling requirement".

$$\text{CDD} = \sum_{d=1}^{N_d} (\gamma_d)(T_{dm} - T_{base_C}), \tag{2.6}$$

where T_{dm} is the daily mean temperature, T_{base_C} is a baseline temperature and γ_d is the indicator function.

- "HDD- Heating Degree Days, a measure of the severity and duration of cold weather used to quantify the heating requirement".

$$\text{HDD} = \sum_{d=1}^{N_d} (1 - \gamma_d)(T_{base_H - T_{dm}}), \tag{2.7}$$

where T_{baseH} is the baseline temperature.

- ELD- Enthalpy Latent Days, given by the equation,

$$\text{ELD} = \frac{1}{24} \sum_{d=1}^{N_d} \sum_{h=1}^{24} (\gamma_h)(Q - Q_b), \tag{2.8}$$

where Q is the enthalpy hourly value (in Kilojoules per Kilogram) and Q_b is the enthalpy located at the reference temperature of 25.6° C. Temperature and humidity dependence is expressed in the model more directly. Also, predicted electricity demand is specified as:

$$\hat{y}_{E_B} = \alpha_0^1 + \alpha_1^1 \theta_t + \alpha_3^1 H_r + \alpha_4^1 V_w + \alpha_5^1 M_s + \alpha_6^1 M_r, \tag{2.9}$$

where θ_t is the monthly Central England Temperature (CET), H_r is the mean monthly relative humidity at Elmdon (village in the Uttlesford district of Essex, England) and α_n^1, $n = 1, ..., 6$. are constants, V_w is the mean monthly wind speed, M_s is the mean monthly hours averaged over a number of sites in England and Wales and M_r is the monthly rainfall in millimeters (mm) summed over a number of sites in England and Wales.

Finally they integrate both relative humidity, degree day variables and predicted electricity demand,

$$\hat{E}_c = \alpha_0^{11} + \alpha_1^{11} CDD + \alpha_2^{11} HDD + \alpha_4^{11} V_w + \alpha_5^{11} M_s + \alpha_6^{11} M_r, \tag{2.10}$$

Hor et al. (2005) also suggest a prediction model with non-weather related variables which extremely affects the electricity demand. The models for both England and Wales include in them the population growth variables like the Gross Domestic Product (GDP). To model electricity demand in France of over 220 substations, Goude et al. (2014) suggest using a generalized additive model with a semi-parametric additive framework at both short- and middle term periods. Among other things, the model approximates the temperature relationship linking the electricity demand and the driving explanatory variables. The degree of information from the paper also highlights that common themes throughout these models state that electricity demand versus some exogenous variables are narrated by the estimate function is clear and the temperature prediction validity is essential. Moreover, empirical results from this paper highlight and provide the semi-parametric additive framework with embedded information, and its ability to capture automatically the consumption series measured on the French grid.

Based on successive experiments conducted, Goude et al. (2014) suggest the following models for short and middle term forecasting respectively:

$$\begin{aligned} y_t = \sum_{j=1}^{7} m_j \mathrm{I}_{\mathrm{DayType}_t = j} + k \mathrm{I}_{\mathrm{SpecialTariff}_t = 1} \\ + g_1(\theta_t) + g_2(\theta_t) + g_3(T_{t-1}) + g_4(T_{t-2}) \\ + \sum_{j=1}^{11} o_j \mathrm{I}_{\mathrm{OffSet}_t = j} + h(toy_t) \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} y_t = \sum_{j=1}^{7} m_j \mathrm{I}_{\mathrm{DayType}_t = j} + K \mathrm{I}_{\mathrm{SpecialTariff}_t = 1} \\ + g_1(\theta_t) + g_2(T_t) + g_3(T_{t-1}) + g_4(T_{t-2}) \\ + \sum_{j=1}^{11} O_j \mathrm{I}_{\mathrm{OffSet}_t = j} + h(toy_t) + l(y_{t-1}) + \varepsilon_t, \end{aligned} \tag{2.12}$$

where y_t is the electricity demand recorded at time t for one instant of the day, DayType_t is the type of day for observation t: 1 for Sunday, 2 for Monday, 3 for Tuesday, Wednesday, Thursday, 4 for Friday, 5 for Saturday and 6 and 7 for Bank holidays, SpecialTarriff_t is a factorial predictor, taking values 0 when there is special tariff at time t and 1 otherwise, θ_t is a smooth temperature which is derived from exponential smoothing of the real temperature, $T_t : \theta_t = (1 - 0.99)T_t + 0.99\theta_{t-1}$, T_{t-1} is the lag temperature of the day before, and T_{t-2} is the lag 2 temperature, OffSet_t is a categorical variable indicating the holidays and daylight saving time, toy_t is the time of the year which is the position of the observation t within the year, from 0 Jan the 1^{st} to 1 Dec 31^{st} and $h(toy_t)$ corresponds to the variation of the yearly cycle of the load per instant of the day.

Goude et al. (2014) suggest in a number of ways the models should be improved, such as considering automatic selection of covariates and even introducing other covariates at different resolutions, and also possibly consider the multivariate analysis methodology.

2.3 Modelling Framework: Bayesian Analysis

Over the past decade, the use of Bayesian statistics has steadily gained its ground and continues to evolve. However, most studies have focused on frequentist point forecasting. The limitation of using frequentist methods is that, it does not take into account uncertainty in parameter estimation. Unlike a frequentist approach, the Bayesian approach is more informative, as it captures uncertainty in the estimation of the parameters and provides a coherent tool for decision making. Bayesian approach also provides a full range of sample inferences which involve the predictive distribution of peak electricity demand, the probability distribution of both time and day-to-day peak electricity demand (Cottet and Smith, 2003).

Nascimento et al. (2012) review a semi-parametric Bayesian analysis which includes a mixture of gamma densities and Peaks Over Threshold distribution as the tail distribution. Additionally, in the bulk distribution they establish the numeral gamma components (Γ), taking into account the "Bayesian Information Criterion (BIC)" or "Deviance Information Criterion (DIC)" based statistics. Nascimento et al. (2012) adopt some parameter restrictions in order to overcome the identity problem arising due to different densities evaluated in the underlying model. The density function of their model is defined as:

$$f(x|\theta, \mathbf{p}, \psi) = \begin{cases} h(x|\mu, \eta, \mathbf{p}), & \text{if } x \leq u \\ [1 - H(u|\mu, \eta, \mathbf{p})]g(x|\psi), & \text{if } x > u, \end{cases}$$

where $H(.|\theta)$ is the distribution of the mixture of Gamma defined as

$$h(x|\theta, \mathbf{p}) = \sum_{j=1}^{k} p_j F_G(x|\mu_j, \eta_j). \tag{2.14}$$

where $\theta = (\mu, \eta)$ denotes the Gamma parameters $\mu = (\mu_1, \mu_2, ..., \mu_k)$ and $\eta = (\eta_1, ..., \eta_k)$, $\mathbf{p} = (p_1, p_2, ..., p_k)$ denotes the mixture weights and f_G is the density of the Gamma distribution given by

$$f_G(x|\mu, \eta) = \frac{(\eta/\mu)^\eta}{\Gamma(\eta)} x^{\eta-1} \exp(-(\eta/\mu)x) \tag{2.15}$$

for $x > 0$.
The $p_j's$ must be non negative and should add up to 1 while $\mu_j's$ and $\eta_j's$ can take any positive value and the $g(x|u, \sigma_u, \xi)$ represents the GPD density function.

Using Bayesian based analysis, photo-voltaic generators, phase load demands and power production wind is forecasted by Bracale et al. (2013). By means of probabilistic load flow, the proposed models are used in short term optimum state approach of a smart grid acquired by means of a probabilistic load flow. The stochastic behaviour of the loads demands and photo-voltaic power production and the uncertainties characterizing the wind are accounted for. However, such models need to be improved so that they are readable and accurate.

Cottet and Smith (2003) discuss how a multiple regression model can be applied as substantial modelling and prediction methods for day-to-day electricity demand. Further on, the evolution appropriate for the disturbances correlation structure to the value at risk (VAR) and the suitable subset of regressors are generally investigated using Bayesian model based analysis. This method is applied to several multiple-equation models of 30 minute periods total system load in New Wales, Australia. Moreover, in the analysis, empirical results show that the Bayesian predictive means for 30 minute periods load compare favorably with point forecasts obtained using iterated generalized least squares estimation of the same models.

Bayesian analysis does not depend on regularity conditions required for maximum likelihood estimation of parameters of extreme value distributions (Coles et al. (2001); Beirlant et al. (2006)). However, for a detailed discussion on the advantages and disadvantages of using Bayesian than frequentist methods see Berger (2013); Bernardo and Smith (2009); Wasserman (2013); among others.

2.4 Modelling Daily Peak Electricity Demand using Extreme Value Theory (EVT)

For the last decade, EVT has been extensively used in various economic sectors, to modelling the tail distribution and extrapolation of the behaviour function. However, there are a few papers in literature devoted to modelling and managing risks in the electricity sector using extreme value models.

A notable paper to consider is that of Hor et al. (2006) in which they outline the use of the EVT techniques to estimate the maximal electricity prediction errors to several decades ahead in order to evaluate risks in long horizon electricity demand predictions, with impacts associated to the climate. The research explores the maximum residuals in terms of return levels and return periods that would occur with a finite time series. Using the number of returns exceeding a given return level, empirical results show that the results are closely in line with what was predicted. On the whole, the significance of this paper is that extreme errors in electricity prediction can be modelled, making sure uncertainty is accounted for.

Chan and Nadarajah (2015) propose the use of the GPD to illustrate how the electricity demand extremes differ with time in the United Kingdom (UK). Additionally, the pivotal idea underlying their study is to allow the GPD parameters to vary linearly and in a non-stationary sinusoidal manner. Essentially, this is done in order to capture some patterns unveiled by the electricity demand data. Initially, they fit four different models, both stationary and non-stationary; to investigate if the seasonality and trend are significant. Among the models considered, the best models are:
Model S2:

$$Y \sim \text{GPD}\left(\sigma = \exp\left[a + b\sin\left(\frac{\pi M}{12}\right) + c\cos\left(\frac{\pi M}{12}\right)\right], \xi\right) \tag{2.16}$$

where ξ is fixed, M denotes the month number and a, b and c are parameter estimates
Model T2:

$$Y \sim \text{GPD}(\sigma = \exp[g + h(X - 2010)], \xi) \tag{2.17}$$

where ξ is fixed and X denotes the year.

Hasan et al. (2013) model extreme temperature of various stations in Malaysia using the generalized extreme value distribution (GEVD). The trend component for each station is detected by performing the Mann-Kendall (MK) trend in the data. They model yearly maximal temperatures by using the stationary and non-stationary GEVD for various stations in Malaysia. In this paper they also use the LMOM (L-moments methods) and MLE to find the underlying model parameter estimates.

Sigauke et al. (2013) investigate extreme peak electricity demand increases on a day-to-day basis in South Africa. In this study the two peaks over threshold distributions are considered. The distributions are the GPD and the Generalized Single Pareto (GSP). Further, to estimate the parameters a Bayesian based analysis is used and posterior predictive tail probabilities. Accordingly, posterior predictive tail probabilities of future observations are estimated using:

$$P(Y_0 > y_0|y, \tau) \propto \int\int \pi(\eta|y) \left(1 + \frac{\xi}{\sigma}(y_0 - \tau)\right)^{-\dfrac{1}{\xi}} d\sigma d\xi \tag{2.18}$$

and

$$P(Y_0 > y_0|y, \tau) \propto \int \pi(\eta|y) \left(1 + \frac{\eta}{1+\eta_\tau}(y_0 - \tau)\right)^{-\dfrac{1}{\eta}} d\eta, -\infty < \eta < \infty \tag{2.19}$$

derived from the GPD and GSP distribution respectively.

Tong et al. (2014) examine a series of prediction errors and their features, by comparing a variety of other probability distribution functions. Further, the authors carry out some discussion and identification of the Student t-distribution's adaptability. Moreover, empirical analysis is based on real data from China to validate and verify the effectiveness of short term electricity prediction based on the t-distribution. On the whole, this is useful as it helps minimise risks associated with the operation of the power-system.

Renard et al. (2006) deal with the non-stationary data making use of the Bayesian analysis. For several models that include extreme value distributions (exponential, generalized Pareto, Gumbel, GEVD) a Bayesian analysis is undertaken. Essentially, the rationale behind this approach is for accounting for non-stationarity in the data by assuming the linear trend (Renard et al., 2006).

Cheng et al. (2014) further introduce a framework for approximating return levels and return periods of both the stationary and non-stationary case of the Bayesian inference analysis. Cheng et al. (2014) implement the framework, on the software package "Non-stationary Extreme Value Analysis (NEVA)" in order to assess the extremes in changing climate and assess the trends in the observations. Empirical results confirm that the climate is non-stationary, and this could be due to climate change mainly influenced by human activities. Eastoe and Tawn (2009) also propose a technique used for modelling non-stationary processes extremes. A comparative analysis is done with other standard methods. Eastoe and Tawn (2009) use a time varying threshold. Empirical results show that the preprocessing approach is superior as compared to its counterparts, the reason being it is prospective to selecting the covariate model more precisely than the standard method. Also, the proposed approach has more dominance in terms of computational simplicity, on account of the transformed process extremes $\{Z_t\}$ being more closer to stationarity than those of the original data $\{Y_t\}$. Eastoe and Tawn (2009) highlight the necessary future work; to study ways in which we can retain threshold stability property while modelling the covariates using GPD parameters.

Sugahara et al. (2009) review an assessment of the rainfall extremes for the period 1933-2005 in Sao Paulo using GPD. Sugahara et al. (2009) consider a time varying threshold, which depends on a specific day annually, denoted as $q_p(t_d)$, $t_d = 1, 2, ..., 365$, composed of p-quantiles of day-to-day rainfall volume acquired from all the present data. In this study, seasonality and trend are non-stationary features present in the Sao Paulo data both integrated as covariates.

Chapter 3

Methodology

The generalized additive regression model and EVT will be used to model hourly electricity demand in this dissertation. Furthermore, a comparative analysis of modelling stationary and non-stationary extreme values of electricity demand will be carried out. In addition to this, to estimate the parameters of the underlying models we will use the Maximum likelihood estimation (MLE) method and also the Bayesian approach.

3.1 Semi-Parametric Additive Model

3.1.1 Setting up the Model

Classical semi-parametric regression is used to develop a nonlinear functional relationship in regression analysis. Further, the semi-parametric additive models are then used to estimate the relationship between the dependent variable and a range of independent variables. The common approach to estimating the relationship follows that the response variable (y) is influenced by some covariates ($x_1, x_2, ..., x_p$) in the linear regression model

$$y_i = \beta_0 + x_{i1}\beta_1 + ... + x_{ip}\beta_p + \varepsilon_i \tag{3.1}$$

for $i \in \{$1,2,...,n$\}$, $\beta_0 \in \mathbb{R}$ is the regression intercept, $\beta_k \in \mathbb{R}$ is the k^{th} predictor regression slope, $k = 1, ..., p$ and $\varepsilon_i \sim N(0, \sigma^2)$. In this dissertation we consider a more general approach and use the semi-parametric additive models

$$y_i = f(x_{i1}, ..., x_{ip}) + \varepsilon_i \tag{3.2}$$

which were introduced in the literature by Ruppert et al. (2003). Equation 3.2 gives the general form of the nonlinear regression model. Equation (3.2) is a special case of equation (3.1). Considering the additivity assumption in (3.1) and substituting the structure of the functional form and additivity form yields.

$$y_i = \beta_0 + f_1(z_{i1}) + f_q(z_{iq}) + x_{i1}\beta_1 + ... + x_{ip}\beta_p + \varepsilon_i \tag{3.3}$$

with $z_i = (z_{ij})$, $i \in (1, 2, ..., n)$, $j \in (1, 2, ..., q)$, consisting of metrically scaled covariates. The model in equation 3.3 is a semi-parametric additive model as suggested by Ruppert et al. (2003). Models of the form given in equation (3.3) are discussed in Hastie and Tibshirani (1990). We will use the model given in equation (3.3) which is a slight modification of the model developed by Fan and Hyndman (2012) model development, and Equation (3.4) is the mathematical expression of interest in this dissertation.

$$\log(y_{t,h}) = \sum_{j=2}^{7} \beta_j \mathrm{D}_{jt} + \sum_{i=2}^{12} \alpha_i \mathrm{M}_{it} + \sum_{m=1}^{M} w_m \prod_{k=1}^{k} [s_{km}(x_{v(k,m)} - t_{km})] + h_p(t) + l_d(x_{t-n}) + u_t \tag{3.4}$$

where $h_p(t)$= holiday effect, $l_d(x_{t-p})$= lagged demand, u_t = error term is assumed to follow the SARIMA model and its general structure is given as,

$$\phi_p(B)\Phi_Q(B^s)u_t = \theta_q(B)\Theta_Q(B^s)e_t \tag{3.5}$$

where e_t is the white noise of the SARIMA model, $u_t = \nabla^d \nabla_s^D x_t$ is the error term, x_t represents the load at time t, s represents the seasonal length, while B represents the lag operator, ∇^d and ∇_s^D are the non-seasonal and seasonal difference operators d and D, respectively. $\nabla^d = (1 - L)^d$ and ∇_s^D , and ϕ_p, Φ_p, θ_q, Θ_Q are the respective polynomial functions of orders p, P, q and Q. The structure of the error term u_t is meant to account for serial correlation (Hyndman and Athanasopoulos, 2014).

3.1.2 Model Fitting

We will estimate the model parameters using MLE method and the Bayesian inference to find estimates that best fit the data to the semi-parametric additive model.

Parameter Estimation of Semi-Parametric Models

In this research we will use the basic frequentist method, or Maximum Likelihood Estimation, and finally also use the Bayesian method.

3.1.2.1 Maximum Likelihood Estimation

Contextually, the likelihood is naturally as function of the parameter θ in terms of parameter estimation. In essence, the model parameters are assumed to be fixed but unknown. This is discussed in detail in Section 3.6,

3.1.2.2 Bayesian parameter Estimation

Under this consideration, with some known prior distribution parameters are assumed to be random variables, and uncertainty about parameters is represented by a probability density function (pdf). Bayesian methods seek to estimate the posterior density $p(\theta|X)$.

A detailed discussion of the two parameter estimation methods, namely MLE and Bayesian parameter estimation is given in Section 3.5.

3.1.3 Evaluation Methods for Model Sufficiency and Prediction Accuracy

3.1.3.1 The evaluation Methods

Generally we use Akaike Information Criterion (AIC) (Akaike, 1998) and Bayesian Information Criterion (BIC) (Schwarz, 1978) as an evaluation method to select the best fitted model. For the most part, the two criterions, AIC and BIC are acquired from definite views. For instance, AIC is meant to reduce the Kullback-Leibler divergence between the true distribution and the estimate from a candidate model, while BIC is meant to select a model that suitably maximizes the posterior model probability (Yong, 2005).

3.1.3.2 Measurement of prediction performance

Generally to examine the prediction precision and accuracy of the model there are a number of measurement statistics that are used namely the root mean square error (RMSE), mean absolute error (MAE) and lastly mean absolute percentage error (MAPE). Furthermore, the three measurement statistics (RMSE, MAPE and MAE respectively) can be computed as follows,

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{3.6}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100, \tag{3.7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{3.8}$$

where n represents observations in testing dataset, y_i is the actual value and $\hat{y}_i$ is the predicted value.

3.1.4 Model Selection

The model specification depends, to some extent on the covariates incorporated in the model. The objective is to build a model that includes appropriate set of exogenous variables which have a causal relationship with the response and also offer sufficient amount of new information about the response (hourly demand). We want to exclude predictor and explanatory variables that provide better results about the response.

3.1.5 Model Diagnostics

3.1.5.1 Checking model assumptions

The residuals are determined by simply subtracting the predicted values from the actual values: $\hat{e} = y_i - \hat{y}_i$. Before we make any model-based conclusions, we will assess if the residuals satisfy the model assumptions. Among a series of assumptions that needs to be satisfied, is that the residuals should be identically independently distributed, i.e., $e_i \sim N(0, \sigma^2)$.

3.2 Daily Peak Electricity Demand Modelling using Extreme Value Theory (EVT)

Classical EVT is used to develop Stochastic models towards real life problems related to occurrence of rare events. Moreover, classical theoretical results are concerned with the stochastic behaviour assume some maxima's (minima's) of random variables sequence are be i.i.d. More importantly, we will model the behavior of the extreme residuals using the GPD for both the stationary and non stationary sequences in this chapter. We will fit the GPD to the positive residuals (under estimated demand). It should be noted that the same procedure can be applied using negative residuals (over estimated demand).

3.3 Generalized Pareto Distribution (GPD)

We use the Peaks Over Threshold (POT) distribution in order to model all the observations above a sufficiently high threshold. Coles et al. (2001) explain that if $x_1, x_2, x_3, ..., x_n$ is a sequence of independent and identically distributed (i.i.d) random variables, then the excess $x - \tau$ over a suitable threshold τ can be approximated by the GPD written as $G(x|\tau, \sigma_\tau, \xi)$.

Let $r_1, r_2, ..., r_{n_\tau}$ be the sequence of hourly peak electricity demand errors above a sufficiently high threshold, τ.

The distribution function of GPD is given by:

$$G(r|\tau,\sigma_\tau,\xi)=\begin{cases}1-\Big[1+\xi\Big(\dfrac{r-\tau}{\sigma_\tau}\Big)\Big]_+^{-\dfrac{1}{\xi}}, & \text{if } \xi\neq 0\\ \\ 1-\exp\Big[-\Big(\dfrac{r-\tau}{\sigma_\tau}\Big)\Big]_+ \end{cases} \tag{3.9}$$

where $r>0$, $\sigma_\tau>0$, $\Big(1+\xi\Big(\dfrac{r-\tau}{\sigma_\tau}\Big)\Big)>0$ and $\tau\in(\tau_+,\tau_-)$

and the probability density function is given by :

$$g(r|\tau,\sigma_\tau,\xi)=\frac{1}{\sigma_\tau}\left[1+\xi\Big(\frac{r-\tau}{\sigma_\tau}\Big)\right]^{-\dfrac{1}{\xi}-1} \tag{3.10}$$

. The shape parameter ξ is essential as it determines the GPD tail distribution. The conditional survival probability function is defined as:

$$P(R>r|R>\tau)=\begin{cases}\left[1+\xi\Big(\dfrac{r-\tau}{\sigma_\tau}\Big)\right]_+^{-\dfrac{1}{\xi}}, & \xi\neq 0\\ \\ \exp\left[-\Big(\dfrac{r-\tau}{\sigma_\tau}\Big)\right]_+, & \xi=0\end{cases} \tag{3.11}$$

where $P(R>r|R>\tau)$ is the probability of the electricity peak demand residuals being above the threshold, $\tau\in(\tau_+,\tau_-)$.

3.4 Non Stationary and Stationary Extremes

Suppose with a univariate marginal distribution F which is characterized by an upper end point $x^F>0$, the process of interest $\{Y_t\}$ is fixed. For details see Eastoe and Tawn (2009). All extreme observations of $\{Y_t\}$ above a sufficiently high threshold τ, $Y_t\sim \text{GPD}\,(\sigma,\xi)$ are called exceedances. Pickands III (1975) shows how the distribution of the excesses, $Y_t-\tau$, converges to the GPD.

The conditional survival function for all the exceedances, assuming that the excesses follow a GPD(σ_τ, ξ) model and the quantile function are given in equations (3.10) and (3.11) respectively.

$$Pr(Y > y + \tau | Y > \tau) = \left[1 + \xi\left(\frac{z - \tau}{\sigma_\tau}\right)\right]_+^{-\frac{1}{\xi}} \tag{3.12}$$

for $y > 0$ and $\sigma_\tau > 0$

$$\hat{y}_p = \tau + \frac{\hat{\sigma}_\tau}{\xi}\left\{\left(\frac{\hat{\sigma}_\tau}{p}\right)^{\xi} - 1\right\}. \tag{3.13}$$

The quantile function given in equation (3.11) is used to estimate high quantiles including return periods.

Eastoe and Tawn (2009) extend the models given in equations (3.9) and (3.10) to the non-stationary case. Suppose the process $\{Y_t\}$ is non-stationary and has some covariates with associated sequences $\{\mathbf{X}_t\}$, $Y_t = \text{GPD}(\sigma(t), \xi(t))$

The rate of exceedance is $\phi_\tau(\mathbf{X}) = pr(Y > \tau | \mathbf{X} = \mathbf{x})$ and the distribution of all observations above the threshold by a GPD for $y > 0$ is,

$$Pr(Y_t > y_t | Y_t > \tau, \mathbf{X}_t = \mathbf{x}_t) = \left[1 + \frac{\xi(t)(r(t) - \tau(t))}{\sigma(t)}\right]^{-\frac{1}{\xi(t)}}. \tag{3.14}$$

3.5 Threshold Selection

This is an extensive ongoing area of research in the literature. Coles et al. (2001) state that more often the threshold selection is repeatedly a trade off between both the bias and variance. Additionally, if the determined threshold is too low, the asymptotic properties underlying the derivation of the GPD are violated. In addition, if the threshold is too high, it prompts less excesses, $(x - \tau)$ to estimate the shape and scale parameter resulting in a higher variance. Thus, threshold selection requires considerable investigation, whether the limiting model provides a sufficiently good approximation versus the variance of the parameter estimates.

There are diagnostic models used to evaluate the model fit. Coles et al. (2001) outlined various three diagnostic plots for the choice of threshold, which include Mean

Residual Plot, parameter stability plot and a more general model fit diagnostics plots.

3.5.1 Mean Residual Plot

This is one of the few methods used in undertaking the threshold selection $\tau \in (\tau_+, \tau_-)$. This technique came to be common after it was introduced by Davison and Smith (1990). Contextually, the mean excess function, $e(t)$ is defined as

$$e(t) = E(X - \tau | X > \tau) \tag{3.15}$$

(Beirlant et al., 2004a). A mean excess plot is defined as a plot of $x_{n-k,n}$ versus the mean excesses $e_{k,n}$ with the latter defined as $e_{k,n}$ (Beirlant et al., 2004a):

$$e_{k,n} := \hat{e}_n(x_{n-k,n}) = \frac{1}{k}\sum_{i=1}^{k} x_{n-i+1,n} - x_{n-k,n} \quad k = 2, ..., n, \tag{3.16}$$

where $\hat{e}_n$ is the empirical function used to estimate e. The optimum threshold τ is selected where the plot starts to follow a linear pattern (Beirlant et al., 2004a).

3.5.2 Parameter Stability Plot

This is the diagnostic plot that is used to assess the model fit in order to determine the threshold. If all observations above the threshold τ follow a GPD with the parameters ξ and σ_τ, for any higher threshold (τ_0), i.e $\tau_0 > \tau$, all the excesses will follow a GPD with ξ and the scale parameter of:

$$\sigma_{\tau_0} = \sigma_\tau + \xi(\tau_0 - \tau) \tag{3.17}$$

Parametrizing the scale parameter σ_{τ_0}, we have;

$$\sigma^* = \sigma_{\tau_0} - \xi\tau_0. \tag{3.18}$$

Given (τ_0) is a sufficiently higher threshold then σ^* does not depend on (τ_0). The parameter stability plot of the GPD means that the threshold (τ) should be chosen where the shape and modified scale parameter remain constant, after taking the sampling variability into account. The stability plot also fits GPD over certain values of thresholds against the shape and scale parameters.

3.5.3 Pareto Quantile Plot

This is defined as the scatter plot of the following points: $(-\log(1-p_i), \log x_i)$ where $i = 1, ..., n$ (Beirlant et al., 1996, 2004a). Several choices of p_i are used in the literature (Beirlant et al., 2004a). In this dissertation $p_i = \frac{i}{n+1}$ is used. Notably, where the plot starts to follow a linear pattern the threshold is considered as the observation found on the vertical axis (y-axis) (Beirlant et al., 2004a).

3.5.4 Extreme Value Mixture Models

The three outlined threshold diagnostic approaches above have one major draw back; they do not take into account uncertainty surrounding the model. To overcome such a drawback (Scarrott and MacDonald, 2012) introduced extreme mixture models. For the past decades, extreme value mixture models have received huge interest from various researchers. Extreme value mixture models provide for an automated approach for both threshold estimation and quantification of uncertainties that result from threshold choice (Hu (2013); Bommier (2014)). Hu (2013) emphasizes that extreme value mixture models have two components; a model for describing all the non-extreme data below the threshold and a traditional technique like EVT. Notably, these models are important in threshold selection, as they account for threshold tail distribution and the bulk distribution below (Hu, 2013). Under these circumstances, all the important information is not wasted as all the available data is taken into consideration. A reasonably comprehensive review and literature of extreme value mixture models can be found in (Hu (2013); MacDonald et al. (2011); Scarrott and MacDonald (2012)).

3.6 Parameter Estimation

3.6.1 Maximum Likelihood Estimation

Let $r_1, r_2, r_3, ..., r_{n_\tau}$ be a sequence of day-to-day peak electricity demand exceedances $\tau \in (\tau_+, \tau_-)$.
i.e. $r_i \sim \text{GPD}(\sigma_\tau, \xi)$, $i = 1, 2, 3, ..., n_\tau$. where n_τ is the total number of exceedances, $\tau \in (\tau_+, \tau_-)$.
Contextually, the likelihood function is given as:

$$L(z|\sigma_\tau, \xi) = \prod_{i=1}^{n_\tau} \frac{1}{\sigma_\tau} \left[1 + \xi\Big(\frac{r_i - \tau}{\sigma_\tau}\Big)\right]^{-\frac{1}{\xi}-1} \tag{3.19}$$

and the log likelihood is given as:

$$\text{log}L(\sigma_\tau, \xi|r) = -n_\tau \text{log}\sigma_\tau - \Big(1 + \frac{1}{\xi}\Big)\sum_{i=1}^{n_\tau} \text{log}\Big[1 + \xi\Big(\frac{r_i - \tau}{\sigma_\tau}\Big)\Big], \xi \neq 0. \tag{3.20}$$

For $\xi = 0$, the log likelihood can be obtained in a comparable way:

$$\text{log}L(\sigma_\tau|r) = -n_\tau\sigma_\tau - \sum_{i=1}^{n_\tau}\Big(\frac{r_i - \tau}{\sigma_\tau}\Big) \tag{3.21}$$

and similarly the same approach is used to estimate the semi-parametric additive model parameters and also for non stationary extremes.

3.7 Daily Peak Electricity Demand Forecasting using Bayesian Approach

In this section, we use the Bayesian based analysis to make inferences about the parameter estimates on day-to-day peak demand in South Africa. Bayesian analysis provide a quantitative framework for representing current knowledge and to rationally integrate new information (Cowles, 2013). Overall, it is established on the idea to express uncertainty of the unknown probability distribution (Raftery, 1995).

In Bayesian analysis the first step starts with reflecting on the probability distribution of the current state of the hourly maximal electricity demand via Bayes' theorem. It takes into account information about parameters, including access to expert information based on historic experience or related experience (Kass and Raftery, 1995).

The major difference between the Bayesian and classical methods is in conditioning, see Cyert and DeGroot (1970); Poirier (1988); among others for a distinction between Bayesian and non Bayesian methods. The hourly maximal electricity demand is modelled through non informative prior distributions. Coles and Powell (1996) comment that if the data on the extremes and scarce, then it may not be possible for an expert to independently formulate prior beliefs about the process. Prior distribution covers the range of plausible values and is fairly flat over that range (Raftery, 1995; Kass and Raftery, 1995).

The second step involves collection of the data to form the likelihood function, $\pi(x|\theta)$. Bayesian statistics incorporates the prior distribution along with the likelihood function to yield the posterior distribution $\pi(\theta|x)$ for parameter estimates, from which inferences can be made. The posterior distribution is given as:

$$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\int_{\Omega} \pi(x|\theta)\pi(\theta)d\theta} \tag{3.22}$$

$$\Rightarrow \pi(\theta|x) \propto \pi(\theta)\pi(x|\theta), \tag{3.23}$$

where, $\pi(\theta)$ is the prior distribution, $\int_{\Omega} \pi(x|\theta)\pi(\theta)d\theta$ is the predictive distribution, $\pi(x|\theta)$ is the likelihood function and the integral is taken over the parameter space Ω.

The posterior distribution represents the combined information of what was known about the parameters beforehand and what the data conveys (Van Dongen, 2006). Inferences about the parameters are based on the posterior distribution, since it summarises all available information (Kass and Raftery, 1995).

The mean $E(\theta|x) = \int_{\Omega} \theta\pi(x|\theta)d\theta$ can serve as an estimate of the parameter θ. The future values of x can also be predicted by using posterior predictive distribution which is given by

$$\pi(x_{n+1}|x) = \int_{\Omega} \pi(x|\theta)\pi(\theta|x)d\theta. \tag{3.24}$$

Chapter 4

Modelling Daily Peak Electricity Demand in Different Hours of the day Using Discrete Time Markov Chain

4.1 Introduction

We present the use of discrete time Markov chain (DTMC) analysis on the day-to-day electricity demand and present the empirical results. Electricity demand exhibits a large degree of randomness, particularly in South Africa. Even more, peak electricity demand forecasting has proven to be a widespread strategic policy concern for many countries globally (Strengers and Maller, 2012). Most compelling evidence, South Africa is faced with challenges that include planning and scheduling maintenance appropriately and electricity shortfalls such as load shedding, power outages and blackouts occurrences. To point out, this chapter presents the effect that peak electricity demand have on the South African electricity market and investigate some relationships between variables that possibly drive peak demand along with others.

Significantly, we model peak electricity demand in different hours of the day using DTMC analysis in order to fully examine and understand which hours of the day are experiencing maximal electricity demand daily.

The organization of this chapter is as follows: We further discuss discrete time Markov chains and some underlying concepts in Section 4.2; Section 4.3 presents some discussion on the empirical results and Section 4.4 presents some concluding remarks.

4.2 Finite Discrete Time Markov Chains (DTMC)

Generally a Markov chain describes some movement of states from one position to another. Additionally, DTMC analysis is a type of Markov chain in which a finite state space exists and one of the advantages of using the Markov chain models is that they capture all dominant factors of uncertainty (Yin and Zhang, 2006). In detail, DTMC involves modelling transition probabilities of demand changes in different hours. We first start to define the two states; positive day-to-day changes (increase) and negative day-to-day changes (decrease), and the transition matrix represents the different states probability distributions. Also, the probability of each $\{i, j\}$ is denoted by p_{ij}.

This tool is characterized by the Markov property which is fully explored in subsection 4.2.2. Furthermore, we calculate the mean return and steady state probabilities of the daily peak electricity demand in different hours, and also their first passage times. Henceforth, the two state problem is then extended to the three state problem, where the daily electricity demand changes in different hours are split into the following states: small increases, extreme increases of the hourly peak demands and last, negative hourly changes are classified as decreases. Primarily, we fit the positive daily changes to the Pareto quantile plot in order to determine a sufficiently high threshold τ. All the exceedances are regarded as extreme positive hourly demand changes and those below τ as small increases. Further, a transition probability matrix is developed from which we calculate the mean return times, steady state probabilities and also the first passage times. We calculate extreme quantiles and Probability of Exceedance (PoE) of the extreme peak electricity demand in different hours using GPD.

4.2.1 Description of the Data

The transition matrices, steady state probabilities, mean return and first passage times are developed from the hourly historical data of electricity demand from Eskom (South African utility company). Eskom is responsible for facilitating and supplying electricity to the whole South African population (about 50 million people) and other parts of Southern Africa. The historical data comprises of the aggregated hourly data (in megaWatts) for the period 2000 to 2010. Aggregated data comes from various sectors (residential , industrial and commercial sector) that contribute to the well being of South Africa's economic growth. Figure 4.1 presents the daily peak demand time series plot, and exhibits both some positive upward trend and strong seasonality. As clearly depicted in Figure 4.1, the data is not stationary. To make the data stationary we take the first difference of the electricity demand data as shown in Figure 4.2. The series first difference is specified as:

$$z_t = \nabla x_t = (1 - \beta)x_t = x_t - x_{t-1} \tag{4.1}$$

where,
z_t denotes day-to-day peak electricity demand changes, and
x_t denotes day-to-day peak electricity demand on day t.

Primarily we determine the threshold to be zero so that we can specify and name the states. After the threshold have been set at zero, the system results in a two-state problem with the underlying states specified as:
state 1: when $x_t - x_{t-1} > 0$ (state of an increase)
state 2: when $x_t - x_{t-1} \leq 0$ (state of a decrease)

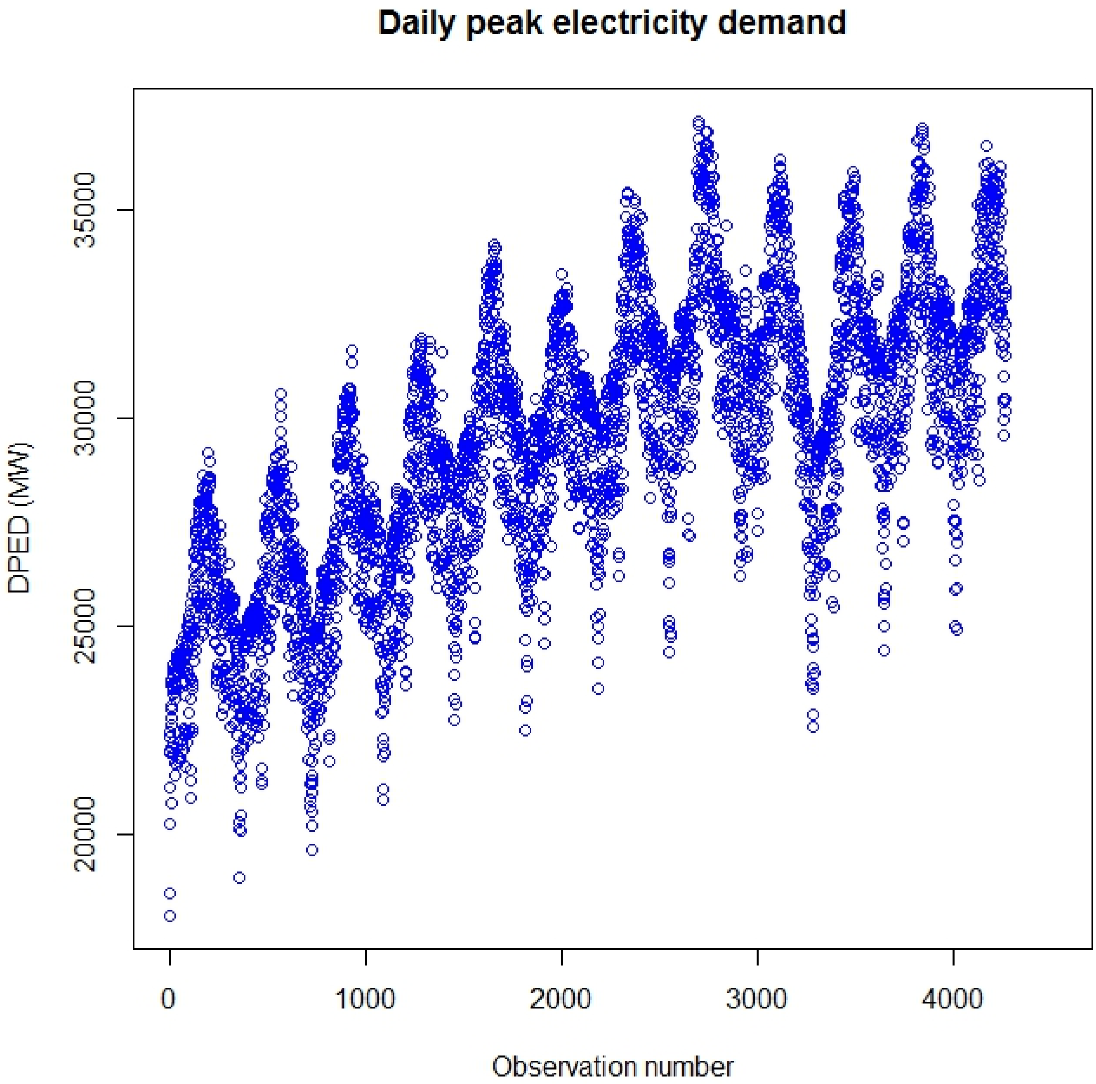


FIGURE 4.1: Daily peak electricity demand time series plot. The observation number is located in the horizontal axis ($x-axis$) and daily peak electricity demand along the vertical axis ($y-axis$)

.

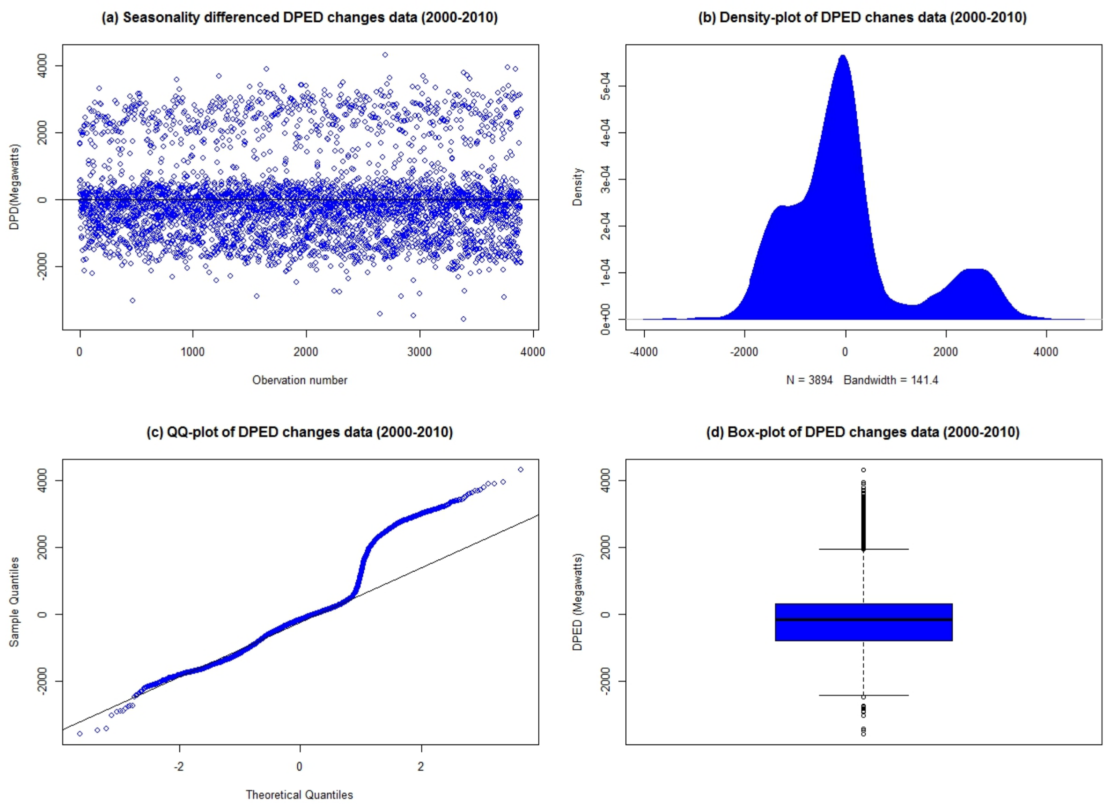


FIGURE 4.2: (a) Seasonally differenced daily peak electricity demand plot (top left panel), (b) DPED changes data density plot (top right panel), (c) DPED changes data QQ-plot (bottom left panel), (d) DPED changes data box-plot (bottom right panel)

Figure 4.2 presents the seasonally differenced daily peak demand time series plot, and daily peak electricity demand changes plots.

4.2.2 Markov Property

Spedicato and Signorelli (2014) defines the Markov property as the distribution of X_{n+1} as a forthcoming state, that depends on the present state only X_n and not on the previous states, $X_{n-1}, X_{n-2}, ..., X_1$. In essence, it is described as a succession of random variables which are characterized with the memorylessness property. Correspondingly, daily peak electricity demand transition probabilities considering an increase state can depend on the present state at that moment and not on the previous state of DPED. The mathematical representation of the property is defined as:

$$Pr(X_{n+1} = x_{n+1|X_1=x_1,...,X_n=x_n}) = Pr(X_{n+1} = x_{n+1}|X_n = x_n). \tag{4.2}$$

4.2.3 DPED Changes Transition Probabilities Estimation

The change from which the system (chain) shifts successively from one state to another is called the transition or commonly the "step" (Spedicato et al., 2015).
Now, we define p_{ij} to be the probability that defines some movement from state i to state j, where $i, j = 1, 2$ and is represented mathematically as:

$$p_{xy} = Pr(X_1 = s_j | X_0 = s_i), i, j = \{1, 2\}, \tag{4.3}$$

where,
state 1: increase change and
state 2: decrease change.
The transition matrix of the two state problem is specified as:

$$\mathrm{P} = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \end{array}.$$

In order to estimate the transition matrix steady state probabilities we use the following

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}, \tag{4.4}$$

where n_{ij} represents the number of transitions observed as there is state movement observed from i to j and $n_i = \sum_{i=1}^{n} n_{ij}$. Moreover we define the state space as the set of all possible DPED states in the system. For example for a two state problem, we have $S = (S_1, S_2)$.

4.2.4 The n-step Transition Probabilities

As an illustration, the n-step transition probabilities formula is given in equation (4.5).

$$P_{ij}^{(n)} = Pr(X_{m+n} = j | X_m = i) = P(X_n = j | X_0 = i) \tag{4.5}$$

The two state problem n-step transition probabilities are specified by the matrix below;

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} p_{11}^{(n)} & p_{12}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} \end{pmatrix} \end{array}.$$

The transition probabilities of the states start to converge when the number of days increase towards some large finite numeral. It converges to some number π_i, after which the chain begins to steadily settle and can be represented mathematically as:

$$r_{di}(n) = P(X_n = i|X_0 = d) = \sum_{k=1}^{n} r_{2k}(n-1)P_{k1} \tag{4.6}$$

$$\lim_{x\to\infty} r_{12}(n) = \sum_{k=1}^{n} \lim_{x\to\infty} r_{dk}(n-1)P_{k1} \tag{4.7}$$

$$\Rightarrow \pi_i = \sum_{k=1}^{n} \pi_2 P_{k1}. \tag{4.8}$$

Now the two state problem steady state probabilities can be seen in equation (4.9)

$$\boldsymbol{\pi} = (\pi_1, \pi_2). \tag{4.9}$$

4.2.5 Mean Return times (MRT)

It is defined as the first visit time that a state takes after a specific number of cycles, specified as t_l. However, this is only valid if the Markov chain system is ergodic (irreducible). The mean return times is specified mathematically as:

$$R_l = E(T_l) = \frac{1}{\pi_l} \tag{4.10}$$

where $\boldsymbol{\pi} = [\pi_1, ..., \pi_r]$ and r is the states number.

4.2.6 First Passage times (FPT)

It is defined as the number of transitions that the Markov chain takes before the cycle of the journey is complete. For instance, we represent the number of transitions for state i to j as T_{ij} and equations (4.11) and (4.12) represents the first passage time probability distribution.

$$h_{ij}^{(n)} = P(T_{ij} = n), \tag{4.11}$$

$$\Rightarrow h_{ij}^{(n)} = P(X_n = j, X_{n-1} \neq j, ..., X_1 \neq j | X_0 = i). \tag{4.12}$$

Now for $n = 1$ in equation (4.11) we have,

$$h_{ij}^{(1)} = p_{ij}. \tag{4.13}$$

Equation (4.14) represents the first passage time for $n \geq 2$ given by:

$$h_{ij}^{(n)} = \sum_{k \in S-j} p_{ik} h_{kj}(n-1). \tag{4.14}$$

4.2.7 Two State Problem

4.2.7.1 Transition matrix and mean return times

We use the R-statistical package "Markovian package" developed by Spedicato and Signorelli (2014). We represent the transition matrix of a two-state problem below:

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 0.4429658 & 0.5570342 \\ 0.3796976 & 0.6203024 \end{pmatrix} \end{array}.$$

As could be seen from the transition matrix, the probability that the next day the current peak electricity demand changes from state 1 and back again is 0.4429658. In detail, there is an approximate 44.3% chance that the system experiences an increase from state 1 and back, and the movement probability from state 1 to 2 is 0.5570342. Also the movement probability experienced by the system from state 1 to 2 is 0.6203024 as can be seen indicated by the transition matrix P. This information is principal during the planning of the power systems operations along with its facilitation. This is to ensure that power outages, blackouts and load shedding do not occur. The two-state problem steady state probabilities are:

$$\pi = (0.4053429, 0.5946571) = (\pi_1, \pi_2).$$

The two-state problem mean return times are :

$R_1 = E(T_1) = \dfrac{1}{0.4026065} = 2.48$ days for state 1 and

$R_2 = E(T_2) = \dfrac{1}{0.5973935} = 1.67$ days for state 2.

The results indicate that, assuming an increase to be the system current state, another increase in about 2.46 days can be expected to be experienced by the system, which is approximately 60 hours. Equally, assuming a decrease (state 2) to be the system current state, then another decrease state can be expected to be seen in about 1.68 days, (approximately 41 hours).

4.2.7.2 First Passage Times (FPT)

Table 4.1: Probabilities of the first passage times (FPT) for the upcoming seven days, assuming that an increase (state 1) is the system current state

Day	Increase (State 1)	Decrease (State 2)
1	0.44296580	0.557034200
2	0.21150455	0.246747100
3	0.13119678	0.109300527
4	0.08138168	0.048416395
5	0.05048125	0.021446807
6	0.03131364	0.009500202
7	0.01942393	0.004208265

A short term summary of the first passage time probabilities (for the next seven days) is depicted in Table 4.1. As the results indicate, assuming state 1 is the system current state, the system probabilities continue to experience state for the next seven days. Meanwhile, assuming state 1 is the system current state, the probability that the next day the system experiences state 1 is 44.3%. In the same way, assuming state 1 is the system current state the probability that the system experiences state 2 is 55.7%. This information is very important to the forecasters to determine which periods are most appropriate to undertake maintenance and monitor the security of the national grid. In order for the economy of the country to continue to grow and remain sustainable, energy security should be a national priority (Sigauke, 2014).

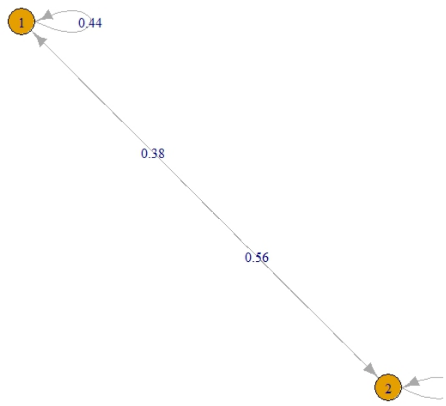


FIGURE 4.3: A two-state problem transition diagram given that the present system state is 1 (increase)

TABLE 4.2: First Passage time probabilities, assuming the system current state is state 2 (decrease) for the next seven days

Day	Increase (State 1)	Decrease (State 2)
1	0.37969760	0.620302400
2	0.23552733	0.211504549
3	0.14609817	0.093689282
4	0.09062505	0.041501148
5	0.05621493	0.018383589
6	0.03487026	0.008143301
7	0.02163010	0.003607204

Table 4.2 provides a summary of the first passage time probabilities (for the next seven days), assuming the system current state is a Decrease (state 2). The first passage times probabilities for the next seven days continue to show some decrease. Practically, for the first and second days, the probability that state 2 will govern the system is approximately 62% and 21.1% respectively.

4.2.8 Two-State Problem Extreme Peaks Modelling

In order to select a sufficiently high threshold, a Pareto quantile plot is based on the daily electricity positive changes. Classifying the states, extreme positive changes are deemed to be all the observations that are above the threshold (state 1) and another state classification is of observations less than the threshold inclusive showing no extreme increase (state 2). A Pareto quantile plot is specified as the scatter plot, consisting of the following points:$(-\log(1-p_i); \log z_i)$ where $p_i = \dfrac{i}{n+1}$ and z_i are daily peak electricity demand changes, for all $i = 1, ..., n$.
As could be seen in Figure 4.4, the tail distribution of the daily peak demand changes' sufficient threshold is exp(7.706613) = 2223 MW.

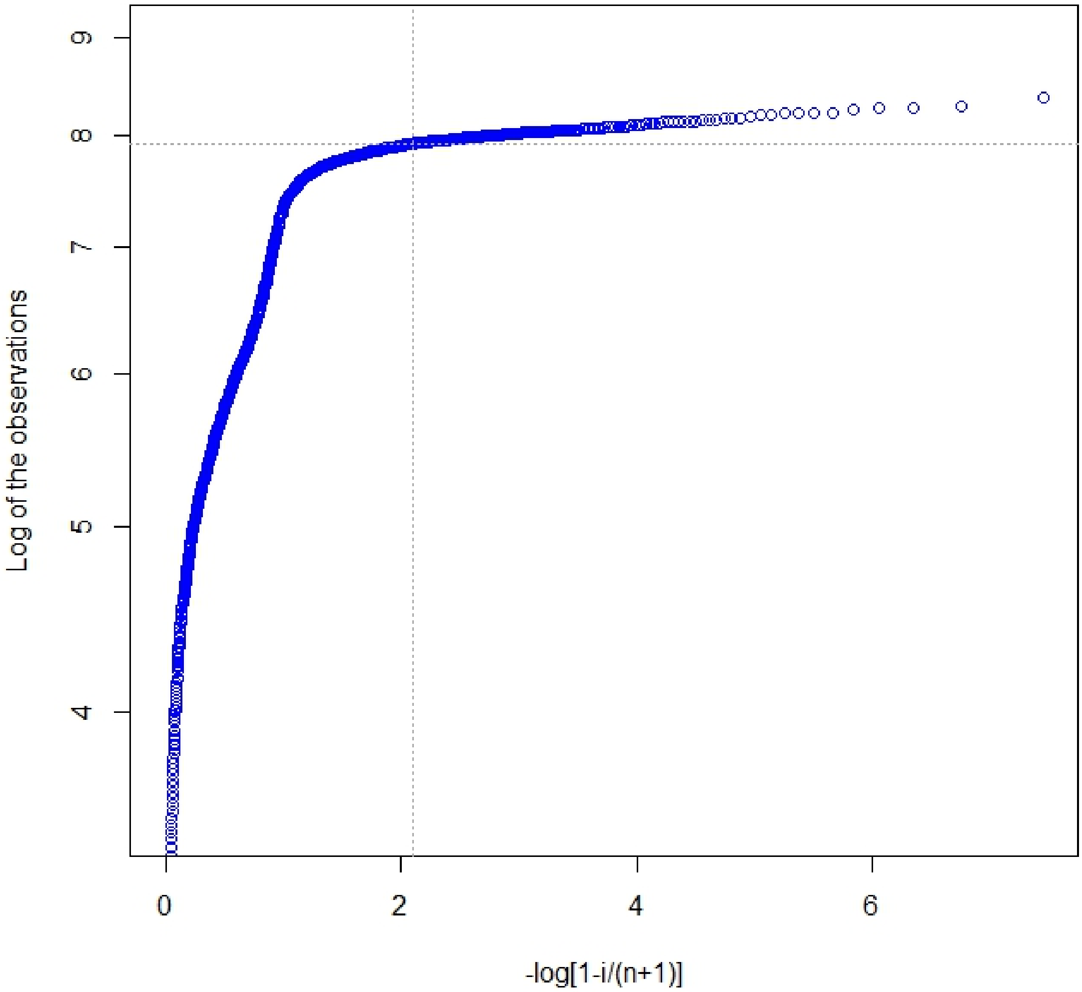


FIGURE 4.4: Daily peak electricity demand changes Pareto quantile plot

We present the transition matrix of the two defined states as:

$$\begin{aligned} &\text{Observations above } \tau : \text{state 1}, \\ &\text{Observations below } \tau : \text{state 2}, \end{aligned}$$

is specified as:

$$P = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} \quad 1 \quad\quad & \quad\quad 2 \end{array} \\ \begin{pmatrix} 0.002415459 & 0.9975845 \\ 0.118999713 & 0.8810003 \end{pmatrix} \end{array}$$

Steady state probabilities of the system are specified as:

$$\pi = (\pi_1, \pi_2) = (0.1065748, 0.8934252)$$

and the mean return times are specified as:
$R_1 = E(T_1) = \dfrac{1}{0.1065748} = 9.38$ days (state 1) and

$R_2 = E(T_2) = \dfrac{1}{0.8934252} = 1.12$ days (state 2).

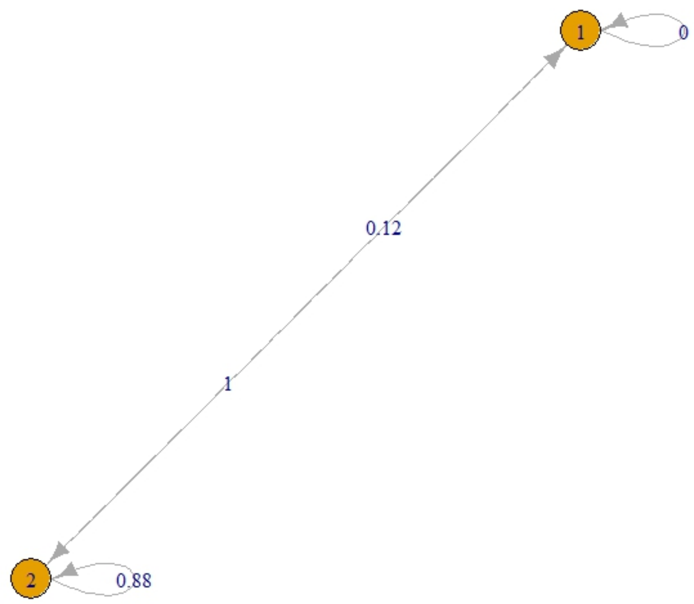


Figure 4.5: Plot of a transition diagram for a two-state Problem when the threshold is 2223 MW

TABLE 4.3: First passage time probabilities assuming the current system state is state 1 ($z_t > 2223$ MW) for next seven days

Day	Extreme increase (State 1)	No extreme increase (State 2)
1	0.002415459	9.975845e-01
2	0.118712269	2.409624e-03
3	0.104585545	5.820349e-06
4	0.092139896	1.405881e-08
5	0.081175276	3.395849e-11
6	0.071515443	8.202534e-14
7	0.063005127	1.981288e-16

Assuming the current system state is "an extreme increase", the first passage time probabilities for the next seven days are shown in Table 4.3. Significantly, it is evident from Figure 4.3 that as the number of days increase the probabilities keep on decreasing.

TABLE 4.4: Probabilities of the first passage time given that state 2 is the present state of the system for an upcoming week ($z_t \leq 2223$ MW)

Day	Extreme increase (State 1)	No extreme increase (State 2)
1	0.11899971	8.810003e-01
2	0.10483878	1.187123e-01
3	0.09236300	2.867446e-04
4	0.08137183	6.926199e-07
5	0.07168861	1.672995e-09
6	0.06315768	4.041051e-12
7	0.05564194	9.760992e-15

Assuming the current system state is "no extreme increase", the probability that the next day we will observe the same "extreme increase" state is approximately 0.11899971, and the probability that we will observe a "no extreme increase" state is approximately 0.8810029. Essentially, there is an approximately 88% probability that we will observe a "no extreme increase" state the next day.

4.2.9 Three-State Problem

We have considered a two-state problem in the previous section. Under the two state problem, we determined the mean return times, first passage times and the steady state probabilities. In this section we will focus on the three state problem, in which there are three states making up the system. Firstly, we categorize the states into two: positive daily changes and negative daily changes. Further, positive daily changes are split into two states. Consequently, we have the following three states:

Observations above τ : (State 1), i.e. $(z_t \geq 2223 \text{ MW})$.
Observations between 0 and τ : (State 2), i.e. $(0 < z_t < 2223 \text{ MW})$,
Observations below zero : (State 3), i.e. $z_t \leqslant 0)$.
The three-state problem transition matrix is given as:

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \end{array}.$$

Below is the transition matrix consisting of the calculated probabilities.

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0.002415459 & 0.5193237 & 0.4782609 \\ 0.077319588 & 0.3376289 & 0.5850515 \\ 0.139956803 & 0.2397408 & 0.6203024 \end{pmatrix} \end{array}.$$

The three-state problem steady state probabilities are given as:

$$\pi = (\pi_1, \pi_2, \pi_3) = (0.1065822, 0.298787, 0.5946308).$$

Three-state problem mean return times are given as:

$$R_1 = E(T_1) = \frac{1}{0.1065822} = 9.38 \text{ days for the extreme increase state } (z_t \geq 2223).$$

The return period is $\simeq$ 9.4 days giving approximately $\dfrac{365}{9.4}$ = 38.8 days in a year of extreme increases.

$$R_2 = E(T_2) = \frac{1}{0.298787} = 3.35 \text{ days (State 2, } 0 < z_t \leq 2223).$$

$$R_3 = E(T_3) = \frac{1}{0.5946308} = 1.68 \text{ days (State 3, } (z_t \leqslant 0).$$

Previously, at the tail distribution the number of observations above the threshold (τ = 2223) was found to be 221. Overall, the data consists of 3894 observations and computing the exceedances proportion we have 0.05418593.

Table 4.5: First passage time probabilities, assuming the current system state is "an extreme increase" ($z_t \geq 2223$) for the next seven days

Day	Extreme increase (State 1)	Small increase (State 2)	Decrease (State 3)
1	0.002415459	0.51932370	0.478260900
2	0.107089761	0.11591306	0.304986329
3	0.106466184	0.10616440	0.122522883
4	0.094579752	0.07369554	0.053660879
5	0.083236500	0.05283866	0.023066929
6	0.073192644	0.03772604	0.009954710
7	0.064355948	0.02695031	0.004292458

Table 4.5 presents the probability distribution of the first passage times probabilities, assuming the current system state is "an extreme increase". So assuming the current system state is "an extreme increase" the probability that we will observe a "small increase" state the next day is 0.51932370. Equally, in order to observe an "extreme increase" state, the probability is 0.002415. The probability that we will observe an "extreme increase (state 1)" is a 0.2% and the probability that we will observe a "decrease (state 2)" is approximately 0.47826.

TABLE 4.6: First passage time probabilities assuming the current system state is a "small increase" ($0 < z_t < 2223$) for the next seven days

Day	Extreme increase (State 1)	Small increase (State 2)	Decrease (State 3)
1	0.07731959	0.33762890	0.585051500
2	0.10798726	0.18041461	0.234509230
3	0.09809608	0.13848964	0.102758511
4	0.08649969	0.09804586	0.044167662
5	0.07607533	0.07011733	0.019061316
6	0.06689164	0.05007920	0.008219177
7	0.05881537	0.03577352	0.003544726

Table 4.6 presents the probability distribution of the three-state problem steady state probabilities assuming the current system state is a "small increase". The probability that we observe an "extreme increase" the next day is 0.07731959. It is highly unlikely that we will observe an "extreme increase" state the next day assuming the current system state is "a small increase". The probability that a "decrease" state is observed is 0.58505 the next day. Practically, the probability that we will observe a "decrease" state assuming the current system state was a "small increase" is 58%.

TABLE 4.7: First passage time steady state probabilities assuming the current system state is a "decrease" ($z_t \leq 0$)

Day	Extreme increase (State 1)	Small increase (State 2)	Decrease (State 3)
1	0.13995680	0.23974080	0.620302400
2	0.10535220	0.22139468	0.207196581
3	0.09123918	0.15355447	0.098906342
4	0.08011351	0.11010864	0.041783319
5	0.07043211	0.07861484	0.018098996
6	0.06192757	0.05616011	0.007798149
7	0.05445047	0.04011626	0.003363701

The probability that we observe a "small increase" assuming the current system state is a "decrease" is 0.23974080. Also, the probability that we observe an "extreme increase" is 0.13995680 the following day. Notably, It is highly unlikely that we will observe an "extreme increase" state assuming the current state is a "decrease".

As represented in Table 4.7, the probability of observing a "decrease" state assuming the current system state is a "decrease" is 0.620302400 the following day. There is a high chance of observing a "decrease" given the current system state is a "decrease" state.

4.2.10 Electricity Demand Data

The frequency analysis of different hours of the day using the historic data of the period 2000 to 2010 is undertaken. As clearly depicted in Figure 4.6 and Table 4.8, high electricity usage in South Africa occur mostly at 19:00 and 20:00. According to the information depicted in Table 4.8, electricity daily usage is 40.5% at 19:00 and 44.7% at 20:00. This information is useful to system operators for both short and long term strategic planning of the national grid, and help ensure there is sufficient energy supply during extreme peak demand periods.

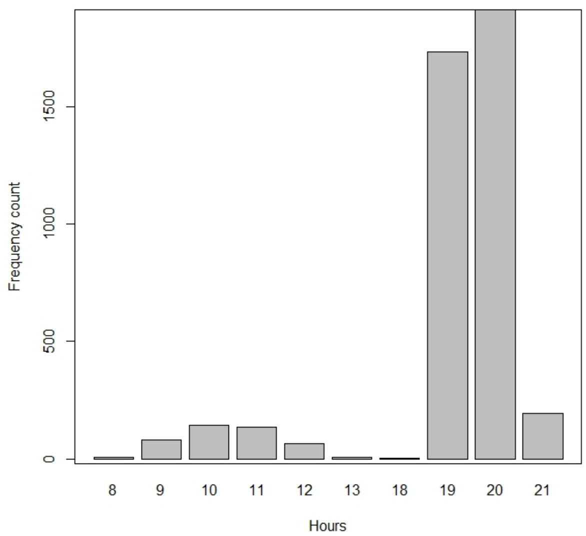


FIGURE 4.6: Bar chart of the daily electricity peak demand occurrences for the period 01 Jan 2000 to 10 Sept 2011

TABLE 4.8: Frequency occurrences of DPED in different hours of the Day

DPED hours	8	9	10	11	12	13	18	19	20	21	Total
Frequency	5	82	142	136	64	5	3	1731	1910	193	4271
Percentage (%)	0.12	1.9	3.32	3.18	1.5	0.12	0.07	40.5	44.7	4.5	100

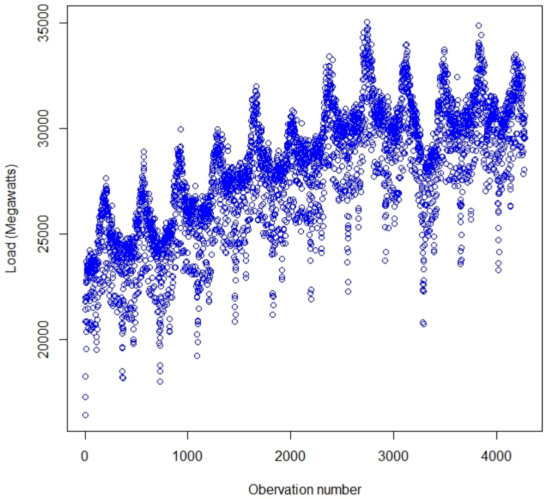


FIGURE 4.7: The daily electricity load consumption for the period 01 Jan 2000 to 10 Sept 2011 at 9:00 a.m.

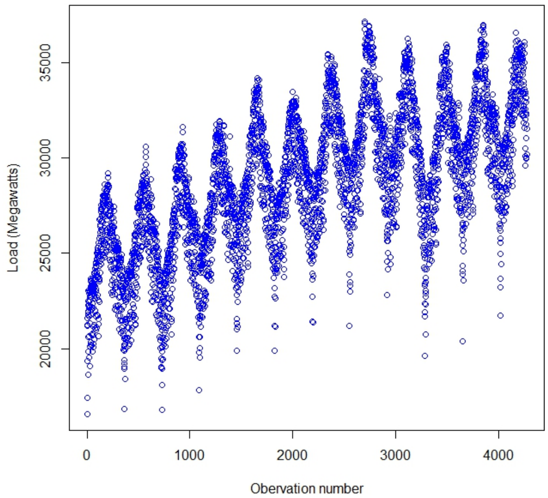


FIGURE 4.8: The daily electricity load consumption for the period 01 Jan 2000 to 10 Sept 2011, at 19:00 p.m.

Figure 4.7 and Figure 4.8 provide evidence that electricity demand is seasonal and there is a clear trend in the data. Figure 4.7 depicts the daily electricity demand in South Africa, at 9:00 in the morning and Figure 4.8 represent the daily electricity demand, at 19:00. It is easy to note that, the 2008 financial crisis impacted the electricity demand pattern in both figures (observations 3500 to 3800), after which it starts to show some smoothness in 2009 (observations 3900 to 4000) (Migon and Alves, 2012).

4.3 Concluding Remarks

The chapter discussed the modelling of daily peak electricity demand in different hours of the day using discrete time Markov chain using South African hourly data of the period 2000 to 2010. Transition probability matrices, steady state probabilities, mean return times and first passage times of the peak daily changes are determined adopting different number of states. In order to determine a sufficiently high threshold we use a Pareto quantile plot. The two-state steady probabilities is 0.1065748 with a 9 days return period of an extreme increase. Equally, the three-state steady probabilities of an extreme increase is 0.1065822 with 9 days return period.

Figure 4.6 depicts some daily electricity peak demand occurrences in South Africa. The 19:00 hours and 20:00 hours experience extreme demand. Since electricity cannot be stored and dependence of consumers on electricity is random, energy forecasting needs careful planning (Hong, 2014). The next chapter develops some regression based short term load forecasting using the 19:00 hours daily demand data for the period 01 Jan 2000 to 10 Sept 2010. The chapter follows the regression framework, focusing on the nonlinear relationship exhibited by logarithmic demand versus various variables that drive demand in South Africa.

Chapter 5

Short Term Load Forecasting

5.1 Introduction

In an effort to ensure an effective operation and scheduling of the power systems, load forecasting is regarded an essential tool (Hyndman and Fan, 2010). Load forecasting is concerned with prediction of future demand for a given forecast time horizon (short, medium and long term). This chapter proposes a probabilistic short term load forecasting model, using South African 19:00 hour demand data for the period 2000 to 2010. Electricity demand patterns differ throughout the 24 hours of the day. Chapter 4 found the hours 19:00 and 20:00 to be the two hours in a day experiencing high electricity usage by consumers. In this chapter we will only focus on the 19:00 hour and the data is for the period 01 January 2000 to 27 September 2010 used to fit and validate short term load forecasting model taking the semi-parametric form. The model investigates whether some nonlinear relationship between the logarithmic load versus variables that drive demand in South Africa exist, and these variables include calendar dummy variables, historical temperature traces and demand (Fan and Hyndman, 2012). The errors from the developed model are assumed to be iid. The developed semi-parametric model is used to forecast the day-to-day 19:00 hour electricity demand, for up to a full week in advance. Despite, the focus being essentially on modelling short term electricity demand at 19:00, it should be noted the same approach can be applied to different hours of the day.

We develop a benchmark model that will be compared to the proposed model in order to accurately assess the better model. To carry this we determine an out-of-sample evaluation to substantiate which model is better than the other. Out-of-sample evaluations are useful as they determine whether the proposed model is potentially forecasting the actual observations accurately.

The chapter is organised as follows: Section 5.2 will cover the scope on the development of the semi-parametric additive model. Section 5.3 will concentrate on the benchmark model development; Section 5.4 gives some comparative analysis between the two models and finally Section 5.5 will concentrate on the general concluding remarks of the whole chapter.

5.2 Models

5.2.1 The Load Forecasting Models

In this scenario, we look at the different features of the models which include electricity demand, predictor variables, dummy variables and calendar effects, temperature effects, interaction effects and errors.

5.2.2 Semi-Parametric Additive Model

Electricity demand is volatile throughout the day and in every hourly period we can obtain proper parameter estimates (Fan and Hyndman, 2012). In this dissertation, we develop a prediction model in order to make forecasts on the 19:00 hour demand. We use the "Heat Degree Days" (HDD_t) including the "Cooling Degree Days" (CDD_t) on day t, reason being we want to investigate how electricity demand is affected by the temperature. We transform the daily electricity demand, we do so by taking natural logarithms. The logarithmic transformation helps to stabilize the volatility in the demand data (Mirasgedis et al., 2006). The proposed model has similar resemblance to the work of Fan and Hyndman (2012) and Sigauke (2014). As an introduction to developing the model, we define the independent variables integrated in the underlying semi-parametric model:

- time t which accounts for the hourly peak electricity demand trend, and Figure 4.8 clearly show some trend and growth attributed to technological advancement and increased electricity consumption by consumers.

- $y_{t,h}$ represent the electricity demand on day t, $t = 1, 2, ..., n$ at hour $h, h = 1, 2, ..., 24$ and in this study we only consider h=19.

- $\bar{y}_t$ represents peak electricity demand average for the past seven days.

- y_t^+ represents the peak electricity demand maxima for the past 24 hours.

- y_t^- represents the peak electricity demand minima for the past 24 hours.

- The "heating degree-days" and "cooling degree-days" (HDD_t and CDD_t respectively) corresponds with day t in which we have observed the hour 19:00 and specified as follows:

$$\text{HDD}_t = \max(T_w - T_t, 0),$$

$$\text{CDD}_t = \max(T_t - T_s, 0),$$

  where T_w, T_s and T_t represent reference temperature and average hourly temperature respectively. T_w and T_s represents some temperature that draws the boundary between what is classified as the summer and winter season of the average day-to-day electricity demand and temperature relationship respectively. The reference temperatures differ from country to country, for example the reference temperature in Australia is 18.5°C (Fan and Hyndman, 2012), Muñoz et al. (2010) use the reference temperatures of 15.0°C and 18.0°C respectively for winter and the summer season. Empirical evidence from a study by Sigauke (2014) shows that for temperature values below 18°C, demand for electricity in South Africa increases significantly while for temperature values above 22°C demand increases slightly. We will use these two temperature values.

- The lagged effects of the "heating degree days" and "cooling degree-days" are specified as HDD_{t-1}, CDD_{t-1}, HDD_{t-2} and CDD_{t-2}

- Past studies have shown a substantial day-to-day seasonal effects in the electricity load series, Mirasgedis et al. (2006); Valor et al. (2001); Muñoz et al. (2010). Essentially, in order to capture such effects, through the dummy variable specifications we incorporate in the model a qualitative variable "*day of the week*". All in all, the dummy variables come to a total number of six excluding only Monday as it is randomly chosen as a base period (Mirasgedis et al., 2006). Any day of the week is appropriate to be used as the base period, without giving the misleading results of the analysis. The dummy variable is introduced as,

$$D_{it} = \begin{cases} 1 & \text{if } i = \text{Tuesday,...,Sunday} \\ 0 & \text{Otherwise,} \end{cases}$$

where t represents the observation number, i.e. $t = 1, 2, .., n$. The index i represents the specific day of the week which are categorised as some dummy variables. They can only take any value in the closed interval $[2, 7]$. D_{it} represents the significant variability of electricity demand at 19:00, is specified with a 1 for correctly locating the t^{th} observation of day i and otherwise D_{it} is 0.

- Past studies confirm that reduction in electricity demand is observable during the holidays. Notably, a study by Ismail et al. (2009) reveal that there is a reduction in electricity demand a day before and after a holiday. For this reason, in order to improve the electricity demand predictions in the proposed model we review events that are not typical such as the holidays including days surrounding them. Practically, during the holidays electricity demand decreases significantly and particularly this is witnessed in the industrial sector (Muñoz et al., 2010). In order to account for the holiday effects that we have in South Africa throughout the year we consider the following dummy variables: H_t, H_{t-1} and H_{t+1} of the reference day t, represented as

$$H_t = \begin{cases} 1 & \text{if } t = \text{a day which is a holiday} \\ 0 & \text{Otherwise,} \end{cases}$$

$$H_{t-1} = \begin{cases} 1 & \text{if } t = \text{a day before a holiday} \\ 0 & \text{Otherwise} \end{cases}$$

and

$$H_{t+1} = \begin{cases} 1 & \text{if } t = \text{a day after a holiday} \\ 0 & \text{Otherwise.} \end{cases}$$

- Electricity demand patterns change throughout the month, and our model takes monthly seasonality effect into account and M_{jt} represents the dummy variable. The base period is randomly chosen to be January, and this means we have eleven dummy variables available to be used.

$$M_{jt} = \begin{cases} 1 & \text{if } j = \text{February,...,December} \\ 0 & \text{Otherwise,} \end{cases}$$

where $t = 1, 2, ..., n$. Assuming that we have identified month j as our observation of concern the M_{jt} equals 1, and otherwise it is identified as zero.

- The differenced Daily peak electricity demand (DifDPED), minimum daily demand (difminDD) and average daily electricity weekly (avADD7) will be incorporated as well in the model. The original load demand data exhibits some strong seasonality, cyclical behavior, a long term trend and is far away from being stationary. In order to make the load data stationary, we difference the original demand data. Differencing means taking the first differences of the original data in order to remove the trend. The first difference of the load data is stationary and completely random.

- u_t is the error term following a SARIMA model. The SARIMA model is given in equation (5.1) with ε_t as a white noise process at time t. $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - \Phi_1 B^7 - \Phi_2 B^{14})(1 - B^7)u_t = \varepsilon_t. \tag{5.1}$$

We consider the following auto-regressive terms in the final model, AR(1), AR(2), AR(7) and AR(14) as a way to account for strong seasonality and reducing the serial auto-correlation that possibly exist in the 19:00 hour daily electricity demand data.

We specify the semi-parametric additive model as,

$$\log(y_{t,h}) = c + \sum_{i=2}^{7} d_i D_{it} + \sum_{j=2}^{12} \lambda_j M_{jt} + \mu H_{t-1} + \delta H_t + \gamma H_{t+1} + \theta \bar{x}_t + \omega x_t^{+} + \kappa x_t^{-}$$

$$+c_1 \mathrm{HDD}_t + c_2 \mathrm{HDD}_{t-1} + c_3 \mathrm{HDD}_{t-2} + c_4 \mathrm{CDD}_t$$

$$+c_5 \mathrm{CDD}_{t-1} + c_6 \mathrm{CDD}_{t-2} + u_t, \quad (5.2)$$

where $c, d_2, ..., d_7, \lambda_2, .., \lambda_{12}, \mu, \delta, \gamma, \theta, \omega, \kappa, c_1, ..., c_6$ are parameters to be estimated. Equation (5.3) is obtained by substituting Equation (5.1) in Equation (5.2).

$$\begin{aligned}
&(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - \Phi_1 B^7 - \Phi_2 B^{14})(1 - B^7)[\log(x_{t,h}) - c \\
&- \sum_{i=2}^{7} d_i D_{it} - \sum_{j=2}^{12} \lambda_j M_{jt} - \mu H_{t-1} - \delta H_t - \gamma H_{t+1} - \theta \bar{x}_t - \omega x_t^{+} - \kappa x_t^{-} - c_1 \mathrm{HDD}_t \\
&- c_2 \mathrm{HDD}_{t-1} - c_3 \mathrm{HDD}_{t-2} - c_4 \mathrm{CDD}_t - c_5 \mathrm{CDD}_{t-1} - c_6 \mathrm{CDD}_{t-2}] = \varepsilon_t.
\end{aligned} \quad (5.3)$$

5.2.3 Model Parameters Estimation

As an estimate the parameters we use the maximum likelihood estimation (MLE). MLE is plausible in terms of retaining the asymptotic efficiency. Out of the 30 predictor variables considered only 14 were selected. The final model is made up of the following variables: the holiday effect, weekly effects and some temperature effects including the heating degree days effects, differenced daily peak electricity demand including the SARIMA model parameter estimates.

A common approach that is used to select the predictor variables is to perform a multiple linear regression and drop all the variables whose *p-values* are greater than the preassigned significance level 5% (Hyndman and Athanasopoulos, 2014). Hyndman and Athanasopoulos (2014) indicate, if multicollinearity exist within the variables then this approach give misleading results. In that instance, we assess the fit of a model by using the cross validation (CV) criterion. To determine the predictive power of the model we use the CV as a useful criterion (Hyndman and Athanasopoulos, 2014). The CV criterion bases the choice of predictor variables on the following; Adjusted R^2 "Akaike's Information Criterion" (AIC), "Corrected Akaike's Information Criterion" (AIC_c) and "Bayesian Information Criterion" (BIC). The following predictor variables are useful in forecasting short term electricity demand at 19:00 hours in South Africa; the day Tuesday, Wednesday, Thursday, Friday and Sunday, month February, holiday effects, differenced DPED, differenced average daily demand (ADD), HDD_{t-1}, HDD_{t-2} and CDD_{t-2}. Table 5.1 presents a summary of the best model CV results. The minimum values of the CV, AIC, AIC_c including the highest Adjusted R^2 value of 90% constitute the best model.

TABLE 5.1: Selection Model Summary

Multiple R^2	Adjusted R^2	CV	AIC	AICc	BIC
0.9049	0.9035	1.3717e+05	2.4400e+04	2.4401e+04	2.4580e+04

5.2.4 Final Model

Table 5.2 shows a summary of the coefficients for variables in the final model. As can be seen, the coefficients of both the holiday effects variables are negative, confirming that electricity demand decreases before, during and after a holiday compared to the normal days. Also, the coefficients of the weekly days, Tuesday, Wednesday and Friday are negative, indicating that electricity demand decreases during such days at 19:00. Further, it can be seen that both the HDD_{t-1} and CDD_{t-2} coefficient values are positive, showing that the effect of previous day's and two days' temperature on the 19:00 hours daily electricity demand is significant. In the end, the final model that is used to make short term forecasts (up to seven days) is given as

$$\begin{aligned}\log(y_{t,h}) = {} & 0.002 + 2.069e-05 * \mathrm{CDD}_{t-2} - 0.007 * \mathrm{DAH} - 0.017 * \mathrm{DBH} + 0.005 \\ & * \mathrm{MFEB} + 0.002 * \mathrm{DFRID} - 0.036 * \mathrm{DH} - 1.139e-06 * \mathrm{DIFADD} \\ & + 2.230e-06 * \mathrm{DIFDPED} - 3.417e-06 * \mathrm{DSUN} + 0.0001 * \mathrm{DTHUR} + \\ & 0.001 * \mathrm{DTUES} + 7.557e-05 * \mathrm{DWED} + 0.003 * \mathrm{HDD}_{t-1} \\ & - 0.002 * \mathrm{HDD}_{t-2} + 0.795 * \mathrm{AR}(1) - 0.0599 * \mathrm{AR}(2) + 0.132 \\ & * \mathrm{AR}(3) - 0.0568 * \mathrm{AR}(4) - 0.458 * \mathrm{SAR}(7) - 0.219 * \mathrm{SAR}(14). \end{aligned} \tag{5.4}$$

As already mentioned the best semi-parametric additive model is selected based on the Adjusted R-squared, AIC, AIC_c and BIC values. Based on results that are summarized in Table 5.2 below, the Adjusted R-squared is 0.674, AIC value is -4.8322 and BIC value is -4.7686. Also, the Durbin-Watson statistic indicate that there is no auto-correlation existing among the residuals. In this study, to determine the efficiency of the model we use the "Root Mean Square Error (RMSE)", "Mean Absolute Percent Error (MAE)" and " Mean Absolute Percentage Error (MAPE)". The out-of-sample validation technique for the best model presents RMSE, MAE and MAPE values as 612.3491, 455.4003 and 1.41% respectively.

TABLE 5.2: Proposed final model parameter estimates

Parameter	Variable	Coefficient	Std. Error	t- statistic	Prob.
c	(Intercept)	0.0015	0.0015	1.032	0.303
d_2	DTUES	0.001	0.001	0.946	0.344
d_3	DWED	7.56E-05	0.001	0.077	0.938
d_3	DTHUR	-0.001	0.001	0.507	0.612
d_5	DFRID	0.002	0.001	1.680	0.093
d_7	DSUN	-3.42E-06	0.001	-0.004	0.997
λ_2	MFEB	0.005	0.006	0.836	0.403
μ	DBH	-0.017	0.004	-4.403	0.000
δ	DH	-0.036	0.006	-6.277	0.0000
γ	DAH	-0.007	0.004	-1.814	0.070
ω	DIFDPED	2.30E-06	8.44E-07	2.721	0.007
θ	DIFADD	-1.14E-06	8.80E-07	-1.294	0.196
c_2	HDD_{t-1}	0.003	0.001	3.548	0.001
c_3	HDD_{t-2}	-0.002	0.001	-3.66	0.0003
c_6	CDD_{t-2}	2.07E-05	0.0002	0.111	0.911
ϕ_1	AR(1)	0.7953	0.0437	18.192	0.000
ϕ_2	AR(2)	-0.060	0.036	-1.685	0.092
ϕ_3	AR(3)	0.132	0.033	4.02	0.001
ϕ_4	AR(4)	-0.057	0.0270	-2.107	0.035
Φ_7	SAR(7)	-0.458	0.036	-12.770	0.0000
Φ_{14}	SAR(14)	-0.219	0.041	-5.367	0.0000
R-squared	0.677		Mean dependent var	-0.001	
Adjusted R-squared	0.674		S.D. dependent var	0.038	
S.E. of regression	0.022		Akaike info criterion	-4.83	
Sum squared resid	0.828		Schwarz criterion	-4.769	
Log likelihood	4411.055		F-statistic	188.351	
Durbin-Watson stat	1.999		Prob(F-statistic)	0.000	

5.2.5 Forecasting results

The training period for the load forecasting model is from January 2005 to January 2010. To evaluate the proposed model performance we used the data for the period February to August 2010.

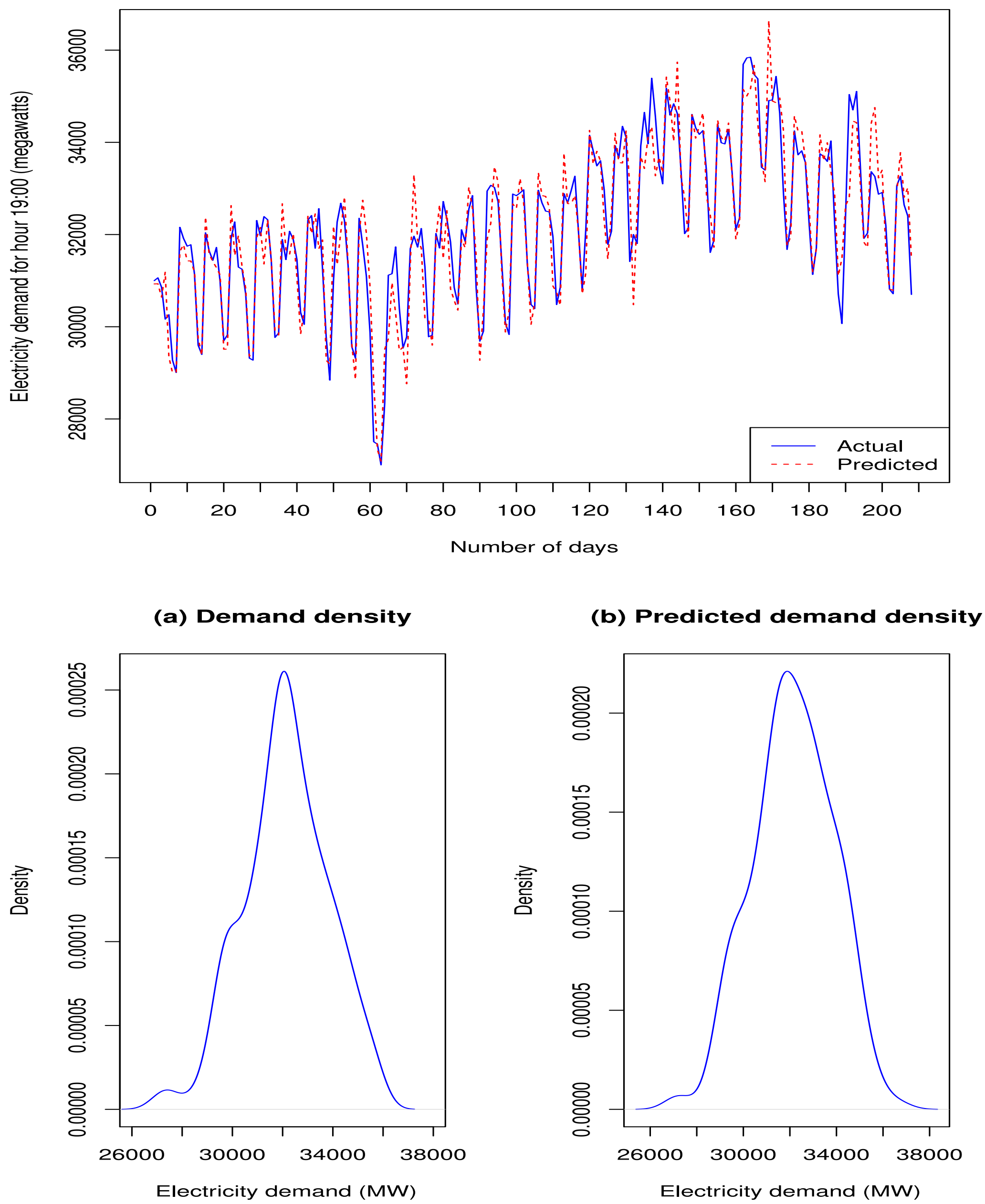


FIGURE 5.1: **Top panel:** Actual and predicted electricity demand in megawatts at hour 19:00 for February to August 2010. **Bottom panel:** (a) Kernel density of the actual electricity demand and (b) Kernel density of the predicted electricity demand at hour 19:00 for February to August 2010.

Figure 5.1 represents the graphical plot of the out-of-sample forecasts for the best semi-parametric additive model for the period February to August 2010. At the end of this chapter in Table 5.4 we find the out-of-sample summary. The bottom panel in Figure 5.1 presents the kernel density representing the density of the actual daily electricity demand at 19:00 and **Bottom right panel** shows the kernel density representing the predicted daily electricity demand at 19:00 hours. Comparatively, it could be seen that both the Actual 19:00 hours electricity demand distribution and the forecasted electricity distribution of the actual 19:00 hours demand distribution have similar features and can be seen in Figure 5.2.

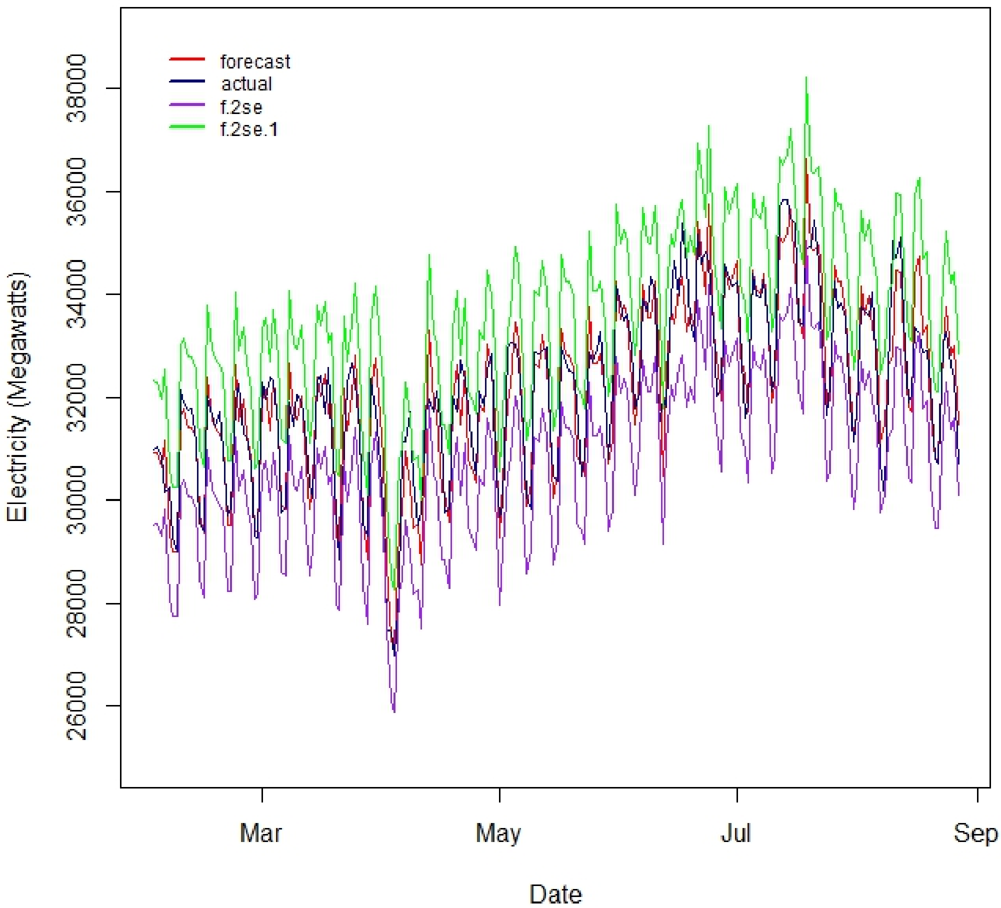


FIGURE 5.2: The 95% prediction interval of the predicted electricity demand at 19:00

Figure 5.2 represents the 95% prediction interval of the electricity demand within which we expect the forecast to lie with a specified probability. As the forecast horizon increases the prediction interval of the forecast increases in length evident from Figure 5.2, see also (Hyndman and Athanasopoulos, 2014).

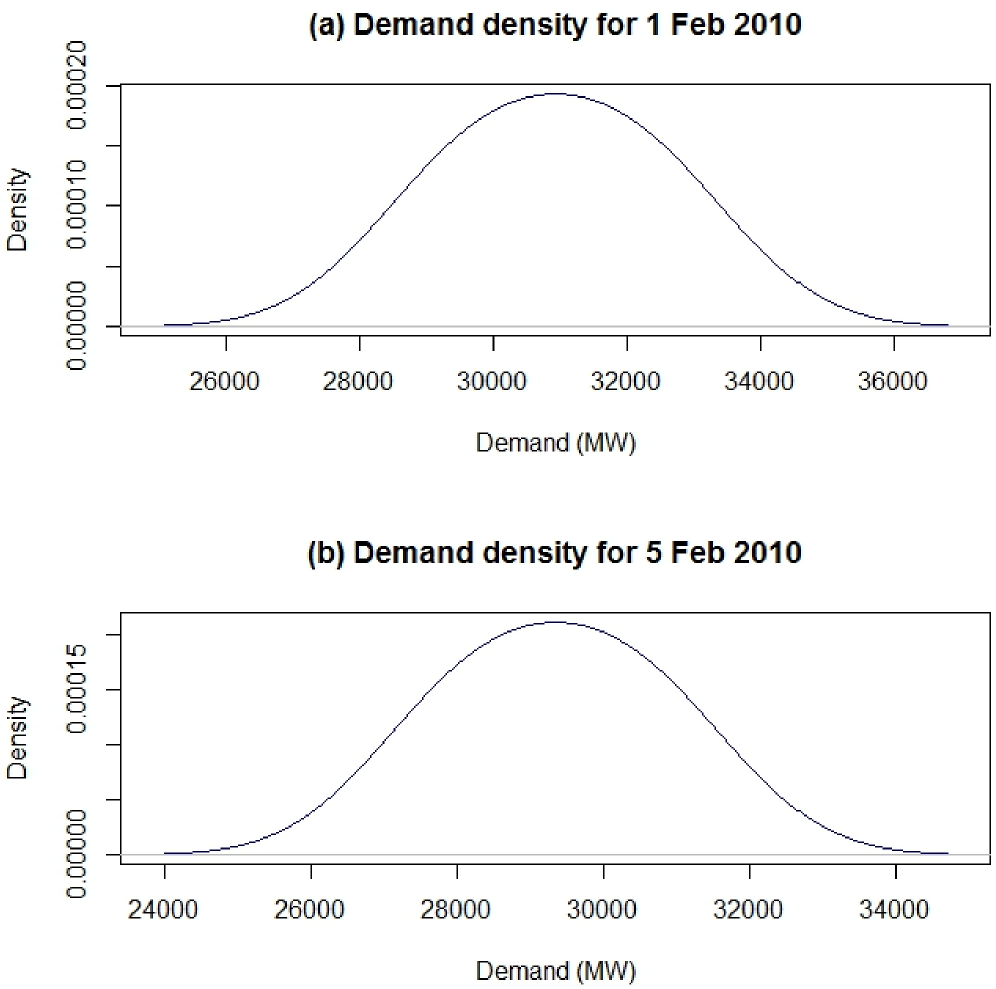


FIGURE 5.3: **Top Panel** (a) Probability density function for 1 February 2010, **Bottom panel** (b) Probability density function for 5 February 2010.

The theoretical demand density for 1 February 2010 represented in Figure 5.3 (a) show that it is highly unlikely that demand at 19:00hrs will be below 26 000MW and highly unlikely that it will be above 36 000MW. It is most likely going to be 30 929.13MW. On the other hand, demand density for 5 February 2010 in Figure 5.3 (b) reveal that that it is highly unlikely that demand at 19:00 will be below 24 000MW and highly unlikely that it will be above 34 000MW.

This information is important for scheduling of electricity, daily system operations and for supply and demand balance of electricity (Taylor, 2012). As was previously stated, short term probabilistic load forecasting is important in modelling electricity. Short term probabilistic models forecast future demand and capture uncertainty of the load forecast and provides a full density of the load forecast (see Figure 5.3). Altogether, this is important to system operators in the dispatching and scheduling of hourly electricity demand particularly during peak period, etc.

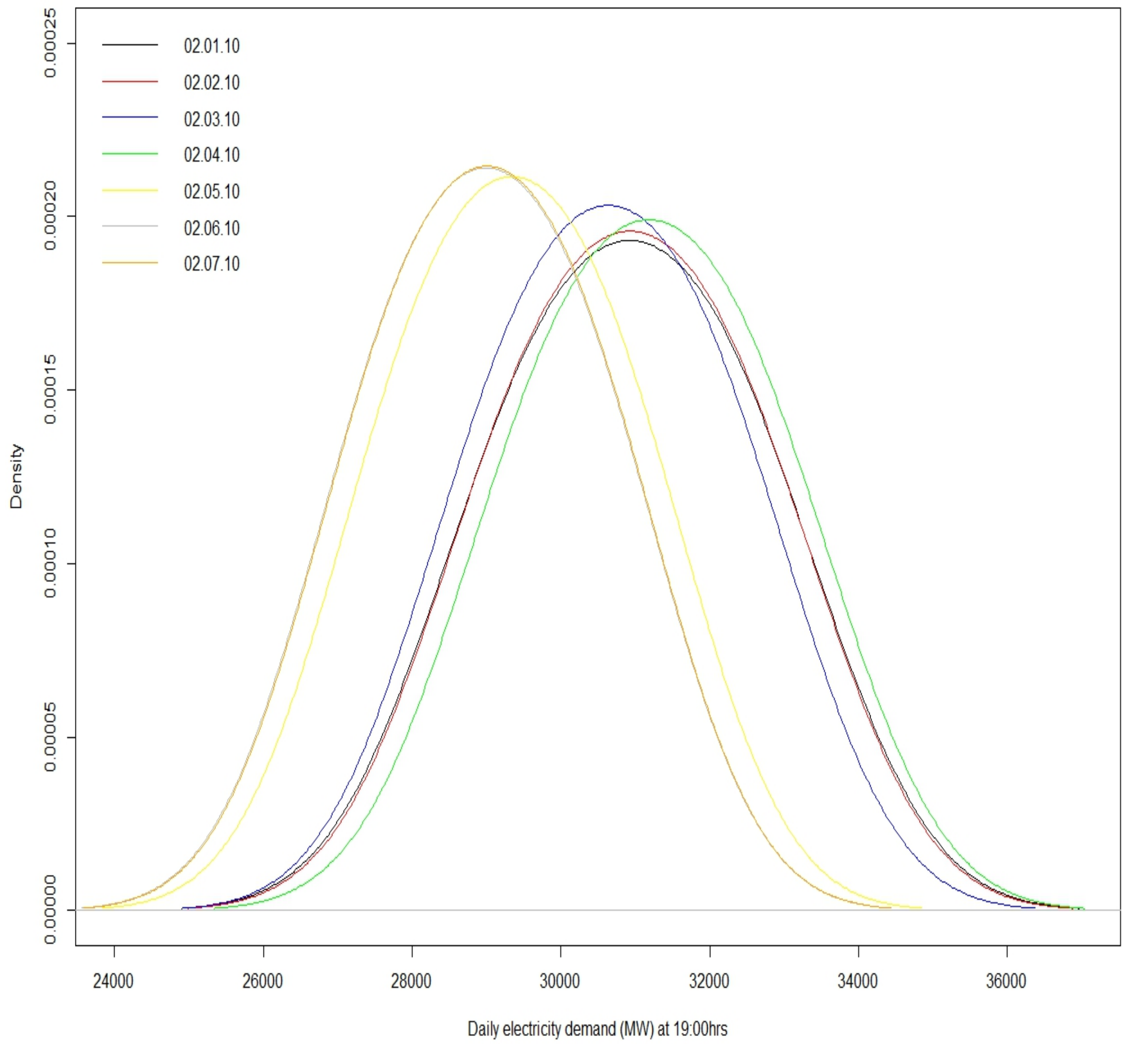


Figure 5.4: Probability density functions of DPED for first seven days of February 2010

Short term load forecasts have increasingly shown to be a huge concern for energy utility companies between hours and weekly days (Cottet and Smith, 2003). Figure 5.4 show demand densities of the first seven days of February 2010 and can be seen that there is demand variability during these seven days. Monday, Tuesday, Wednesday, Thursday, Friday and Saturday were on the following respective dates, 1st, 2nd, 3rd, 4th, 5th, 6th and 7th February 2010. It can be seen that it is highly unlikely that demand at 19:00hrs on Friday, Saturday and Sunday will be below 24 000MW and highly unlikely that it will be above 34 000MW despite some demand variability exhibited. In the same way, it can be seen that it is highly unlikely that demand at 19:00 on Monday, Tuesday, Wednesday and Thursday will be below 26 000MW and highly unlikely that it will be above 36 000MW. Forecasts prove to vary significantly as can be seen in Figure 5.4, because usage patterns are more volatile on different days (Cottet and Smith, 2003). This information serves as a useful analysis in operational decisions of the the power systems.

5.3 Benchmarking Model

The Previous section discussed the fitting of the semi-parametric additive model using the 19:00 hours daily electricity demand short term forecasts. In particular, the developed model, captures variables that drive electricity demand in South Africa such as temperature (°C), calendar effects and lagged demand. Further, as a result of seasonal variation exhibited in electricity demand both hourly, daily, weekly, monthly and annually we account for the residual correlation in the model. The model is useful to system operators during periods of scheduling maintenance, dispatching electricity to end users while ensuring the national grid is not vulnerable to disturbances, load shedding and outages.

If a model gives good forecasts it does not mean the model is highly significant (Fan and Hyndman, 2012). Consequently, a benchmark model is developed, from which we compare the two models. The benchmark model in this study is based on a notable paper by (Hong et al., 2011) with some slight modifications. Hong et al. (2011) highlights that regression analysis has been recognized as a proper and rightful methodology to produce benchmarking models. Subsection 5.3.1 will review the benchmark model.

5.3.1 Benchmark Model

In this subsection the benchmark multiple regression model incorporates the polynomial regressors, the calendar effects, temperature effects and interactions between the predictor variables. The benchmark model used is represented as:

$$Y_i = \beta_0 + \sum_{k=1}^{7} \kappa_k D_{it} * DifH19_{it} + \sum_{k=1}^{3} \alpha_k (18 - T_{it})^k + \eta_t \tag{5.5}$$

where β_0 is the intercept, $\alpha_0, ..., \alpha_2$, $\kappa_0, ..., \kappa_7$ are the parameters, $DifH19_{it}$ is the differenced electricity demand at 19:00, D_{it} represent the weekly dummy variables, T_{it} is the average daily temperature at time t (in °C), η_t is error term which is assumed to follow a SARIMA model if not white noise as in the semi-parametric additive model. The benchmark model is fitted using the following predictor or explanatory variables:

- Temperature is a major variable that affects daily electricity demand patterns throughout the year. This subsection presents two sets of data that will be used in the benchmark model: Temperature and load demand which can further be divided as, average daily temperature (in °C), daily temperature at 19:00 hours (°C), daily peak electricity demand (in Megawatts) and daily electricity demand at 19:00 hours (MW). There are various ways to incorporate the temperature information in the model such as including current hour temperature and previous hours temperature (Hong et al., 2011)

  In Figure 5.5, **Top panel** (a) shows the average daily temperature for the period 2000-2010, (b) shows the daily temperature at 19:00 hours, **Bottom panel** (c) shows the daily peak electricity demand and panel (d) shows the daily electricity demand at 19:00 hours, both exhibiting positive linear trend.

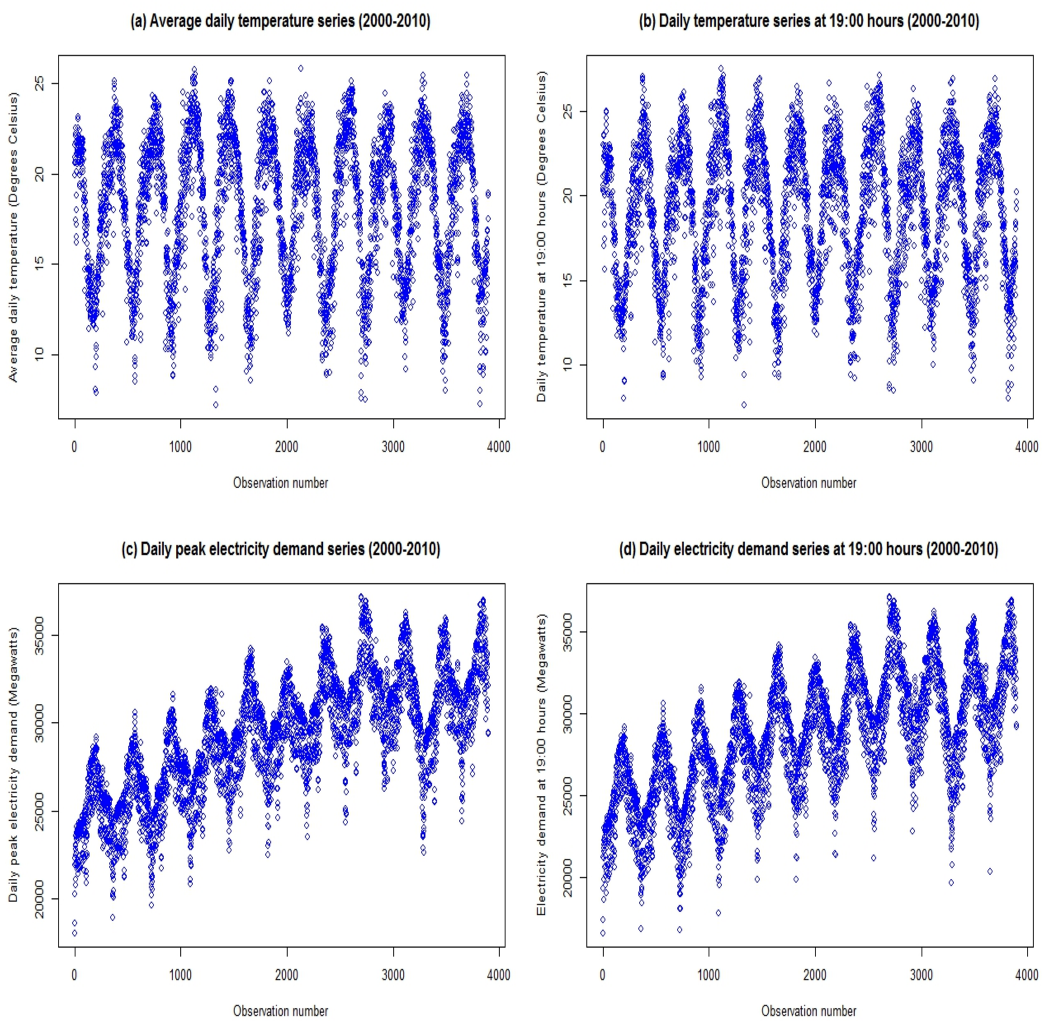


FIGURE 5.5: **Top panel:** (a) Average Daily Temperature series (2000-2010). (b) Daily Temperature series at 19:00 hours (2000-2010). **Bottom panel:** (c) Daily Peak Electricity Demand (2000-2010). (d) Daily electricity demand at 19:00 hours series (2000-2010).

- Temperature

  Over the past years, studies have been extensively conducted to investigate the effect of temperature on the electricity demand. Examples include Muñoz et al. (2010), Hong et al. (2011), Hyndman and Fan (2010) and many others. In Figure 5.6, **Top panel** (a) shows the scatter plot of electricity demand at 19:00 hours against peak temperature at 19:00, (b) shows the scatter plot of average daily temperature versus daily electricity demand at 19:00, **Bottom panel** (c) shows the scatter plot of the average minimum temperature versus daily electricity demand at 19:00, and panel (d) shows scatter plot of the average maximum temperature versus the daily electricity demand at 19:00 hours. The 3^{rd} ordered polynomials of the temperature are used to predict the electricity demand in the benchmark model.

Past studies have shown that the nonlinear relationship between load and temperature is modelled using cooling degree days and heating degree days (Muñoz et al., 2010; Mirasgedis et al., 2006); among others.

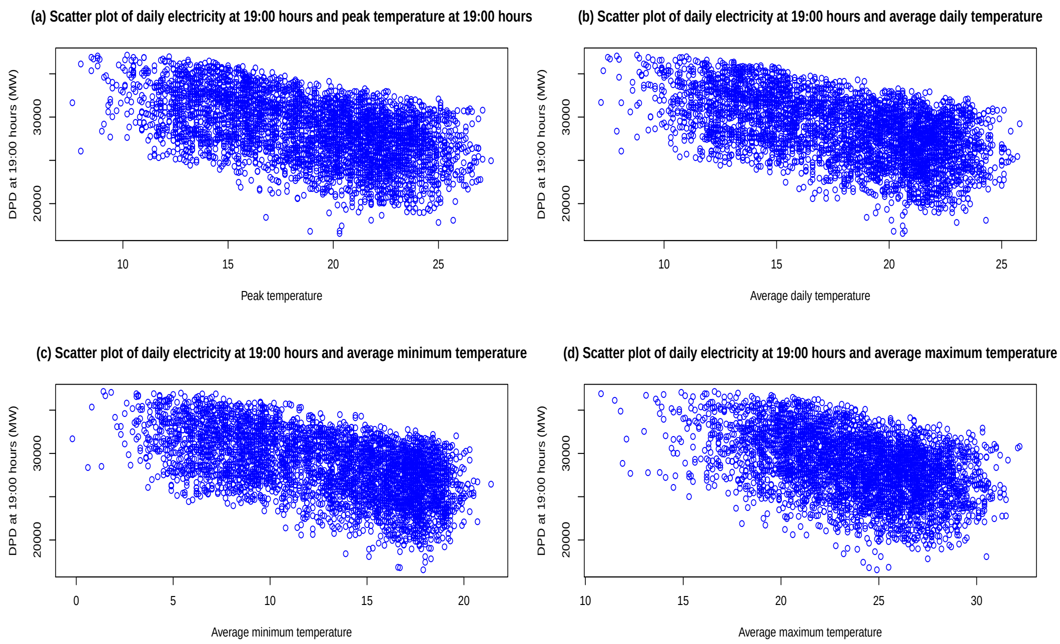


FIGURE 5.6: **Top panel:** (a) Daily peak electricity demand at 19:00 hours and peak electricity demand at 19:00 hours scatter plot. (b) Daily electricity demand at 19:00 hours versus the average daily temperature scatter plot. **Bottom panel:** (c) Daily electricity demand at 19:00 hours versus the average minimal temperature scatter plot. (d) Daily electricity demand at 19:00 hours versus the average maximal temperature scatter plot.

- *Interaction effects*

  Practically, a relationship may arise among two or more variables. The interaction term arises when variables simultaneously influence the outcome variable. In view of short term electricity forecasting, time of the day, month and the temperature effects may have some important implications in electricity demand.

- η_t is the error term which follows the SARIMA model.

5.3.2 Empirical Results on the Benchmark Model

Table 5.3: Benchmark model parameter estimates

Parameter	Variable	Coefficient	Std. Error	t- statistic	Prob.
β_0	(Intercept)	-0.001143	0.001652	-0.691622	0.4893
d_2	Dtues*DifD19	7.81E-06	1.00E-06	7.814705	0.0000
d_3	Dwed*DifD19	6.06E-06	1.38E-06	4.389601	0.0000
d_5	Dfrid*DifD19	1.35E-06	5.18E-07	2.600449	0.0094
d_7	Dsun*DifD19	3.50E-06	1.39E-06	2.518364	0.0119
κ_1	DifD19*AvgTemp	0.000442	0.000414	1.067435	0.2859
κ_2	DifD19*AvgTemp^2	-1.70E-05	9.07E-06	-1.869415	0.0617
κ_3	DifD19*AvgTemp^3	0.000134	5.53E-05	2.427886	0.0153
ϕ_1	AR(1)	0.829062	0.022374	37.05423	0.0000
ϕ_2	AR(2)	-0.098744	0.028794	-3.429299	0.0006
ϕ_3	AR(3)	0.140112	0.028726	4.877491	0.000
ϕ_4	AR(4)	-0.061090	0.022661	-2.695868	0.0071
Φ_7	SAR(7)	-0.494127	0.022523	-21.93891	0.0000
Φ_{14}	SAR(14)	-0.254639	0.022111	-11.51626	0.0000
R-squared	0.66011		Mean dependent var	0.000481	
Adjusted R-squared	0.657937		S.D. dependent var	0.037546	
S.E. of regression	0.021959		Akaike info criterion	-4.792407	
Sum squared resid	0.976965		Schwarz criterion	-4.753833	
Log likelihood	4902.255		F-statistic	302.6836	
Durbin-Watson stat	1.998784		Prob(F-statistic)	0.000000	

Table 5.3 presents the benchmark model summary estimates together with their respective p-values. The interaction between Tuesday, Friday and the differenced demand at 19:00 is significant in predicting electricity demand. Again the interaction between the electricity demand at 19:00 with the 3^{rd} degree polynomial of average daily temperature is significant, influencing the demand at 19:00. The Adjusted R-squared of the model is 0.658, AIC value is -4.792, BIC value is -4.754. The out-of-sample accuracy measures (RMSE, MAE, MAPE) of the benchmark model are 615.326, 458.452 and 1.43% respectively.

5.4 Comparative analysis of the models

In order to determine the best short term forecasting model between the semi-parametric additive and benchmark models, MAPE, MAE and RMSE are normally used (Muñoz et al., 2010; Mirasgedis et al., 2006). The information provided in Table 5.4 summarizes the error measures for out-of-sample evaluation of the models. Comparatively, it could be seen that semi-parametric additive model is the better model among the two models.

TABLE 5.4: Out-of-sample evaluation of the models (validation)

Performance Criteria	Semi-parametric additive model	Benchmark model
RMSE	612.345	615.326
MAE	455.400	458.452
MAPE(%)	1.41	1.43

5.5 Concluding Remarks

This chapter proposed a probabilistic short term load forecasting model using the 19:00 hours daily electricity demand data for the period 2000 to 2010, and is a semi-parametric additive based model. The kernel density is used to estimate the underlying probability density distribution at 19:00 hours in Figure 5.1. (Silverman, 1986). In Figure 5.1, both **Bottom panels** (a) and (b) show that the data is not normally distributed. The model incorporates some nonlinear relationship between electricity demand and predictor variables that drive demand in South Africa. Empirical results show that, the Adjusted R^2 of the model is 0.6882, AIC value is -4.8322 and BIC value is -4.769.

The second part of the chapter discussed the development of the benchmark model, which was compared to the semi-parametric additive model. Empirical results show that the semi-parametric additive model is a better fitting model and will be used to make short term forecasts. In order to foster and assess risk uncertainty we begin to explore the behaviour of the residuals. Modelling of tail behaviour of residuals through the EVT approach is then covered in Chapter 6 to assess risk uncertainty in the forecasts and ultimately evaluate hourly peak event risks.

Chapter 6

Residual Demand Analysis

6.1 Introduction

This chapter reviews the residual demand series generated from the best fitting model using EVT. Primarily, the main concern in this chapter is to model the probability distribution of the correlated residuals. The residuals demand is the forecast error that represents the portion of the demand that has not been entirely captured by the model (Hor et al., 2008). Among techniques that have been used to assess extreme tail residuals include "Auto-regressive Conditional Heteroskedasticity (ARCH)", "Generalized Auto-regressive Conditional Heteroskedasticity (GARCH)" models (Hor et al., 2006) and a normal distribution (Xie et al., 2015). However, in this dissertation we model residual extreme tails using EVT in order to assess the levels of risk in short term load forecasting and evaluate hourly peak event risks.

EVT is a statistical tool that has been used for the past decades to describe the occurrence of rare and unusual events. EVT is largely applicable in areas like finance and insurance, Embrechts (1999), engineering, Castillo (2012), and risk management, McNeil (1999) and in many other areas.

In this dissertation, we focus mainly on the under demand prediction (positive residuals) and less on the over demand prediction (negative residuals). The reason being that under estimation demand may lead to national grid instability, blackouts and load shedding because of unmet demand (Fan and Hyndman, 2012). The chapter discusses an application of GEVD and GPD in modelling the upper tail of the residuals demand distribution.

6.2 Generalized Extreme Value Distribution (GEVD)

In situations when we model the extreme tail distributions EVT is considered a robust framework (Gencay and Selcuk, 2004; Hor et al., 2008). In this section we fit GEVD distribution to a series of demand residuals and also find the extreme residuals return levels and return periods. In what follows, we consider a stationary and a non-stationary model in GEVD framework. The GEVD approach adopted in this dissertation is based on a paper by Hor et al. (2008). The GEVD is the limiting distribution of the Fisher-Tippett theorem (Beirlant et al., 2004a). The Fisher-Tippett Theorem is given as:

Theorem 6.1. *Fisher-Tippett Theorem*
Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables and the common distribution $F(x)$ and $M_n = max(X_1, X_2, ..., X_n)$, If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ for $n \geq 1$, such that

$$Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \rightarrow G(x)$$

as $n \rightarrow \infty$ for a non-degenerate distribution function, then G belongs to the GEVD family of distributions, which include Type I, Type II and Type III, also known as Gumbel, Frećhet and Weibull classes respectively.

In what follows, we briefly discuss the three classes of GEVD. The GEVD is a flexible family of distributions and is given as

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}\right\} \tag{6.1}$$

(Coles et al., 2001) defined on $\left\{x : 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0\right\}$, where $\mu \in \mathbb{R}$ represents the location parameter, $\sigma > 0$ represents the scale parameter and $\xi \in \mathbb{R}$ represents the shape parameter, also known as the Extreme Value index (EVI). The rate of decay distribution F is determined by the EVI sign.

6.2.1 Gumbel Class of Distributions

When $\xi = 0$, the Gumbel distribution is given as:

$$G_{\xi}(x) = \exp\left\{ - \exp\left[1 - \left(\frac{x-\mu}{\sigma}\right)\right]\right\}, \xi = 0. \tag{6.2}$$

As the limit $\xi \to \infty$, the right tail of F decays exponentially with an infinite end point (McNeil and Frey, 2000; Coles et al., 2001).

6.2.2 Frećéhet Class of Distributions

When $\xi > 0$, the Frećéhet distribution is given as:

$$G_{\xi}(x) = \exp\left\{ - \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \xi > 0 \tag{6.3}$$

and then the infinite upper end point of these distributions is heavy tailed (Coles et al., 2001).

6.2.3 Weibull Class of Distributions

The Weibull distribution is given as:

$$G_{\xi}(x) = \exp\left\{ - \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \xi < 0, \tag{6.4}$$

The Weibull class of distributions is characterized by a finite upper end point and it is light tailed (Coles et al., 2001).

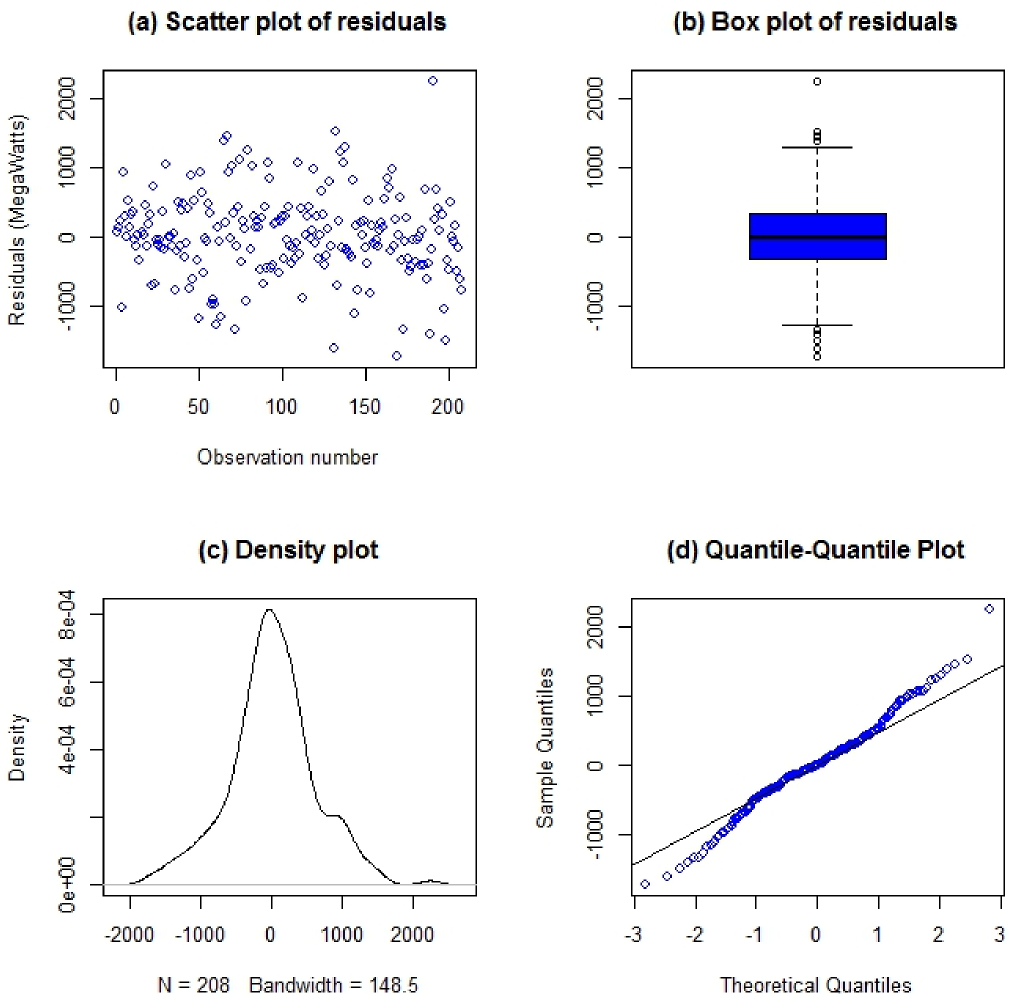


Figure 6.1: (a) Positive and negative residuals scatter plot, (b) Positive and negative residuals box plot, (c) Positive and negative residual density plot and (d) qq-norm plot depicting the positive and negative residuals

In accordance with what Figure 6.1 depicts, the forecast errors cannot be modelled using the normal distribution but can only be described through tail distribution models. The distribution tails are often too long and decays much slow than the normal distribution (Shenoy and Gorinevsky, 2014). We will then fit the GEVD to a series of demand residuals. Assuming that the series of hourly demand residuals are independent and identically distributed (i.i.d) random variables, with the unknown distribution function; $F(x) = Pr\{X_n \leq x\}$. The sample size of the residuals is then divided into b large blocks of equal size. Then, the cumulative distribution function (C.D.F) of $M_n(x)$ is

$$Pr\{M_n(x) < x\} = Pr\{X_1 \leq x, ..., X_n \leq x\} = \prod_{i=1}^{n} F(x_i) = \{F(x_i)\}^n \tag{6.5}$$

and the exceedance probability is given as

$$Pr\{M_n(x) > x\} = 1 - \{F(x_i)\}^n. \tag{6.6}$$

6.2.4 Model Choice and Parameter Estimation

It is very important to consider selecting the best model by looking at the simplicity of the model possible. The model should explain as much of the variation within the data as possible (Coles et al., 2001). The four GEVD models (1 stationary model and 3 non-stationary models) considered are as follows:

- Model 1: stationary model (μ, σ and ξ are constants).
- Model 2: non-stationary model ($\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2$, σ and ξ are constants).
- Model 3: non-stationary model ($\mu(t) = \beta_3 + \beta_4 t$. $\sigma = \exp(\sigma_0 + \sigma_1 t)$, $\xi(t) = \xi$)
- Model 4: non-stationary model ($\mu(t) = \beta_5 + \beta_6 t$, $\sigma(t) = \sigma$, $\xi(t) = \xi$).

Model 1 assumes that all the three parameters do not depend on time, Models 2, 3 and 4 assume non-stationarity. The maximum likelihood estimation (MLE) along with the GEVD log likelihood represented in Equation (6.7) below will be used to estimate the parameters:

$$l(\mu, \sigma, \xi : \mathbf{x}, t) = -m\log\sigma - \left(\frac{1}{\xi} + 1\right)\sum_{i=1}^{m} \log\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]_+ - \sum_{i=1}^{m}\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}} \tag{6.7}$$

where m is the number of block maxima $x_1, x_2, ..., x_m$.

6.2.5 Parameter Estimation

The R statistical package *ismev* by Stephenson (2006) is used for obtaining the parameter estimates for both stationary and non stationary models.

6.2.5.1 Model 1

Fitting the model without trend the parameters ξ, μ and σ are given in Table 6.1 with their respective standard errors as:

Table 6.1: Parameter estimates for stationary GEVD Model (Model 1)

$\hat{\mu}(se)$	$\hat{\sigma}(se)$	$\hat{\xi}(se)$	-log L
-222.1462(45.645)	614.5153(30.203)	-0.2137(0.025)	1633.423

Since $\hat{\xi} < 0$, this belongs to the Weibull class of distributions and the right endpoint is finite. The right end point is

$$\hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} = -222.1462 - \frac{614.5153}{-0.2137} = 2653.4518.$$

The Quantile-Quantile (QQ) and Probability-Probability (PP) plots given in Figure 6.2 show that the Weibull class fits the residuals well. Further, the 95% parameter confidence interval for the parameter estimates are presented in Table 6.2.

Table 6.2: 95% Confidence intervals of Model 1 parameter estimates

$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
(-311.610, -132.683)	(555.317, 673.714)	(-0.262, -0.166)

6.2.5.2 Model 2

Now fitting the GEVD model to allow some quadratic trend in μ, we have the parameter estimates as:
$\hat{\beta}_0 = -431.976(139.0899)$ $\hat{\beta}_1 = 5.495(2.972)$ $\hat{\beta}_2 = -0.027(0.0135)$ then the parameter estimates of Model 2 are given below with standard errors in parentheses:

$$\hat{\mu(t)} = -431.9763 + 5.4947t - 0.02715t^2,$$

$$\hat{\sigma} = 585.534(27.234),$$

and

$$\hat{\xi} = -0.190(0.0231).$$

The negative log-likelihood value of Model 2 is 1634.081. The Residual Probability (RP) and Residual-Quantile (RQ) Plots given in Figure 6.3 illustrate the goodness of fit for Model 2.

We use the deviance statistic defined below to compare Model 1 and Model 0

$$D = 2\{l_1(M_1) - l_0(M_0)\},$$

where $l_1(M_1)$ and $l_0(M_0)$ are defined as the "maximised log-likelihood" respectively under the alternative model (M_1) and null model (M_0). Comparing the stationary model (null model) and the non-stationary (alternative model) we have

$$D = 2\{1634.081 - (1633.423)\} = 0.658,$$

which is small compared to the Chi-squared test, $\chi_1^2(0.05) = 3.841$. There is insufficient evidence to conclude that allowing for a quadratic dependence does improve on the stationary model (Model 1).

6.2.6 Diagnostic Plots for GEV Models (Model 1 and Model 2)

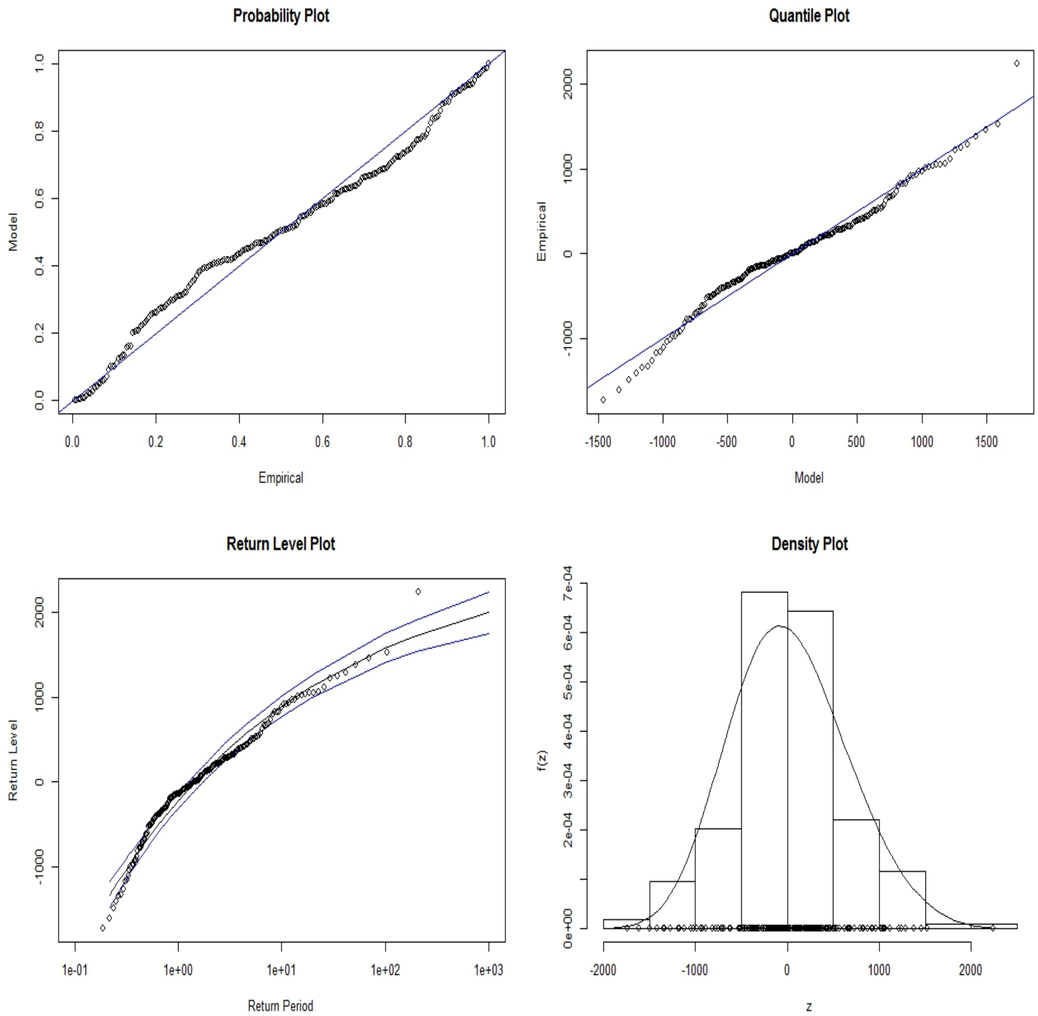


FIGURE 6.2: Model diagnostics for Model 1 demonstrating both the positive and negative residuals applied to the GEVD, (a) Top left panel: Probability plot, (b) Top right panel: Quantile plot, (c) Bottom left panel: Return level plot and (d) Bottom right panel: Density plot

Figure 6.2 confirms that stationary GEVD model (Model 1) is best fitted as the points lie on the unit diagonal. A visual inspection of the plots in Figure 6.2 show that most of the residual observations lie approximately within desired unit diagonal.

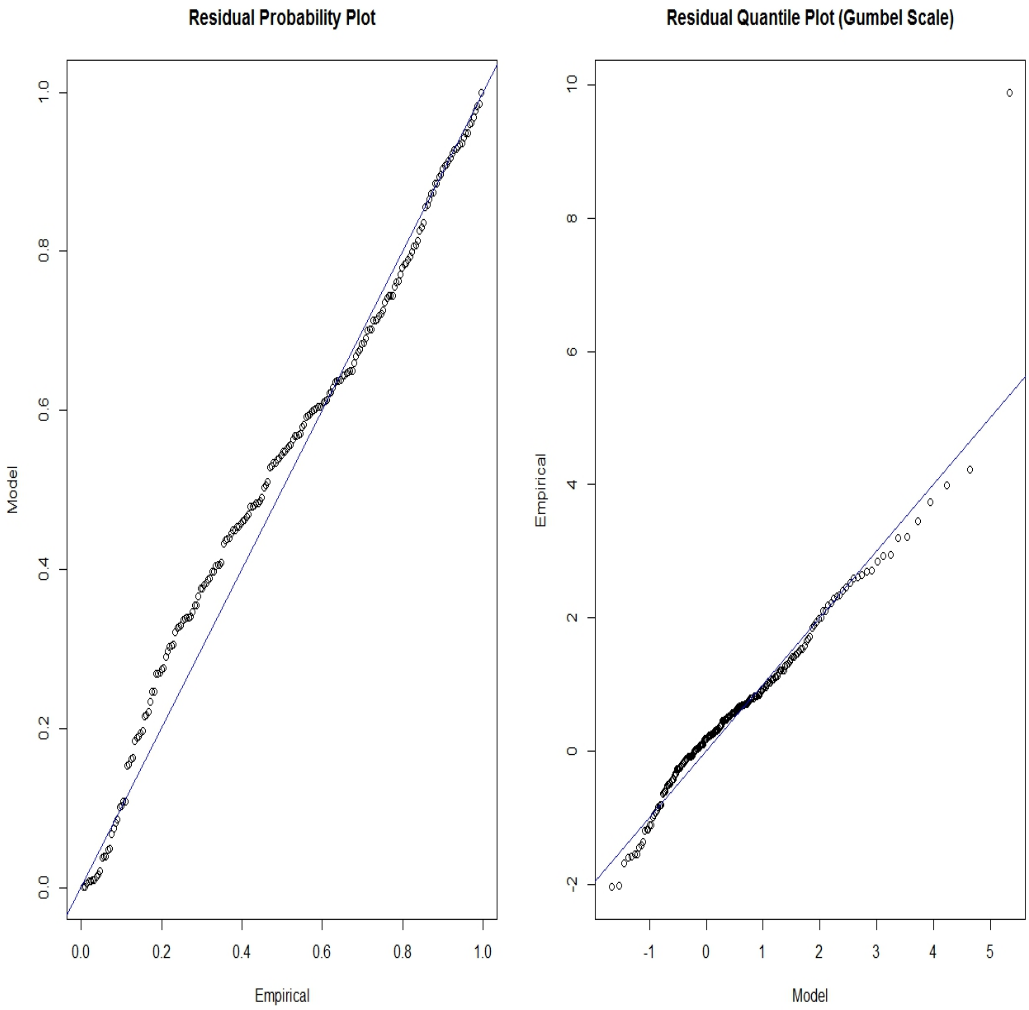


FIGURE 6.3: GEVD diagnostic plots for goodness of fit, allowing quadratic trend in μ.

From Figure 6.3 it is evident from the two plots that the non-stationary GEVD model (Model 2) allowing for a quadratic trend is fairly adequate for the residuals data.

6.2.6.1 Model 3

The GEVD was applied on the residuals data, allowing for a linear trend in the σ and μ yielding the following parameter estimates (with standard errors in parentheses), $\hat{\beta_3} = -118.1351(77.8731)$, $\hat{\beta_4} = -0.8902(0.7163)$, $\hat{\sigma_0} = 488.0182$ and $\hat{\sigma_1} = 1.1936$, then

$$\hat{\mu(t)} = -118.1351 - 0.8902t, t = 1, ..., 208$$

$$\hat{\sigma(t)} = \exp(488.0182 + 1.1936t)$$

$$\hat{\xi} = -0.2401$$

with the negative log-likelihood as 1629.978. Thus, comparing to the stationary model (model 1), we get

$$D = 2\{1629.978 - (1633.423)\} = -6.89,$$

which is small relative to $\chi_1^2(0.05) = 3.841$. Thus, it can be seen that the linear trend in the log scale parameter and location parameter is not evident.

6.2.6.2 Model 4

A GEVD model on the residuals, allowing for a linear trend on μ, the parameter estimates with standard errors in parentheses $\hat{\beta}_5 = -277.8901$, $\hat{\beta}_6 = 0.1665$

$$\hat{\mu(t)} = -277.8901 + 0.1665t, t = 1, ..., 208$$

$$\hat{\sigma} = 612.9777(31.3879)$$

$$\hat{\xi} = -0.2086(0.0272)$$

with the negative log-likelihood value as 1634.067. The Deviance statistic is

$$D = 2\{1634.067 - (1633.423)\} = 1.288,$$

which is small relative to $\chi_1^2(0.05) = 3.841$. As can be seen, the linear trend in μ is not evident.

6.2.7 Diagnostic Plots for GEVD Models (Models 3 and 4)

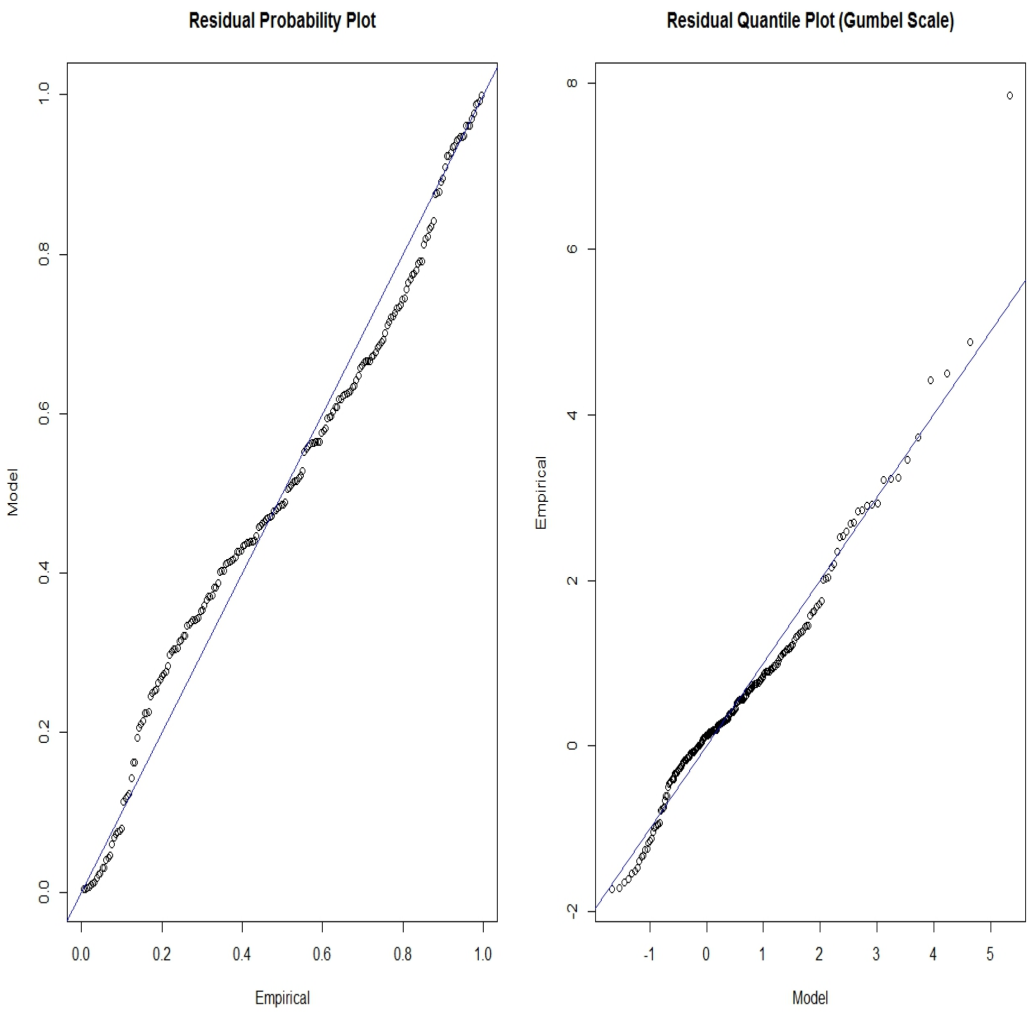


FIGURE 6.4: GEVD diagnostic plots for model 3, allowing some linear trend in μ and σ

.

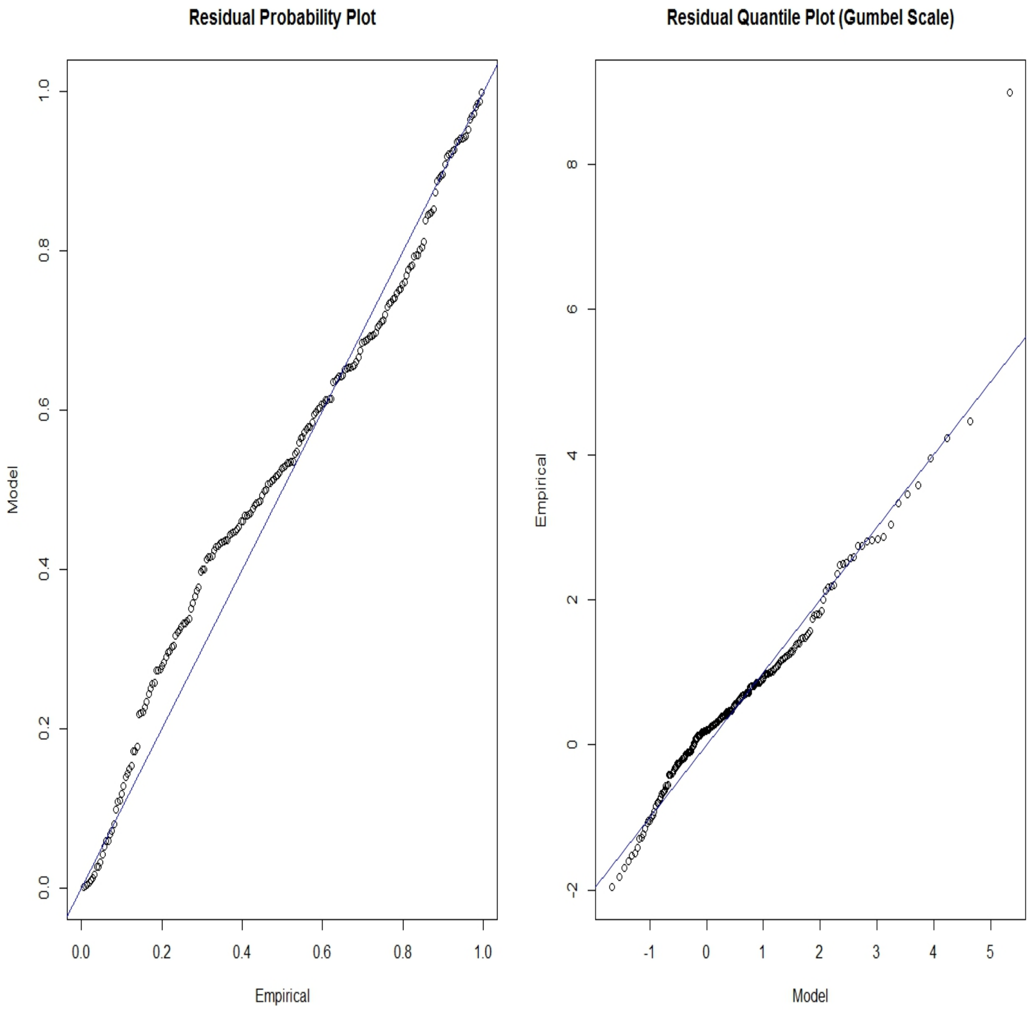


Figure 6.5: GEVD diagnostic plots for model 4, allowing some linear trend in μ

Figure 6.2 presents a visual inspection of the diagnostic plots of the four models. It could be seen that Figure 6.2 shows out most of the residual observations for both the QQ and PP plots approximately on the unit diagonal to those in Figure 6.3, 6.4 and 6.5.

6.2.8 Return Level and Return Period Estimation

We have modelled the upper tail of residuals by fitting a GEVD model in order to make inferences. We then use such models to calculate the N-year return levels. The inverse of GEVD function represents the quantile function (return levels). We represent the return level of $1/p$ period as,

$$z_p = \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\xi}}[1 - [-\log(1-p)]]^{-\xi}, & \text{if } \xi \neq 0 \\ \hat{\mu} - \sigma\log[-\log(1-p)], & \text{if } \xi = 0. \end{cases} \tag{6.8}$$

Further, we explore the return levels of the stationary GEVD model and their respective confidence intervals in this section. Evidently, as the return periods increase the residuals return levels increase as well. Moreover, we can seen that as the return period increase the confidence intervals (CI) increase as well. Prior to information depicted in Table 6.3 one would expect, for example, that extreme residuals generated from the best fitting model will exceed almost 763.96 for every decade on average, and will exceed almost 1436.15 for every century on average. The 95% confidence intervals for the return level are 875.74-1001.62 and 1577.67-1797.67, respectively.

Table 6.3: 95% confidence intervals, return levels for several return periods for stationary GEVD.

Return Period (year)	Return Level	Lower Bound	Upper Bound
10	763.964	875.742	1001.602
100	1436.151	1577.665	1797.666
250	1614.447	1769.622	2041.290
500	1728.994	1891.533	2204.729
810	1799.288	1966.273	2303.208
1000	1827.012	1996.570	2344.715

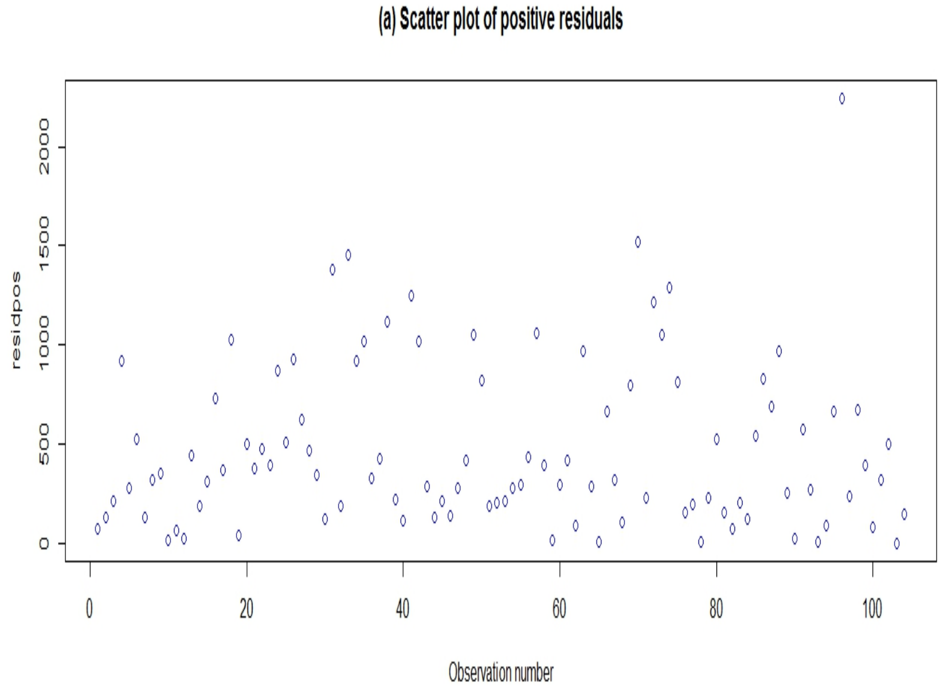
6.3 Generalized Pareto Distribution (GPD)

Past studies have shown that fitting the GEVD model to a data set discards relevant information (Gilleland and Katz, 2006; Balkema and De Haan, 1974) among others. The GPD was first introduced by Pickands III (1975) as exceedances approximate distribution of a random variable X. The random variable is above a sufficiently high threshold τ. This distribution is normally called the "Peaks Over Threshold (POT)" approach. As the threshold tends to approach the right end point, the exceedances distributions converge to the GPD Pickands III (1975). It is characterised by the following cumulative distribution function (Matthys and Beirlant, 2003);

$$G_{\xi}(x) = \begin{cases} 1 - \Big[1 + \xi\Big(\frac{x-\tau}{\sigma_{\tau}}\Big)\Big]_{+}^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ \\ 1 - \exp\Big[-\Big(\frac{x-\tau}{\sigma_{\tau}}\Big)\Big]_{+} & \text{if } \xi = 0, \end{cases} \tag{6.9}$$

where $x > 0$, $\sigma_{\tau} > 0$, $\left[1 + \xi\left(\frac{x-\tau}{\sigma_{\tau}}\right)\right] > 0$ and σ_{τ} is the scale parameter.

In order to determine the GPD tail distribution the shape parameter ξ is deemed an important tool (Beirlant et al., 2004a). The GPD takes different forms for different values of ξ. If $\xi < 0$, then GPD has a truncated, extreme Weibull-type tail, if $\xi > 0$, then the GPD transforms into a Pareto type tail, and also if $\xi = 0$, then the GPD transforms into a Gumbel tail (Beirlant et al., 2004a). After determining some sufficiently high thresholds the GPD will be fitted to both the positive and negative residuals shown in Figure 6(a) and 6(b). This will enable us to estimate extreme electricity demand under and over predictions at 19:00hours. More importantly, in some sections only positive residuals will be undertaken. However, the negative residuals can be used as well when using similar approaches.



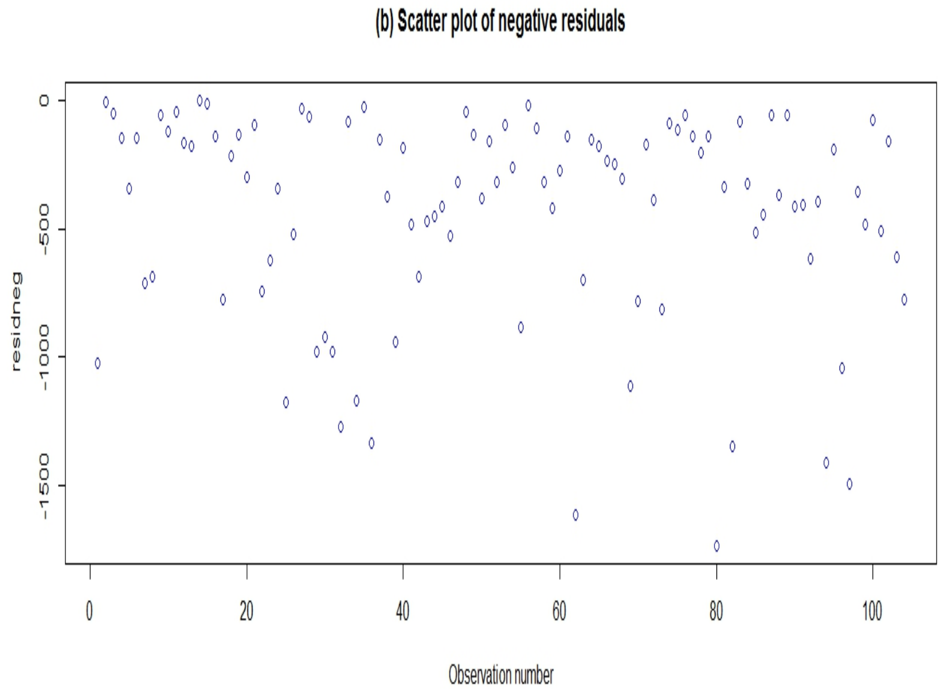


Figure 6.6: (a) Positive residuals plot (b) Negative residuals plot

6.4 Threshold Selection

6.4.1 Introduction

Threshold selection is an area of ongoing research in literature (Hu, 2013). There is always a trade-off between the model bias and its variance during the threshold selection Coles et al. (2001). It means selecting a higher threshold may discard too much data leading to estimates possessing high variance. In contrast, the bias of the estimates result from selection of a very low threshold (Gilleland and Katz, 2006; Coles et al., 2001). The model bias is error from invalid assumptions in the model. In this dissertation we use a variety of threshold selection methods.

6.4.2 Pareto Quantile and Mean Excess Plots

Various graphical plots are commonly used for choosing a threshold. Coles et al. (2001) outlined three graphical plots for threshold selection which include the parameter stability plot and mean residual plot. The Pareto quantile plot and more general model fit diagnostics plots are discussed in Beirlant et al. (2004a). The drawback of using all these methods, is that uncertainty surrounding the model is not taken into account (Scarrott and MacDonald, 2012). The Pareto quantile plot given in Figure 6.7 is used to obtain the threshold on the positive residuals and negative residuals. The sufficiently high threshold is $\tau = \exp(6.769899) = 871.22$ and there are 19 exceedances.

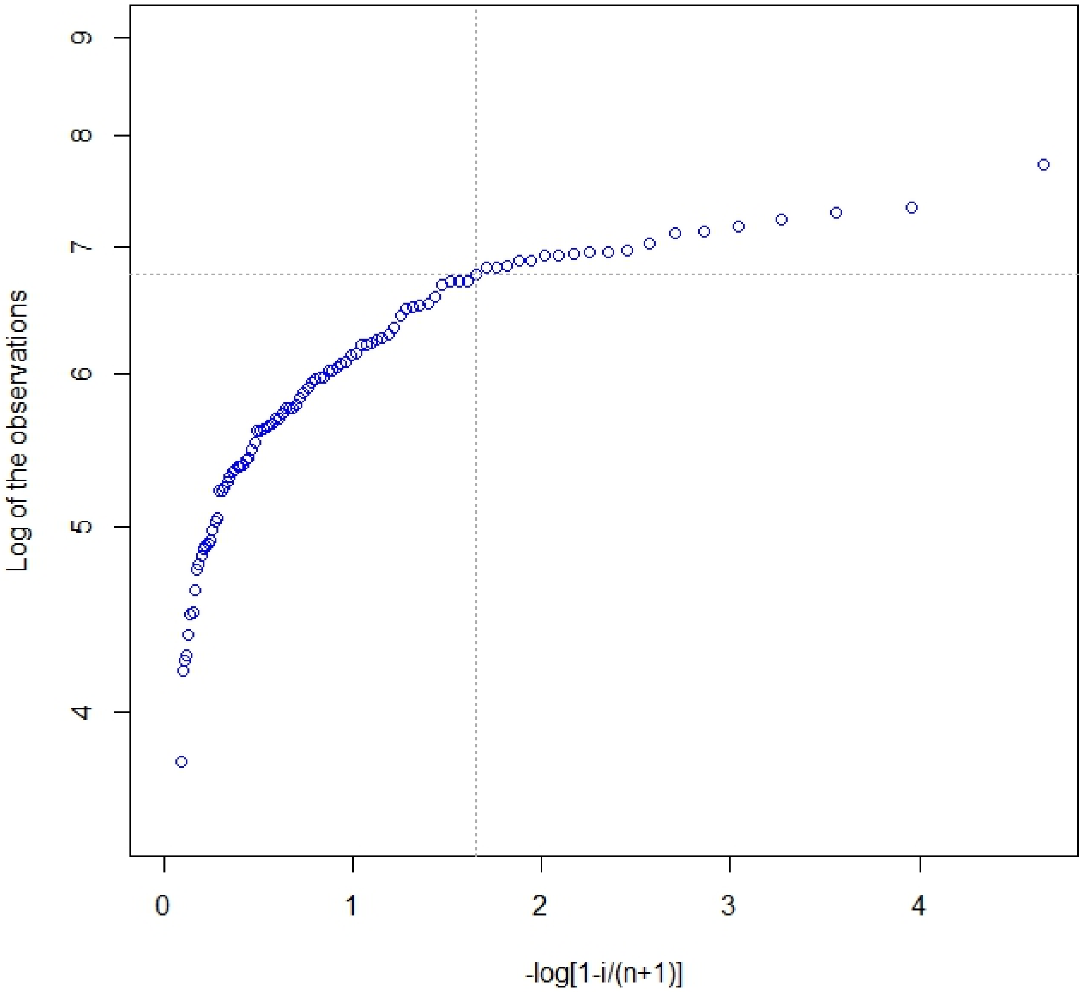


FIGURE 6.7: Positive residuals Pareto quantile plot

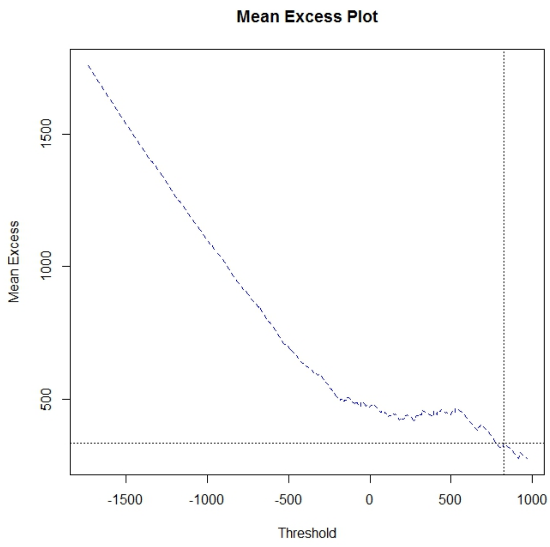


FIGURE 6.8: Positive residuals mean excess plot

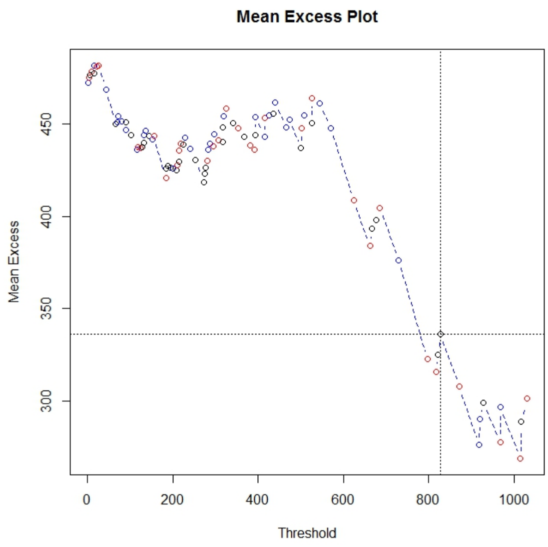


FIGURE 6.9: Negative residuals mean excess plot

In accordance with what Figure 6.8 depicts, to obtain the positive residuals threshold we use the mean excess plot and we have determined the threshold as $\tau = 827.4472$ with 20 exceedances. Similarly, Figure 6.9 depicts the mean excess plot being used on the negative residuals, and the estimated threshold is approximately $\tau = -814.0311$ and there are 86 exceedances. In what follows, we use the extreme value mixture models in the next section to estimate a sufficiently higher threshold.

6.4.3 Extreme Value Mixture Models

Scarrott and MacDonald (2012) used the extreme value mixture models to select the threshold. The "extreme value mixture models simultaneously describes the bulk of the distribution and the tail, containing the threshold as a parameter" (MacDonald et al., 2011). The evmix package in R by Hu et al. (2013) is used to implement extreme value mixture models and other threshold diagnostics. In order to assess model fit, diagnostic plots are used.

To approximate the entire data distribution function fully we use extreme value mixture models (Scarrott and MacDonald, 2012). The mixture models for the bulk distribution are made up of three classifications, which include the parametric, semi-parametric and non-parametric estimators (Scarrott and MacDonald, 2012). In this dissertation, we will only cover two existing parametric form models; bulk model based tail fraction approach and parameterised tail fraction approach.

The parametric mixture models consist of the parametric model below the threshold and a threshold tail model above the threshold. The extreme mixture models distribution function of the parametric form of the bulk component can be defined as (Hu et al., 2013):

Bulk Model Based Tail Fraction

$$F(x|u,\theta,\sigma_u,\xi,\phi_u) = \begin{cases} H(x|\theta), & x \leq u \\ H(u|\theta) + (1 - H(u|\theta)) \times G(x|u,\sigma_u,\xi), & x > u. \end{cases} \tag{6.10}$$

Parameterised Tail Fraction

$$F(x|u,\theta,\sigma_u,\xi,\phi_u) = \begin{cases} (1-\phi_u) \times \dfrac{H(x|\theta)}{H(u|\theta)}, & x \leq u \\ (1-\phi_u) + \phi_u \times G(x|u,\sigma_u,\xi), & x > u \end{cases} \tag{6.11}$$

where ϕ_u is the exceedances proportion and $0 < \phi_u < 1$. $H(.|\theta)$ could be a parametric distribution such as gamma, Weibull or normal distribution, and θ is the bulk distribution parameter vector. $G(.|u,\sigma_u,\xi)$ represents the GPD distribution function where σ_u is the scale parameter, u represents the threshold and ξ is the shape parameter.

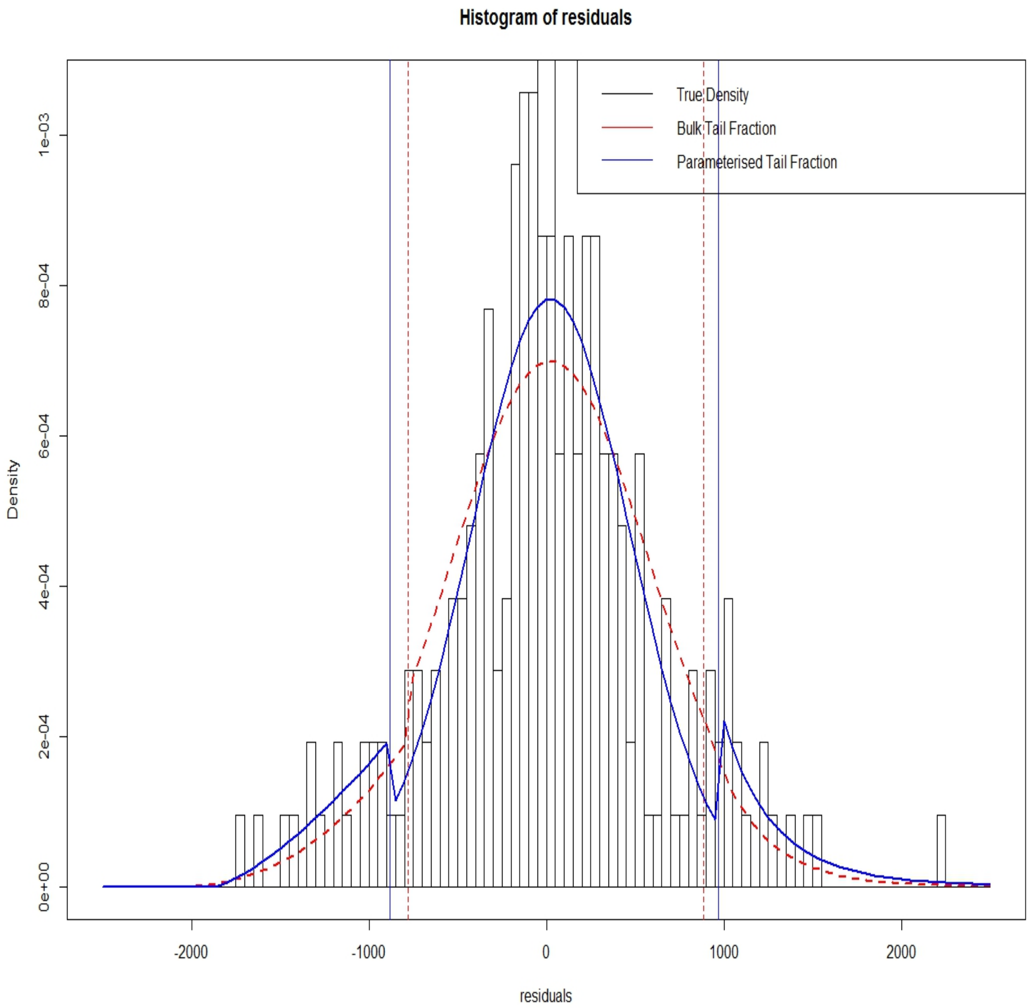


FIGURE 6.10: Density function of the extreme value mixture model (parametric) with bulk model based tail fraction approach with parameter vector and parameterised tail fraction approach with parameter, where the vertical dashed lines indicate the thresholds for the two different approaches.

Figure 6.10 shows a two tail mixture model, where the normal distribution (GNG) describes the bulk and the GPD has been fitted to both tails separately (Hu et al., 2013). Where μ= mean , β= standard deviation, u_l= scalar lower tail threshold, σ_{u_l}=scalar lower tail GPD scale parameter (positive), ξ_l= scalar lower tail GPD shape parameter, ξ_r= scalar upper tail GPD shape parameter, u_r= scalar upper tail threshold, σ_{u_r}=scalar upper tail GPD shape parameter. The corresponding estimates of the vector is Θ=($\mu = 22.2467, \beta = 568.722, u_l = -781.4266, \sigma_{u_l} = 401.5377, \xi_l = -0.2866, u_r = 886.0329, \sigma_{u_r} = 283, \xi_r = 0.03959$) on left and Θ=($\mu = 47.8436, \beta = 454.0504, u_l = -914.749, \sigma_{u_l} = 429.9785, \xi_l = -0.4298, u_r = 966.8171, \sigma_{u_l} = 292.584$) on the right.

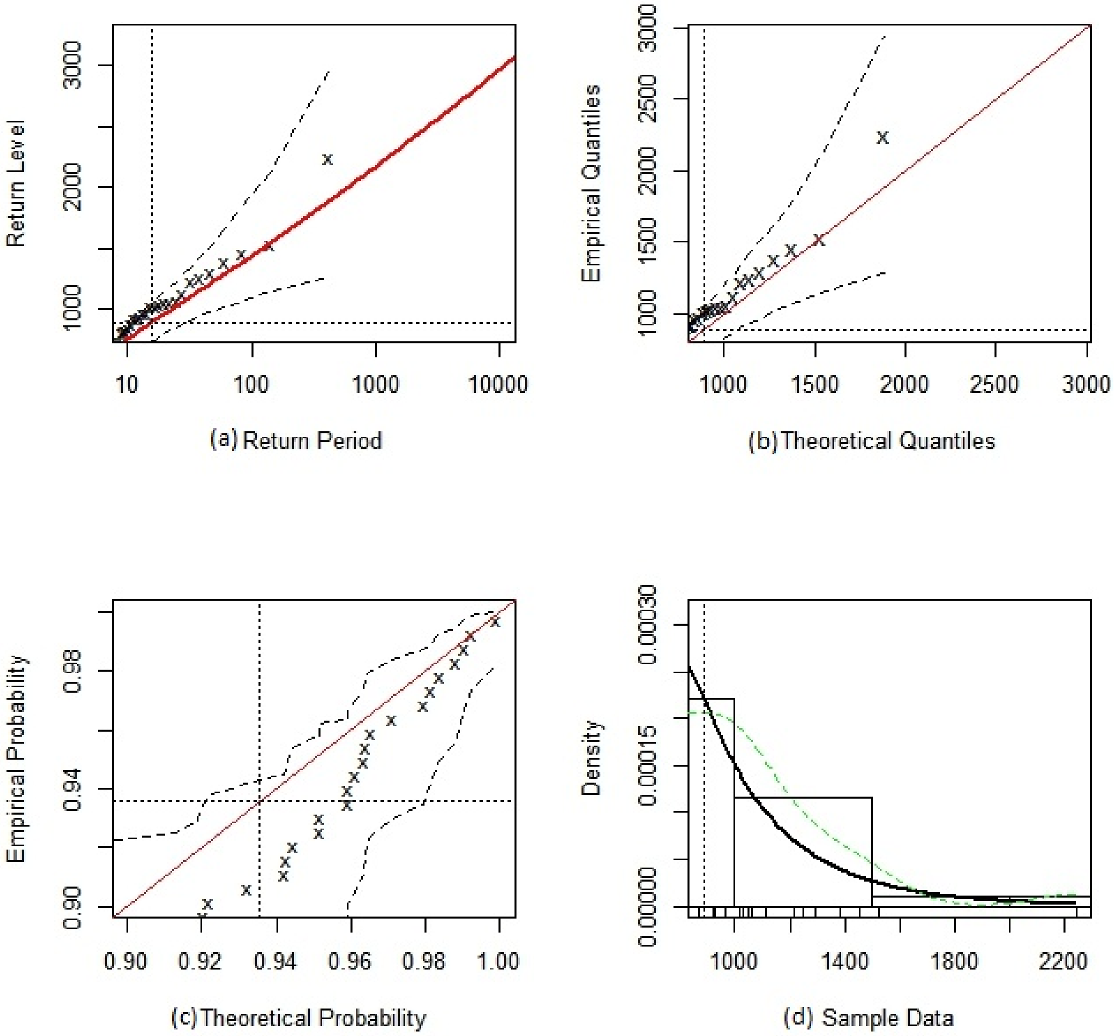


FIGURE 6.11: Bulk model based tail fraction approach diagnostics plots, (a) Top left panel: Return level plot, (b) Top right panel: Quantile plot, (c) Bottom left panel: Probability plot and (d) Bottom right panel: Density plot

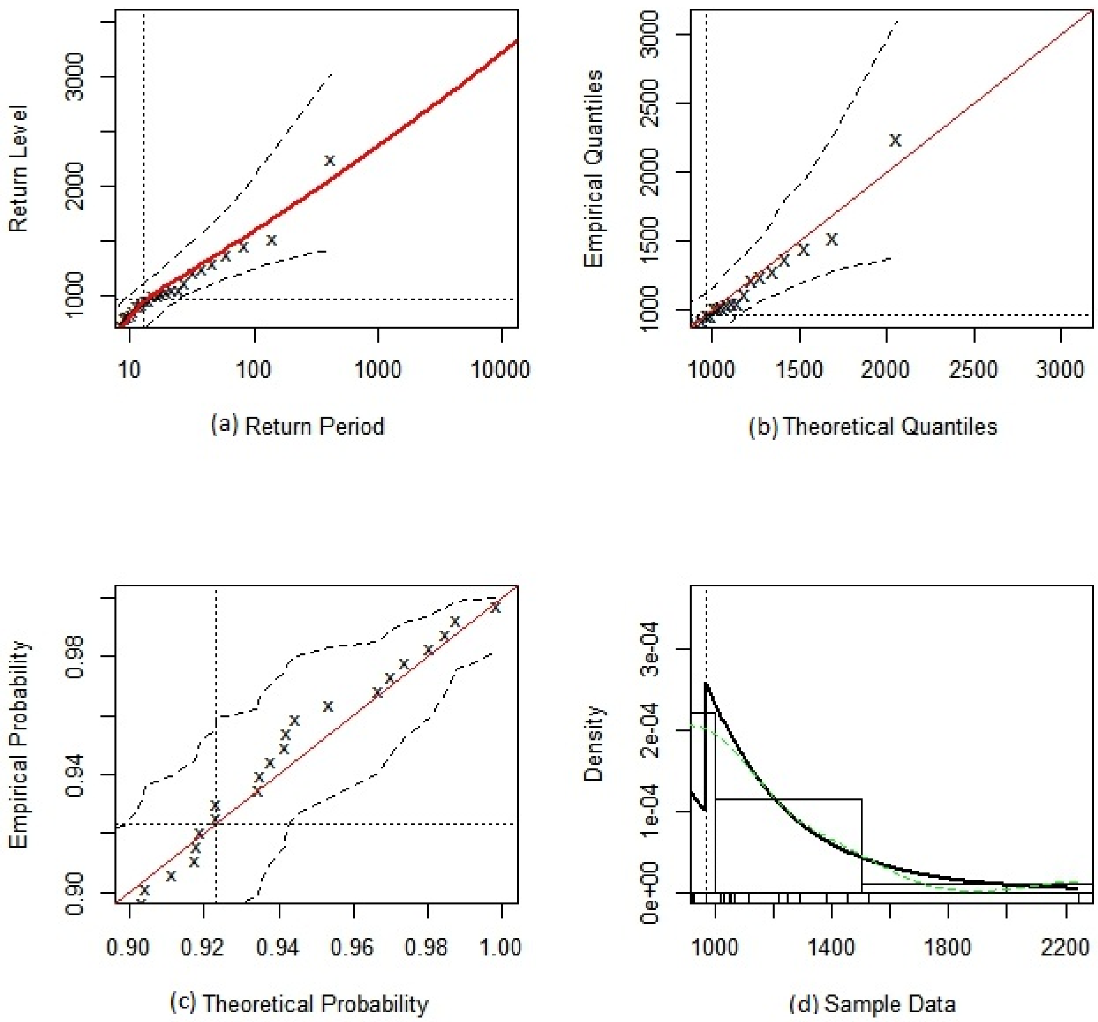


FIGURE 6.12: Parameterised tail fraction approach diagnostics plots, (a) Top left panel: Return level plot, (b) Top right panel: Quantile plot, (c) Bottom left panel: Probability plot and (d) Bottom right panel: Density plot

Figure 6.11 and 6.12 represent the parametric diagnostics plots of the two approaches, bulk and parameterised model tail fraction approach respectively. Prior to information depicted in Figure 6.10, the bulk tail model is represented with red dashed line and the parameterised tail model with the blue lines. Using the bulk tail model, the scalar lower threshold and the scalar upper tail threshold are given as $u_l = -781.4$ and $u_r = 886.03$ respectively. Using the parameterised tail fraction, the scalar lower threshold and the scalar upper tail threshold are given as $u_l = -914.4$ and $u_r = 966.82$, respectively. Comparatively, the diagnostic plots in Figure 6.12 confirm that the bulk tail fraction approach is poorer than the parameterised tail fraction approach when looking at their respective diagnostic plots. In Figure 6.12 the points are aligned along the diagonal line better than the points in Figure 6.11.

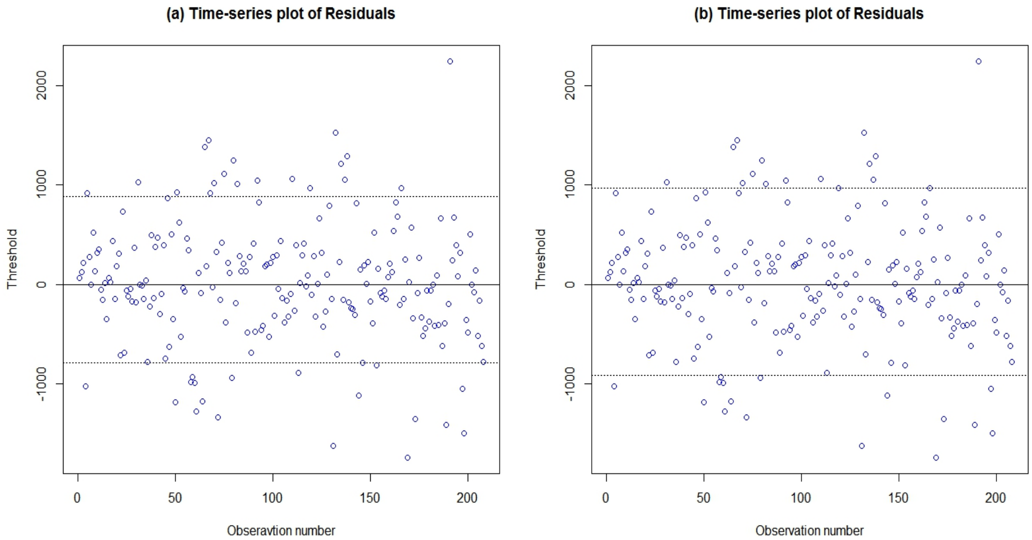


FIGURE 6.13: (a) Time series plot of residuals showing the upper and lower thresholds estimated using extreme value mixture model approach (Parameterised tail fraction approach), (b) Time series plot of residuals showing the upper and lower thresholds estimated using extreme value mixture model approach (bulk model based tail fraction approach)

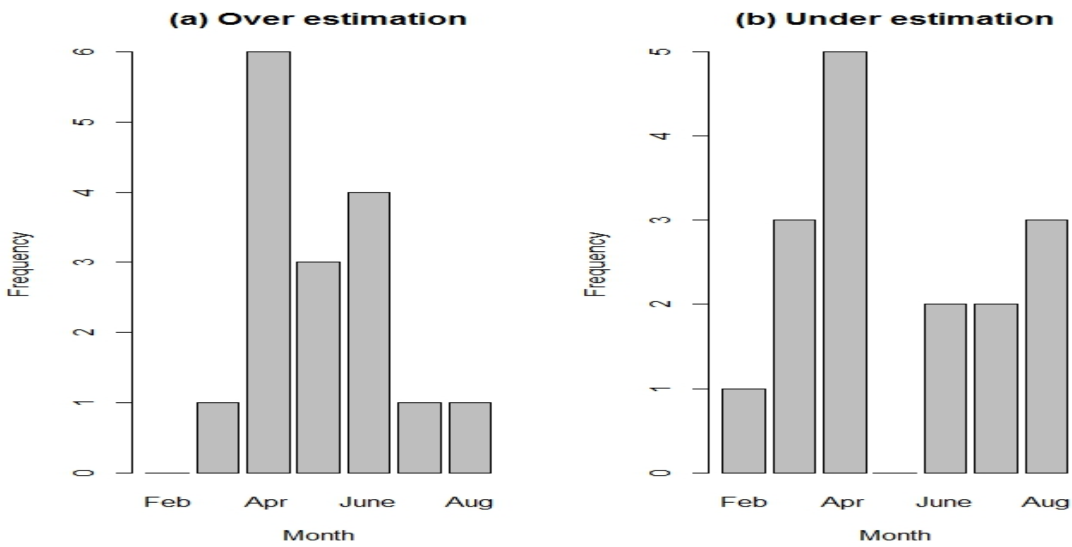


FIGURE 6.14: Frequency analysis of the occurrences of (a) over estimation and (b) under estimation residuals demand above the thresholds.

Figure 6.13 represent a time series plot of residuals showing the upper and lower thresholds and Figure 6.14 represent a frequency analysis of the occurrences of over and under estimation above the negative and positive thresholds respectively. Figure 6.14 indicates that a lot of over and under estimation occurred during the month of April. February had the least number of over estimated demand occurrences while May had the least number of occurrences for under estimated demand. This information is important to system operators for optimal acceleration of electrical energy scheduling in South Africa.

6.5 Modelling Extreme Stationary Series

Generalized Pareto Distribution is a flexible tool used to model exceedances. We adopt the stationary case in which all the positive exceedances are modelled. In accordance, the same procedure can be undertaken with negative residuals. As mentioned earlier, positive exceedances are representing under predictions and negative exceedances represent over predictions. Applying POT to the positive exceedances over some threshold 967, the parameter estimates are obtained as

$$\hat{\sigma} = 234.1232(96.8782)$$

and

$$\hat{\xi} = 0.1560(0.3261),$$

respectively, with standard errors parenthesised. Further, we see the negative log likelihood value is 99.2363. Figure 6.15 shows diagnostic plots from the proposed fitting positive exceedances, and suggest that the stationary GPD reasonably fits the positive exceedances well. The positive estimated shape parameter is indicative that the heavy tailed distribution exist.

Similarly applying POT to the negative exceedances over a threshold -914.75, we get the parameter estimates as

$$\hat{\sigma} = 1395.658(2.001e - 06)$$

and

$$\hat{\xi} = -1.527(0.001)$$

respectively, with S.E in parentheses. The value of the negative log likelihood is 579.4521. Figure 6.16 presents the negative exceedance diagnostic plots which suggest that the stationary GPD reasonably fits the data well. The negative estimated shape parameter is indicative that a finite right distribution end point exist.

Figure 6.14 indicates that the GPD qualifies to model the positive threshold exceedances. The return level estimates for residuals lie within the 95% confidence interval. As we approach towards the end of the GPD section, an estimate of return levels for both the positive and negative exceedances over thresholds of 966.7 and -914.75 respectively are estimated.

Further, we explore the model fit of the residuals by including time varying parameters in the GPD model.

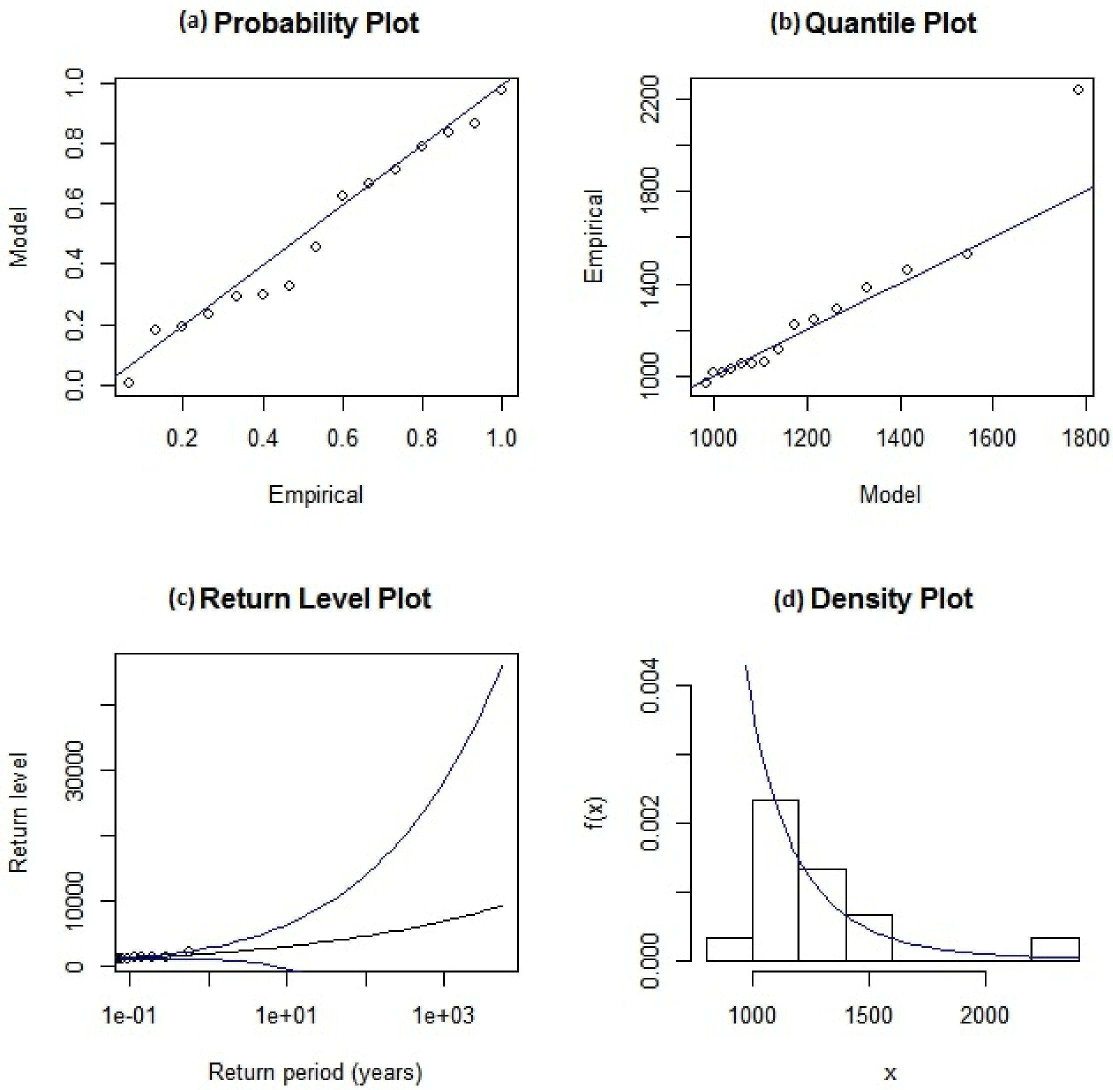


FIGURE 6.15: Diagnostic plots used to assess GPD (only for positive residuals) goodness–of–fit, (a) **Top left panel**: Return level plot, (b) **Top right panel**: Quantile plot, (c) **Bottom left panel**: Probability plot and (d) **Bottom right panel**: Density plot

6.6 Modelling Extreme Non-Stationary Series

Fitting the non-stationary GPD model to the exceedances, which allows some linear trend through the "log scale parameter" the fitted model gives.

$$\sigma\hat{(t)} = \exp(17.536 + 3.723t)$$

$$\hat{\xi} = -0.414$$

with the negative log likelihood as 97.680. The estimated shape parameter of the positive residuals is clearly negative and it indicates that the distribution is classified as Weibull. Figure 6.15 shows the diagnostics model for a non stationary model allowing some linear trend on the "log scale parameter". We thus conclude that the non stationary model is not well fitted based on the diagnostic plots.

To affirm our findings, we compare the non stationary model which allows for linear trend through time on the log scale parameter with the stationary model (model 1), then we get

$$D = 2\{97.6975 - (99.2248)\} = -3.0546,$$

which is small relative to $\chi_1^2(0.05) = 3.841$. Evidently, we see that there is no linear trend in the log–scale parameter.

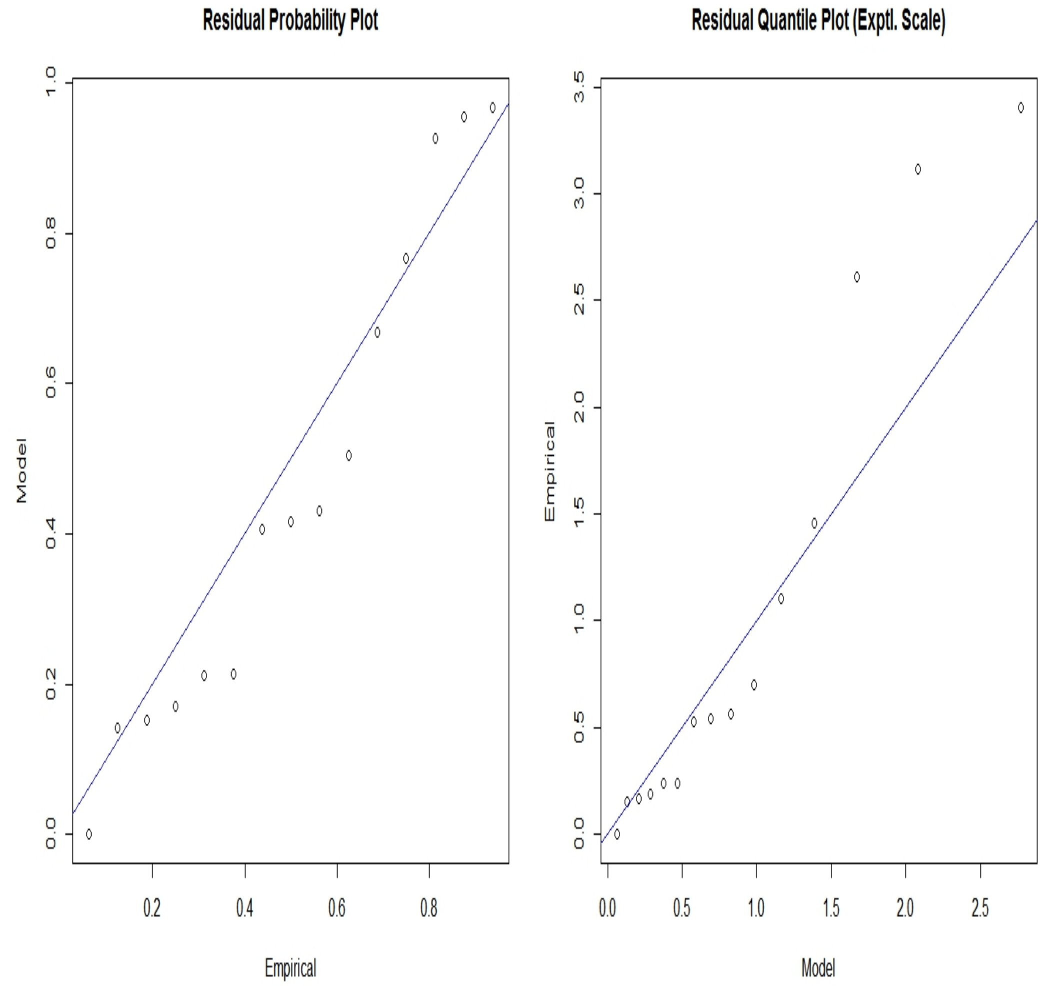


FIGURE 6.16: (a) Left panel: Residuals probability plot and (b) Right panel: Residuals Quantile plot

Comparatively, positive residuals are well fitted using the stationary GPD as compared to the non-stationary GPD model evident from Figure 6.15. It is clear that, the truncated Weibull type distribution does not fit the non-stationary GPD data well and the Pareto type fits residuals to the stationary GPD very well. A similar analysis can be done with over predictions (negative residuals).

6.6.1 Return Levels and Confidence Limits

The GPD 95^{th} quantile estimate is given as

$$\hat{x}_{0.05} = 966.7 - \frac{234.12}{0.1599}(\text{-log}(0.95)^{0.1599} - 1) = 2418.646.$$

6.7 Bayesian Estimation and Simulation Study

The primary objective of a Bayesian inference is to represent the uncertainty of any scientific experiment through probability distributions dependent on some unknown parameters. Expert information about these unknown parameters is modelled through probability distribution, known specifically as prior information which we denote by $\pi(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the parameter space, that is $\boldsymbol{\theta} = \{\theta_1, ..., \theta_p\}$. As sample data, $\mathbf{x}$ is made available, the information about the model parameters is expressed as the 'Likelihood function' $\pi(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{x} = \{x_1, ..., x_n\}$. The prior distribution of θ can take any particular value based on the additional information being updated. Applying Bayes' theorem, "proportional to the distribution of the observed data given the model parameters", we have $\pi(\boldsymbol{\theta}|\boldsymbol{x})$. This new updated belief about our prior beliefs is called the posterior distribution (Darnieder, 2011). The posterior distribution is given as:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{6.12}$$

After we have observed the data, the beliefs about the parameter space θ are expressed through the posterior distribution and can be rewritten (Bolstad, 2010) as;

$$\pi(\theta_1, \theta_2, ..., \theta_p|x_1, x_2, ..., x_n) \propto \pi(\theta_1, \theta_2, ..., \theta_p) \times \pi(x_1, x_2, ..., x_n|\theta_1, \theta_2, ..., \theta_p)$$

or,

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{6.13}$$

where $\pi(\boldsymbol{\theta})$ is the prior, $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood function and $\boldsymbol{\theta}$ is the vector of the parameters to be estimated and in a GPD scenario, the parameter $\boldsymbol{\theta}$ corresponds to (ξ, σ). Over the parameter space $\boldsymbol{\theta}$ is where we take the integral.

6.7.1 Prior Elicitation

The specification of the prior $\pi(\boldsymbol{\theta})$ is the most important aspect of using the Bayesian estimation. So, we must first specify a suitable prior, which is often characterized by expert opinion. There are two types of priors to help carry out objectively valid results (DeGroot and Schervish, 2002). They are non-informative priors such as flat priors and informative priors which include conjugate and Jeffrey's priors, e.t.c. In this dissertation, a common type of priors that will be used is non-informative prior. In particular, we will look at Jeffreys' prior developed by Jeffreys (1961).

6.7.1.1 Jeffreys' Prior

This is the standard objective prior specifically for Bayesian inferences. In order to obtain the Jeffreys's Prior we start by finding the Fisher Information matrix determinant and take the square root of the determinant. This is mathematically represented as,

$$J(\xi, \sigma) \propto \sqrt{(|I(\xi, \sigma)|)} \tag{6.14}$$

where $I(\xi, \sigma)$ is the Fisher information matrix (Jeffreys, 1961) and

$$I(\boldsymbol{\theta}) = E\left\{ -\frac{\partial^2 \log f_X(\boldsymbol{X}|\boldsymbol{\theta})}{\partial^2 \boldsymbol{\theta}} \right\}. \tag{6.15}$$

For a detailed discussion, see Beirlant et al. (2004a).

6.7.2 Markov Chain Monte Carlo (MCMC)

In order to avoid the complexity of the integration computation in equation 6.12 we simulate the posterior distribution realisation using Markov Chain Monte Carlo (MCMC) (Smith, 2005). In particular, the Metropolis algorithm will be used.

6.7.2.1 Metropolis Algorithm

The Metropolis algorithm involves simulation of a sequence. For a detailed literature discussion on the algorithm see Beirlant et al. (2004b); D'Agostini (2003); among others. We use the Bayesian analysis approach to estimate the GPD parameters estimates. The R-statistical package 'evdBayes' by Gilleland et al. (2013) is used for obtaining the estimates. We use the maximum likelihood parameter estimates along with their corresponding standard errors as the initial solution. In order to maximize the posterior distribution, we use simulated annealing (SANN). For an extensive and detailed discussion of SANN, see Bandyopadhyay et al. (2008) and Van Laarhoven and Aarts (1987);

among others. The Bayesian parameter estimates are thus given as,

$$\hat{\mu} = -210.79,$$

$$\hat{\sigma} = 619.and\hat{\xi} = -0.22.$$

The negative sign of the parameter estimate $\hat{\xi}$, through the EVI is indicative that the data can be modelled using a Weibull class since it has an upper bound.
The starting value $\boldsymbol{\theta}_0 = (-222.146, 614.515, -0.2136)$ is relatively poor as it is not close to the posterior distribution. To derive the starting value, we use the "mposterior" function in 'evdBayes' to maximise the posterior distribution and it is given as $\boldsymbol{\theta}_0 = (-217.57, 608.92, -0.21)$.

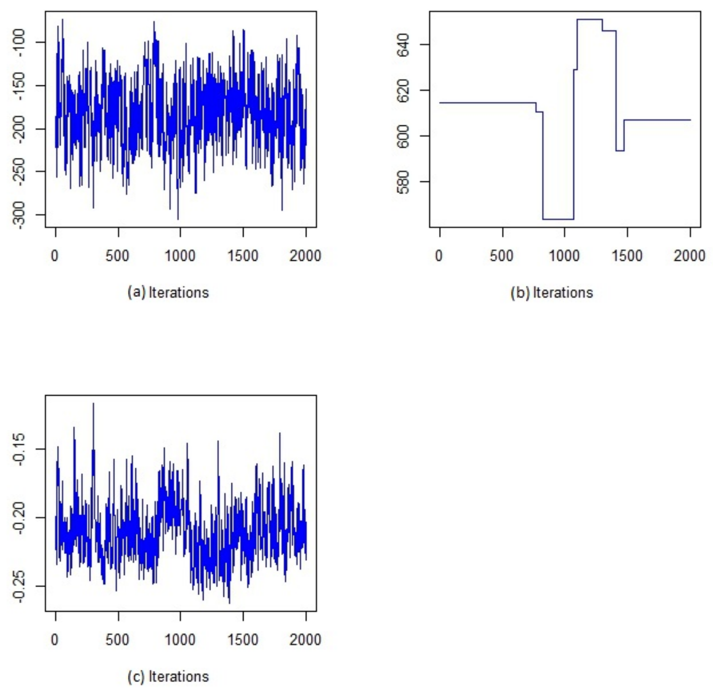


FIGURE 6.17: Generalized Extreme Value parameters MCMC realizations of the residuals data.

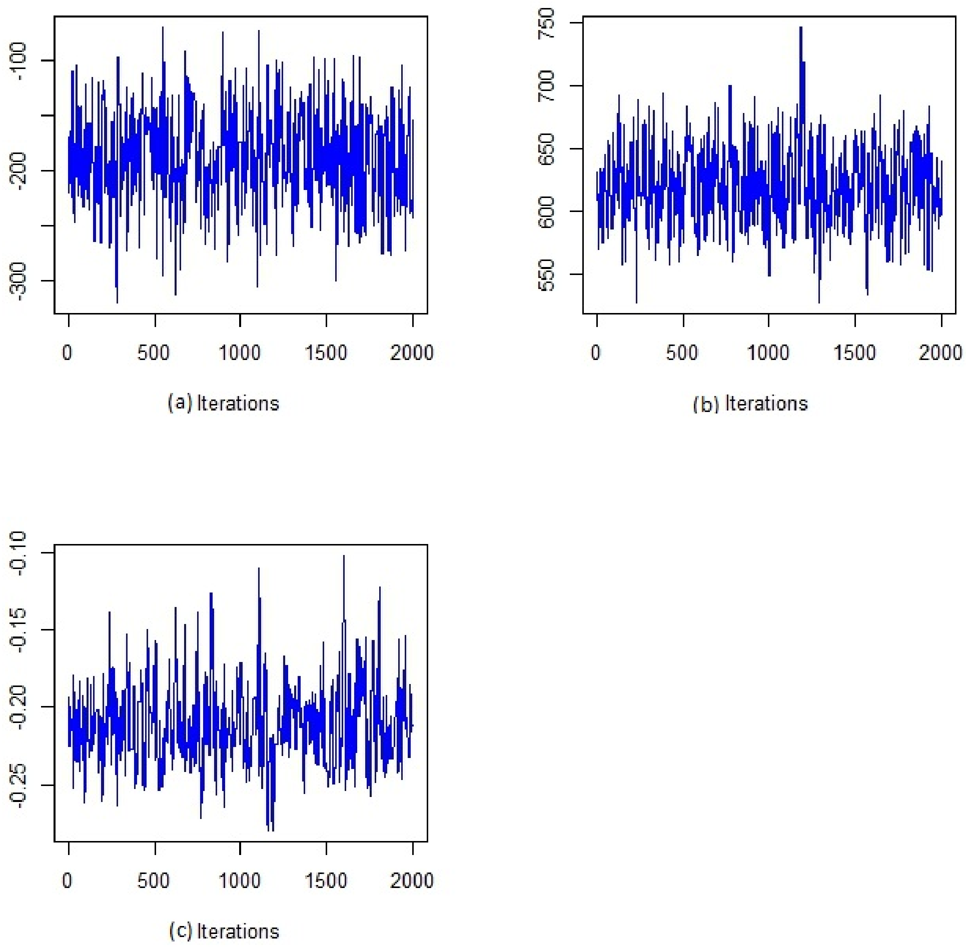


FIGURE 6.18: Generalized extreme value parameters MCMC realizations using different starting values than those used to produce the realizations of Figure 6.17

.

Based on conclusions presented, Figure 6.18 shows an improvement in convergence compared to Figure 6.17. To generate the iterations in Figure 6.18, the starting value $\boldsymbol{\theta}_0 = (-216.04, 611.42, -0.22)$ has been used and clearly it can be seen that there is a smaller burn in period. The burn in period is defined as the initial part of the chain. Further, the proposal standard errors $\boldsymbol{s} = (106.09, 0.11, 0.05)$ lead to improved mixing properties.

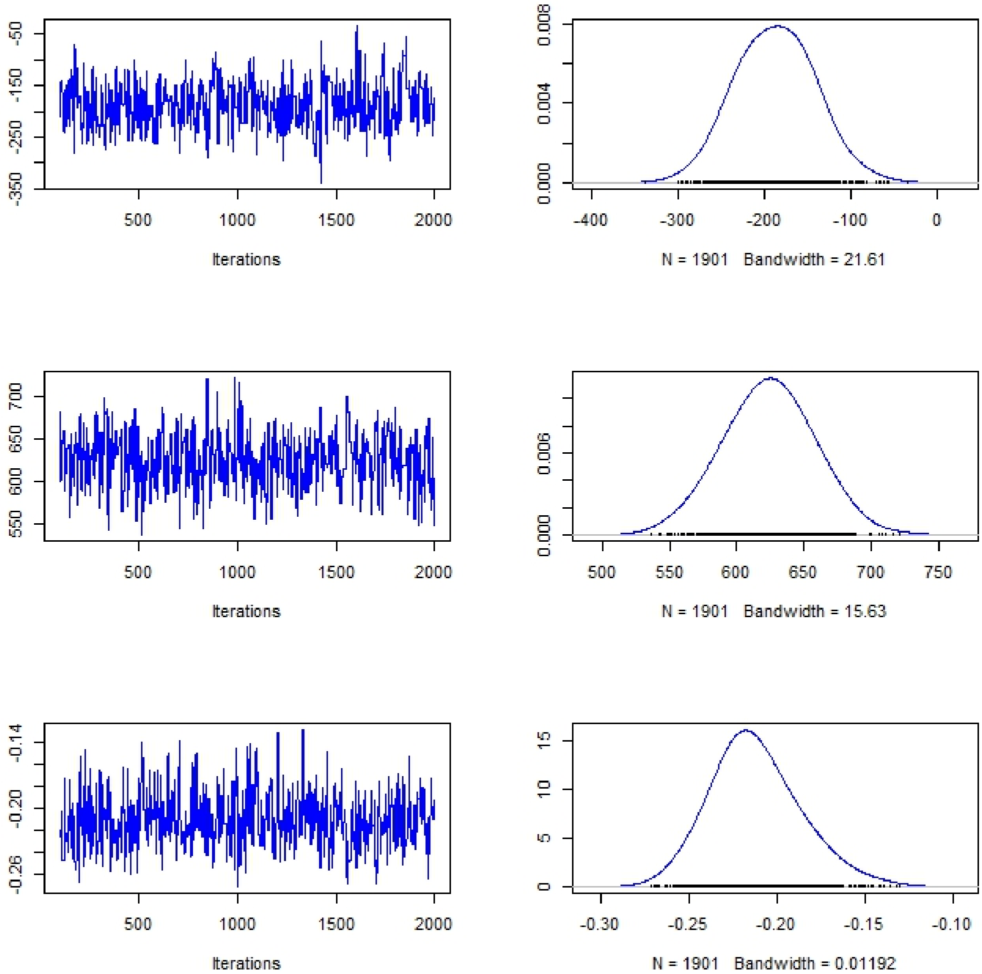


FIGURE 6.19: **Left Panels** (a) Trace plots of the three parameters and their marginal posterior densities, **Right panels** (b) Marginal posterior densities for the generalized extreme value parameters μ, σ and ξ respectively.

Figure 6.19 illustrates estimates of the marginal posterior densities with their corresponding trace plots.

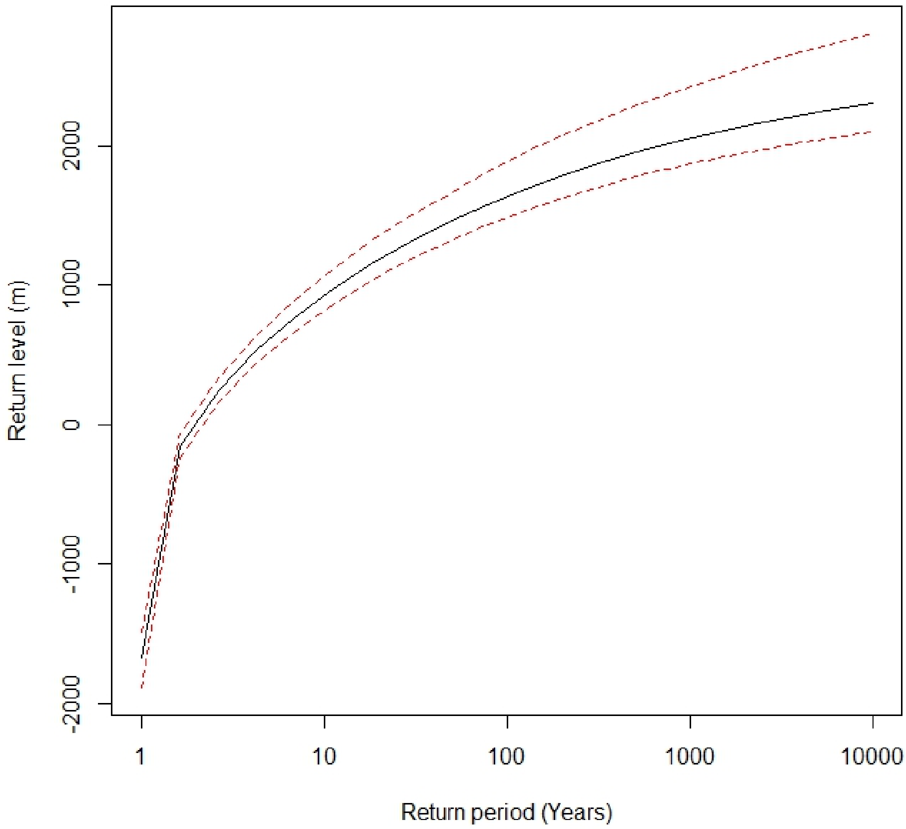


Figure 6.20: The Posterior distribution return level plot demonstrating the 95% Bayesian credible intervals in the form of the dashed lines

.

6.8 Simulation Study

In order to assess the the different estimators performance we conduct a simulation study by examining the performance of their Extreme Value Index (EVI) estimators. There are steps that we follow in conducting a simulation study which include the following:

- A sample size of n is first chosen, we will test the performance of different estimators. We use R2OpenBUGS package in R, and the program will run in OpenBugs software.
- simulate from a Frechét distribution where the EVI is $1/\sigma$ prior.
- sort observations into n ordered statistics. We then call the residuals data, picking out only the positive residuals.
- Hill estimate is used as a starting point for the Bayesian chains.
- Run the Bugs model on 5000 iterations, nburnin= 2000 and n.thin= 2. The number of iterations can be reduced.
- we run the simulations in parallel.

The code to the simulation results in Table 6.4 is given in the Appendix.

TABLE 6.4: Simulation results on the three methods

Bayesian GPD	Median hill	MLE GPD
0.1995	0.2317	0.1596

Table 6.4 represents different estimates of the EVI using different estimators. However, this alone cannot tell which is the better method since the value of EVI is unknown in this case and so we compare the methods of estimation to themselves. Generally, the Median Hill estimator has a large bias, so we would expect it to overestimate the EVI. As can be seen, the Bayesian estimate is in the middle of the two others, so the MLE could be underestimating the EVI. So in order to decide which method among the three is best in estimating parameters we will conduct another simulation study from a Frechét distribution.

Simulation Results

We choose a distribution with a true underlying EVI in the second simulation study conducted. We consider data simulated from a Frechét distribution with a tail index of 0.5. Consequently, the estimator that will be be able to estimate the value approximately closer to 0.5 and also have the minimal mean square error (MSE) value is considered the better estimator. As a basis of the study, we will only simulate from one distribution which is the Frechét distribution. In order to build a strong case, we will restrict the case to the following: consider 200 samples we make a table for when $n = 1000$, $k = 5\% * 1000$, $k = 15\% * 1000$, $k = 25\% * 1000$ and $k = 50\% * 1000$ and since we are only concerned with the Frechét distribution, we will choose the EVI range as $0 < \xi \leq 1$. This will be used as a basis to conclude which is the best method among the three.

Further, we calculate the mean of the square α errors of the simulated EVIs, yielding the mean square error (MSE). MSE is widely known and more often used to quantify and assess the performance when conducting a simulation study.

TABLE 6.5: For 200 samples, $\alpha = 0.1$, $n = 1000$, $k = 5\% * 1000, 15\% * 1000$ and $25\% * 1000$, EVI and MSE results of the three estimators are as follows

n	k	α	Median Hill(MSE)	Bayesian GPD (MSE)	MLE GPD (MSE)
1000	50	0.1	0.472 (13.90%)	0.042 (0.72%)	0.068 (2.90%)
1000	150	0.1	0.357 (6.60%)	0.030 (0.64%)	0.065 (0.86%)
1000	250	0.1	0.303 (4.10%)	0.030 (0.65%)	0.064 (0.51%)
1000	500	0.1	0.232 (1.70%)	0.011 (0.82%)	0.017 (0.84%)

For samples of size=1000 and $k = 50$, the method which yielded the lowest MSE, is the Bayesian estimator, with $100 \times \text{MSE} = 0.72$. Notably, we can make similar statements with respect to other sample sizes. Generally considering a clear and intuitive interpretation from Table 6.5 we can confirm that Bayesian estimator is more captivating as it provides a way of taking uncertainty into consideration (DeGroot and Schervish, 2002). We now present tables similar to Table 6.5 for all the other sample sizes. At the end of this series of tables we will provide further interpretation of the results.

Table 6.6: For 200 samples, $\alpha = 0.1$, $n = 1000$, $k = 5\% * 1000, 15\% * 1000$, $25\% * 1000$ and $50\% * 1000$, EVI and MSE results of the three estimators are as follows

n	k	α	Median Hill(MSE)	Bayesian GPD (MSE)	MLE GPD (MSE)
1000	50	0.25	1.17 (84.30%)	0.12 (3.60%)	0.21 (3.80%)
1000	150	0.25	0.86 (40.50%)	0.15 (2.40%)	0.20 (1.33%)
1000	250	0.25	0.76 (26.40%)	0.18 (1.35%)	0.21 (0.74%)

Table 6.7: For 200 samples, $\alpha = 0.1$, $n = 1000$, $k = 5\% * 1000, 15\% * 1000$, $25\% * 1000$ and $50\% * 1000$, EVI and MSE results of the three estimators are as follows

n	k	α	Median Hill (MSE)	Bayesian GPD (MSE)	MLE GPD (MSE)
1000	50	1.0	4.65 (13.40%)	1.04 (7.10%)	0.98 (7.80%)
1000	150	1.0	3.55 (6.52%)	0.97 (2.44%)	0.99 (2.43%)
1000	250	1.0	3.05 (20.36%)	0.97 (1.64%)	0.98 (1.65%)

Table 6.8 presents the summary of results from Tables 6.5 to 6.7.

Table 6.8: Simulation Results Summary

k size	α	Optimal estimator	100×MSE
50	0.1	Bayesian GPD	0.72
150	0.1	Bayesian GPD	0.64
250	0.1	MLE GPD	0.51
500	0.1	Bayesian GPD	0.82
50	0.25	Bayesian GPD	3.6
150	0.25	MLE GPD	1.33
250	0.25	MLE GPD	0.74
50	1.0	Bayesian GPD	7.1
150	1.0	MLE GPD	2.43
250	1.0	Bayesian GPD	1.64

Table 6.8 reports only the estimators which yielded the lowest overall MSE. As can be seen the Bayesian estimator is the best estimator among the three estimators.

Additional remarks

In order to estimate the parameters we have considered both the frequentist and Bayesian methods. Comparatively, results obtained from the two methods are fairly the same but the Bayesian approach is more appropriate since it accounts for parameter estimation uncertainty (DeGroot and Schervish, 2002).

6.9 Power Outage Risk Assessment and Management

It is the goal for every power system decision maker to have the national grid operate both efficiently and economically (Shenoy and Gorinevsky, 2014). To understand the behavior of power system across South Africa, we model the short term probabilistic load forecasting with an objective to understand the behavior of various variables versus the logarithmic hourly electricity demand. This information is not only important to the system operators for making power system operational decisions but in providing some insights into the South African electricity market now and in the future.

Electric load forecasting can be used as a crucial tool for demand management in South Africa. In particular, load forecasting can be used to assist in planning how much electricity must be reduced during system peak hours in order to meet the electricity demands every day. In chapter 4 it was discovered that peak hourly electricity demand of both sectors combined is observed at 19:00 and 20:00. From a practical point of view, with this information known, then high electricity demand should be reserved, monitored and managed well at such periods in order to avoid unnecessary occurrences such as power outages. Over the past decade, South Africa has experienced electricity shortages influenced by increased electricity demand usage. Linares and Rey (2013) also establish that South Africa in 2008, 2009 suffered electricity shortfalls which were caused mainly by drought, fuel disruption or growth demand.

Security concerns pertaining to electricity supply surrounding power systems cannot be underestimated today. Peak electricity demand planning and management is a critical concern for system managers and the power grid energy efficiency (Linares and Rey, 2013; Shenoy and Gorinevsky, 2014). In the next section, in order to reduce the possibility of power outages occurring in South Africa we focus on evaluation of peak event risks of under demand predictions. Power outages need to be avoided as they trigger some large economic and social costs (Linares and Rey, 2013).

6.9.1 Power Outage Risk Assessment

Practically, extreme forecasting errors that result due to power outages pose a serious risk to the utility companies in terms of costs as they have to buy the shortfall or resort to load shedding (Shenoy and Gorinevsky, 2014).

6.9.1.1 Outage risk (r)

Shenoy and Gorinevsky (2014) define outage risk as "the expected number of outages that will occur in a year". Prior to some residual demand analysis undertaken in the previous chapter we have seen that the residuals are i.i.d. Now, the expected number of outages experienced can be computed as

$$n_{out} = r \times N_h, \tag{6.16}$$

where n_{out} is the number of outages per year, N_h represents the number of hours or days in a year (Shenoy and Gorinevsky, 2014). Since the forecasting model is predicting the day ahead forecast, $N_h = 365$ days and as can be observed from the Quantile plot in Figure 6.14 there is one extreme outage, and so we assume $n_{out} = 1$. So,

$$\text{risk parameter}(r) = \frac{n_{out}}{N_h} = \frac{1}{365} = 0.003$$

Notably, the risk parameter is determined $r = 0.00273973$ for a single outage in a year. This information is very important to system managers as they can use it to describe how much power on average can be reserved for the next day above the forecast (Shenoy and Gorinevsky, 2014). Electricity usage in South Africa is influenced by a lot of factors, and makes the electricity demand unpredictable and hence cannot be accurately forecasted. To model the residuals we use EVT to forecast the exceedance risk level.

6.10 Concluding Remarks

In this chapter we explored the use of EVT in modelling residual demand series in South Africa. We used the Generalized Extreme Value Distribution (GEVD) to model residuals and to model the positive residuals we used the Generalized Pareto Distribution (GPD). A GEVD and GPD comparative analysis is undertaken for fitting the residuals. Various graphical plots are used to estimate a sufficiently higher threshold, and the extreme mixture models are more appropriate in determining the threshold as they provide a way of taking uncertainty into account (DeGroot and Schervish, 2002).

GEVD and the GPD provide a good fit for residual demand analysis with information the empirical results provide. Furthermore, empirical results reveal that the GPD is easier to use as it has only two parameters to estimate instead of three as is the case with the GEVD. It is significant to model residuals in order to determine the level of risks and also accurately assess the frequency and future load forecasts level (Hor et al., 2008).

In order to estimate the GEVD parameters we used both the frequentist estimation method and Bayesian method. The Bayesian provide a way to account for risk uncertainty (MacDonald et al., 2011). A simulation study is conducted to confirm that Bayesian method is indeed a better than the frequentist approach. Finally, we discussed the power outage risk assessment and management in South Africa as a result of residual demand analysis. This information is important for load forecasters in South Africa to determine consistent, reliable supply schedules and also assess hourly event risks.

Chapter 7

Conclusion

7.1 Introduction

We presented a semi-parametric additive approach in modelling short term probabilistic electricity demand in South Africa in this dissertation. The dissertation started with modelling the daily electricity demand using discrete time Markov chain analysis, the second part of the dissertation which is Chapter 5 dealt with short term probabilistic load forecasting and residual analysis. This dissertation provides useful models which are important for operational decisions of power system operators. Chapter 5 dealt with residual demand analysis and this information helps in assessing the level of risk uncertainty in the short term load forecasting and evaluating hourly event risks in South Africa.

7.2 Summary and Concluding Remarks

Modelling daily maximum electricity demand in different hours of the day using discrete time Markov chain analysis (DTMC) was discussed in Chapter 4. The Chapter's contribution in this dissertation goes as far as contributing towards understanding the impact of daily extreme electricity demand in South Africa. According to findings in Chapter 4, high electricity consumption is mostly observable within 19:00 hours to 20:00 hours. This is due to continuous operation of the commercial sector (clinics, retail stores, street lights, e.t.c) even in the evenings, some factories still operating at night and the electricity consumption by the households. This information is important particularly in the power system operations, maintenance areas and making sure there is proper and efficient generating capacity (Cottet and Smith, 2003).

Chapter 5 was based on results obtained in Chapter 4, in order to develop a short term load forecasting on the hour with frequent high electricity demand. Contextually, we developed a probabilistic short term load forecasting model in this dissertation and we used the 19:00 hours daily data for the period of 10 years (01 January-10 September 2010). Nevertheless, we can still use the 20:00 hours data to develop a short term probabilistic model using the same approach, and likewise for any hours of the day. Chapter 5 followed a regression framework, in which we investigate some form of relationship that exist between various exogenous variables versus the 19:00 hours daily demand in South Africa. We formulated the semi-parametric additive model using 30 nonlinear regressors but the majority of these variables were not significant (see Table 5.2). There were only 14 significant regressors, which included variables such as calendar effects, weekly and monthly seasonal effects, Cooling Degree Days (CDD_t), Heating Degree Days (HDD_t), past information on the 19:00 hours temperature and demand. A benchmark model using the 19:00 hours data was developed from which some performance evaluation was carried out against the proposed model, documenting the effect of temperature on load and interactions that exist within variables and drive demand. In comparison, an out-of-sample evaluation on the two models was undertaken. Findings from the out-of-sample evaluation confirmed that the short term probabilistic forecasting model is better than the benchmark model based on the MAPE, MAE and RMSE. There is some evidence that here in South Africa temperature sensitivity has a strong effect on the electricity demand (Sigauke, 2014).

Chapter 6 provided an extension of both Chapters 4 and 5. Some special aspect involved in Chapter 6 is modelling the tail distribution of positive residual demand (underestimation) for risk uncertainty assessment of the forecasting model developed in Chapter 5 and further assess the hourly event risks. This information is particularly useful to system operators knowing that the under demand prediction forecasts can lead the power system to power outages, load shedding, blackouts and ultimately threaten to destabilize the national grid. Although we analysed under demand residuals in this dissertation it is important to note that over demand (overestimation) residual risk analysis can be done using the same modelling approach. EVT was used to model positive residual demand analysis. GEVD and GPD were both used to analyse the positive residual demands. Findings from Chapter 6 suggest GPD is robust in modelling the tail distribution of positive residual demand (underestimation). Further, we estimate the positive residuals data parameters using the Bayesian approach. In comparison, the results obtained from the frequentist approach and those obtained from the Bayesian approach are fairly closer to each other. However, Bayesian approach is deemed favourable as it accounts for parameter estimation uncertainty and even more, is informative (DeGroot and Schervish, 2002). The last part of Chapter 6 looked at the power outage risk assessment in South Africa based on findings from Chapter 4, 5 and 6. In this study, it was shown that accurate electricity demand timing and forecast at different hours of the day can help in scheduling maintenance and planning.

Finally, it was also shown that to quantify and assess basis for risk for forecasting uncertainty we model under and over electricity demand.

7.3 Limitations of the Dissertation

The hourly electricity demand data comes from various sectors of the South African economy which include residential, commercial and agricultural sector. Consequently, it remains a challenge to know with absolute certainty how significant each individual economic sector is on everyday demand.

7.4 Future Research

The findings of this dissertation provide possible areas for future research as follows:

- Investigate how monthly electricity demand is affected by some weather related variables (Hor et al., 2005). This will be carried out using the South African historic data.
- Future research will look at probabilistic description and modelling of daily maximum electricity demand using Poisson point process.
- Since a great deal of this study is an industry application, future research will consider use of informative priors. Elicitation of a system managers' prior beliefs about the various quantities should actually yield a significant improvement in the accuracy of the results.
- The simulation study was brief and only a few simulations carried out. Thus, future research would look into a larger simulation study and consider variety of distributions.
- In this dissertation, we used an aggregated national data. Future research aims to focus on individual provinces as electricity patterns exhibited by all provinces differ. For example the electricity pattern in coastal areas is expected to differ with inland provinces. This information is important to the provincial power system managers for strategic and operational decisions.

References

Abu-El-Magd, M. A. (1983). *Short-term load demand modelling and forecasting*, PhD Thesis, McMaster University.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, pages 792–804.

Bandyopadhyay, S., Saha, S., Maulik, U., and Deb, K. (2008). A simulated annealing-based multi-objective optimization algorithm: Amosa. *IEEE Transactions on Evolutionary Computing*, 12(3):269–283.

Bathurst, G. N., Weatherill, J., and Strbac, G. (2002). Trading wind generation in short term energy markets. *IEEE Transactions on Power Systems*, 17(3):782–789.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004a). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004b). Bayesian Methodology in Extreme Value Statistics. *Statistics of Extremes: Theory and Applications*, pages 429–459.

Beirlant, J., Vynckier, P., and Teugels, J. L. (1996). Tail Index Estimation, Pareto Quantile Plots Regression Diagnostics. *Journal of the American statistical Association*, 91(436):1659–1667.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Bernardo, J. M. and Smith, A. F. (2009). *Bayesian theory*, volume 405. John Wiley & Sons.

Bolstad, W. M. (2010). *Understanding computational Bayesian statistics*, volume 644. John Wiley & Sons.

Bommier, E. (2014). Peaks-over-threshold modelling of environmental data.

Bracale, A., Caramia, P., Carpinelli, G., Di Fazio, A. R., and Varilone, P. (2013). A Bayesian-based approach for a short-term steady-state forecast of a smart grid. *IEEE Transactions on Power Systems*, 4(4):1760–1771.

Castillo, E. (2012). *Extreme value theory in engineering*. Elsevier.

Chan, S. and Nadarajah, S. (2015). Extreme value analysis of electricity demand in the U.K. *Applied Economics Letters*, 22(15):1246–1251.

Cheng, L., AghaKouchak, A., Gilleland, E., and Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, 127(2):353–369.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Coles, S. G. and Powell, E. A. (1996). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review/Revue Internationale de Statistique*, 64(1):119–136.

Cottet, R. and Smith, M. (2003). Bayesian modeling and forecasting of intraday electricity load. *Journal of the American Statistical Association*, 98(464):839–849.

Cowles, M. K. (2013). *Applied Bayesian statistics: with R and OpenBUGS examples*, volume 98. Springer Science & Business Media.

Cyert, R. M. and DeGroot, M. H. (1970). Multiperiod decision models with alternating choice as a solution to the duopoly problem. *The Quarterly Journal of Economics*, 84(3):410–429.

D'Agostini, G. (2003). Bayesian inference in processing experimental data: principles and basic applications. *Reports on Progress in Physics*, 66(9):1383.

Darnieder, W. F. (2011). *Bayesian methods for data-dependent priors.* PhD thesis, The Ohio State University.

Davison, A. and Kuonen, D. (2002). An Introduction to the Bootstrap with Applications in R. *Statisticial Computing and Statistical Graphics Newsletter*, 13(1):6–11.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442.

DeGroot, M. H. and Schervish, M. J. (2002). Probability and statistics.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45.

Embrechts, P. (1999). Extreme value theory in finance and insurance. *Manuscript, Department of Mathematics, ETH, Swiss Federal Technical University.*

Fan, S. and Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1):134–141.

Gencay, R. and Selcuk, F. (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting*, 20(2):287–303.

Gilleland, E. and Katz, R. W. (2006). Analyzing seasonal to interannual extreme weather and climate variability with the extremes toolkit. In *18th Conference on Climate Variability and Change, 86th American Meteorological Society (AMS) Annual Meeting*, volume 29. Citeseer.

Gilleland, E., Ribatet, M., and Stephenson, A. G. (2013). A software review for extreme value analysis. *Extremes*, 16(1):103–119.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151.

Goude, Y., Nedellec, R., and Kong, N. (2014). Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE Transactions on Power Systems*, 5(1):440–446.

Gross, G. and Galiana, F. D. (1987). Short-term load forecasting. *IEEE Transactions on Power Systems*, 75(12):1558–1573.

Hasan, H., Salam, N., and Kassim, S. (2013). Modeling annual extreme temperature using generalized extreme value distribution: A case study in Malaysia. In *Proceedings of the 20th National Symposium on Mathematical Sciences: Research in Mathematical Sciences: A Catalyst for Creativity and Innovation*, volume 1522, pages 1195–1203. AIP Publishing.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC Press.

Hong, T. (2010). *Short term electric load forecasting*, PhD Thesis, North Carolina State University.

Hong, T. (2014). Energy forecasting: Past, present, and future. *Foresight: The International Journal of Applied Forecasting*, (32):43–48.

Hong, T., Wang, P., and Willis, H. L. (2011). A naïve multiple linear regression benchmark for short term load forecasting. In *IEEE Transactions on Power Systems*, pages 1–6. IEEE.

Hor, C., Watson, S., Infield, D., and Majithia, S. (2008). Assessing load forecast uncertainty using extreme value theory. *16th PSCC, Glasgow*.

Hor, C.-L., Watson, S. J., and Majithia, S. (2005). Analyzing the impact of weather variables on monthly electricity demand. *IEEE Transactions on Power Systems*, 20(4):2078–2085.

Hor, C.-L., Watson, S. J., and Majithia, S. (2006). Daily load forecasting and maximum demand estimation using ARIMA and GARCH. In *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, pages 1–6. IEEE.

Hu, Y. (2013). *Extreme Value Mixture Modelling with Simulation Study and Applications in Finance and Insurance*, M.Sc., University of Canterbury.

Hu, Y., Scarrott, C., and Scarrott, M. C. (2013). Package 'evmix'.

Hyndman, R. J. and Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.

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

Hyndman, R. J. and Fan, S. (2014). Monash Electricity Forecasting Model.

Ismail, Z., Yahaya, A., and KA, M. (2009). Forecasting peak load electricity demand using statistics and rule based approach. *American Journal of Applied Sciences*, 6(8):1618–1625.

Jeffreys, H. (1961). Theory of probability, harold jeffreys. *International series of monographs on physics.*

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kou, P. and Gao, F. (2014). A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises. *International Journal of Electrical Power & Energy Systems*, 55(1):144–154.

Linares, P. and Rey, L. (2013). The costs of electricity interruptions in spain. are we sending the right signals? *Energy Policy*, 61:751–760.

Liu, J. M., Chen, R., Liu, L.-M., and Harris, J. L. (2006). A semi-parametric time series approach in modeling hourly electricity loads. *Journal of Forecasting*, 25(8):537–559.

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6):2137–2157.

Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, 13(3):853–880.

McNeil, A. J. (1999). Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum.*

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300.

McSharry, P. E., Bouwman, S., and Bloemhof, G. (2005). Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Transactions on Power Systems*, 20(2):1166–1172.

Migon, H. S. and Alves, L. C. (2012). Modeling and forecasting intraday electricity load.

Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., Lalas, D., Moschovits, M., Karagiannis, F., and Papakonstantinou, D. (2006). Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31(2):208–227.

Morita, H., Kase, T., Tamura, Y., and Iwamoto, S. (1996). Interval prediction of annual maximum demand using grey dynamic model. *International Journal of Electrical Power & Energy Systems*, 18(7):409–413.

Muñoz, A., Sánchez-Úbeda, E. F., Cruz, A., and Marín, J. (2010). Short-term forecasting in power systems: a guided tour. In *Handbook of Power Systems II*, pages 129–160. Springer.

Nascimento, F. F., Gamerman, D., and Lopes, H. F. (2012). A semi-parametric Bayesian approach to extreme value estimation. *Statistics and Computing*, 22(2):661–675.

Oberhofer, W. and Haupt, H. (2005). The asymptotic distribution of the unconditional quantile estimator under dependence. *Statistics & probability letters*, 73(3):243–250.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131.

Poirier, D. J. (1988). Frequentist and subjectivist perspectives on the problems of model building in economics. *The Journal of Economic Perspectives*, pages 121–144.

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). *Sociological Methodology*, 25:111–195.

Renard, B., Lang, M., and Bois, P. (2006). Statistical analysis of extreme events in a non-stationary context via a Bayesian framework: case study with peak-over-threshold data. *Stochastic environmental research and risk assessment*, 21(2):97–112.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semi-parametric Regression.* Number 12. Cambridge university press.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Shenoy, S. and Gorinevsky, D. (2014). Risk adjusted forecasting of electric power load. In *American Control Conference (ACC), 2014*, pages 914–919.

Sigauke, C. (2014). *Modelling Electricity Demand in South Africa*, PhD Thesis, University of Free State.

Sigauke, C., Verster, A., and Chikobvu, D. (2013). Extreme daily increases in peak electricity demand: Tail-quantile estimation. *Energy Policy*, 53:90–96.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Singh, G., Chauhan, D., Chandel, A., Parashar, D., and Sharma, G. (2014). Factor Affecting Elements and Short term Load forecasting Based on Multiple Linear Regression Method. In *International Journal of Engineering Research and Technology*, volume 3. ESRSA Publications.

Smith, E. (2005). *Bayesian modelling of extreme rainfall data.* PhD thesis, University of Newcastle upon Tyne.

Spedicato, G. and Signorelli, M. (2014). The markovchain package: A package for easily handling discrete markov chains in r.

Spedicato, G. A., Kang, T. S., Thoralf, M., Spedicato, M. G. A., and Rcpp, L. (2015). Package 'markovchain'.

Stephenson, M. A. (2006). The ismev package.

Strengers, Y. (2012). Peak electricity demand and social practice theories: reframing the role of change agents in the energy sector. *Energy Policy*, 44:226–234.

Strengers, Y. and Maller, C. (2012). Materialising energy and water resources in everyday practices: Insights for securing supply systems. *Global Environmental Change*, 22(3):754–763.

Sugahara, S., Da Rocha, R. P., and Silveira, R. (2009). Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology*, 29(9):1339–1349.

Taylor, J. W. (2012). Short-term load forecasting with exponentially weighted methods. *IEEE Transactions on Power Systems*, 27(1):458–464.

Tong, X., Chen, Q., Fan, J., Yan, Q., and Kang, C. (2014). Adaptability verification and application of the t-distribution in short-term load forecasting error analysis. In *Power System Technology (PowerCon), 2014 International Conference on*, pages 145–150. IEEE.

Valor, E., Meneu, V., and Caselles, V. (2001). Daily air temperature and electricity load in spain. *Journal of Applied Meteorology*, 40(8):1413–1421.

Van Dongen, S. (2006). Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology*, 242(1):90–100.

Van Laarhoven, P. J. and Aarts, E. H. (1987). *Simulated annealing: theory and applications*, volume 37. Springer Science & Business Media.

Wang, H., Yang, K., Xue, L.-y., and Liu, S. (2015). The Study of Long-term Electricity Load Forecasting Based on Improved Grey Prediction. In *Proceedings of the 21st International Conference on Industrial Engineering and Engineering Management 2014*, pages 441–444. Springer.

Wasserman, L. (2013). *All of statistics: a concise course in statistical inference*, volume 1. Springer Science & Business Media.

Willis, H. L. and Northcote-Green, J. E. (1983). Spatial electric load forecasting: a tutorial review. *IEEE Transactions on Power Systems*, 71(2):232–253.

Xie, J., Hong, T., Laing, T., and Kang, C. (2015). On normality assumption in residual simulation for probabilistic load forecasting.

Yin, G. G. and Zhang, Q. (2006). *Discrete-time Markov chains: two-time-scale methods and applications*, volume 55. Springer Science & Business Media.

Yong, Y. (2005). Can the strengths of AIC and BIC be shared. *Biometrika*, 92(4):937–950.

Zhang, H., Wang, Z., Gao, F., and Zhou, D. (2012). A New Method for the Daily Electricity Consumption Forecast of Energy Intensive Enterprise. In *Power and Energy Engineering Conference (APPEEC), 2012 Asia-Pacific*, pages 1–4. IEEE.

Appendix A

Discrete time Markov chain analysis R code

```
DTMC <- read.csv("C:/Users/user/Desktop/ZT.csv")
attach(DTMC)
names(DTMC)
attach(Code1)
head(Code1)
tail(Code1)
win.graph()
x=ts(DPD)
plot(x,xlab="Observation number", ylab="DPD (megawatts)", col="blue")
par(mfrow=c(2,2))
y=ts(Zt)
plot(y,xlab="Observation number", ylab="Daily Changes (MW)",
     col="blue", main="(a) Daily changes in peak demand")
plot(density(y),col="blue",xlab="Daily changes (MW)",
     main= "(b) Density of daily changes")
qqnorm(y, col="blue", main=" (c) Normal QQ-plot of daily changes")
qqline(y)
boxplot(y, col="blue", main="(d) Box-plot of daily changes")
```

```
# Two states
names(DTMC)
attach(DTMC)

library(markovchain)
xChar<-as.character(Two.state)
mcX<-markovchainFit(xChar)$estimate
mcX
#Steady state probabilities
statesNames=c("1","2")
markovB<-new("markovchain", states=statesNames, transitionMatrix=
               matrix(c( 0.6540478, 0.3459522,
                         0.2331505, 0.7668495),
                      nrow=2, byrow=TRUE, dimnames=list(statesNames,
                                                        statesNames)
               ))
steadyStates(markovB)

#First passage time
simpleMc<-new("markovchain", states=c("1","2"), transitionMatrix=
                matrix(c( 0.6540478, 0.3459522,
                          0.2331505, 0.7668495),
                       nrow=2,byrow=TRUE))
firstPassage(simpleMc,"1",7)



simpleMc<-new("markovchain", states=c("1","2"), transitionMatrix=
                matrix(c( 0.6540478, 0.3459522,
                          0.2331505, 0.7668495),
                       nrow=2,byrow=TRUE))
firstPassage(simpleMc,"2",7)

# plot of a transition diagram
statesNames=c("1","2")
```

```
markovB<-new("markovchain", states=statesNames, transitionMatrix=
               matrix(c( 0.4295830, 0.5704170,
                         0.3846971, 0.6153029),
                      nrow=2, byrow=TRUE, dimnames=list(statesNames,
                                                        statesNames)
               ))
plotMc(markovB)



#-----------THRESHOLD USING PACKAGE LAEKEN-------------------------


attach(abovezero)
head(abovezero)
win.graph()
library(laeken)
summary(lnZt)
attach(paretoqp)
head(paretoqp)
win.graph()
library(laeken)
paretoQPlot(lnZt,w = NULL, xlab ="-log[1-i/(n+1)]",
            ylab ="Log of the observations", main="",
            interactive = TRUE,ylim=c(3.5,9),col="blue")

#----------------------- Three State Problem -----------------------
attach(threestates1)
head(threestates1)
win.graph()
xChar<-as.character(sorted)
library(markovchain)
mcX<-markovchainFit(xChar)$estimate
mcX
#Steady state probabilities
statesNames=c("1","2","3")
```

```
markovB<-new("markovchain", states=statesNames, transitionMatrix=
               matrix(c( 0.4354540, 0.009882644, 0.5546634,
                         0.4900000, 0.020000000, 0.4900000,
                         0.3392157, 0.032156863, 0.6286275),
                      nrow=3, byrow=TRUE, dimnames=list(statesNames,
                                                        statesNames)
               ))
steadyStates(markovB)

#First passage time
simpleMc<-new("markovchain", states=c("1","2","3"), transitionMatrix=

                matrix(c( 0.4354540, 0.009882644, 0.5546634,
                          0.4900000, 0.020000000, 0.4900000,
                          0.3392157, 0.032156863, 0.6286275),
                       nrow=3,byrow=TRUE))
firstPassage(simpleMc,"3",7)

# plot of a transition diagram
statesNames=c("1","2","3")
markovB<-new("markovchain", states=statesNames, transitionMatrix=
               matrix(c( 0.4354540, 0.009882644, 0.5546634,
                         0.4900000, 0.020000000, 0.4900000,
                         0.3392157, 0.032156863, 0.6286275),
                      nrow=3, byrow=TRUE, dimnames=list(statesNames,
                                                        statesNames)
               ))
plotMc(markovB)

#example for predict method
mcFit<-markovchainFit(data=sorted)
predict(mcFit$estimate, newdata=c("2"),n.ahead=10)
#example of summary
summary(simpleMc)
```

```
#----------------- FPT given current state is increase ------------
head(DTMC)
win.graph()
#GPD=c(0.2004,0.0628,0.0195,0.0060)
#GSP=c(0.1966,0.0688,0.0261,0.0106)
g_range=range(0.0002,0.6)
plot(d,type="o",axes=F,col="blue", ylab="Probability",
     xlab="Day")
lines(i,type="o",pch=22,lty=2,col="red")
axis(1,at=1:10, lab=c("1", "2", "3", "4", "5", "6", "7", "8", "9",
                      "10"))
axis(2, las=1, at=0.05*12:g_range[2])
legend("topright", col=c("blue", "red"),lty=c(1,2),cex=1,
       legend=c("Decrease","Increase"))
box()


#--------------FPT given current state is decrease----------------


win.graph()
#GPD=c(0.2004,0.0628,0.0195,0.0060)
#GSP=c(0.1966,0.0688,0.0261,0.0106)
d = c(0.6229,  0.2114,0.0929,0.0408,0.0179,0.0079,0.0035,
      0.0015,0.0007,0.0003)
i = c(0.3771,0.2349,0.1463,0.0911,0.0568,0.0354,0.022,
      0.0137, 0.0085, 0.0053)
g_range=range(0.0002,0.65)
plot(d,type="o",axes=F,col="blue", ylab="Probability",
     xlab="Day")
lines(i,type="o",pch=22,lty=2,col="red")
axis(1,at=1:10, lab=c("1", "2", "3", "4", "5", "6", "7", "8", "9",
                      "10"))
```

```
axis(2, las=1, at=0.05*12:g_range[2])
legend("topright", col=c("blue", "red"),lty=c(1,2),cex=1,
       legend=c("Decrease","Increase"))
box()

# Three state problem (FPT given current state is a small increase
#par(mfrow=c(2,2))
s = c(0.4355,0.1930,0.1288,0.0841,0.0550,0.0359,0.0235)
e = c(0.0099,0.0221,0.0227,0.0223,0.0218,0.0212,0.0207)
d=c(0.5547, 0.2464, 0.1101, 0.0492, 0.0220, 0.0098, 0.0044)
g_range=range(0.0042,0.56)
plot(s,type="o",axes=F,col="blue", ylab="Probability",
     xlab="Day")
lines(e,type="b",pch=19,lty=2,col="red", lwd=2)
lines(d,type="b",pch=22,lty=3,col="black", lwd=2)
axis(1,at=1:7, lab=c("1", "2", "3", "4", "5", "6", "7"))
axis(2, las=1, at=0.02*24:g_range[2])
legend("topright", col=c("blue", "red", "black"),lty=c(1,2,3),cex=1,
       legend=c("Small increase","Extreme increase", "Decrease"))
box()

# Three state problem (FPT given current state is an extreme increase
s = c(0.4900, 0.1760, 0.1157, 0.0756, 0.0494, 0.0323, 0.0211)
e = c(0.02, 0.0206, 0.0224, 0.0221, 0.0216, 0.0211, 0.0206)
d=c(0.49, 0.2816, 0.1264, 0.0565, 0.0252, 0.0113, 0.0050)
g_range=range(0.0042,0.50)

# Three state problem (FPT given current state is a decrease
s = c(0.3392, 0.2290, 0.1496, 0.0978, 0.0639, 0.0418, 0.0273)
e = c(0.0322, 0.0236, 0.0223, 0.0217, 0.0212, 0.0207, 0.0202)
d = c(0.6286, 0.2039, 0.0926, 0.0414, 0.0185, 0.0083, 0.0037)
g_range=range(0,0.60)

win.graph()
```

```
#par(mfrow=c(2,2))
p=c(0.9318, 0.4421, 0.2088, 0.0982, 0.0460, 0.0214, 0.0099, 0.0046)
g_range=range(0.0042,1.0)
plot(p,type="o",axes=F,col="blue", ylab="Probability",
     xlab="Extreme increases (MW)")
axis(1,at=1:8,
     lab=c("3000", "3200", "3400", "3600", "3800", "4000", "4200",
           "4400"))
axis(2, las=1, at=0.05*20:g_range[2])
box()

win.graph()
Pr=c(0.9318,  0.4421,  0.2088,0.0982,0.046,0.0214,0.0099,
     0.0046)
g_range=range(0.004,0.94)
plot(Pr,type="o",axes=F,col="blue", ylab="Probability",
     xlab="Extreme increases (MW)")
axis(1,at=1:8, lab=c("3000","3200","3400","3600","3800","4000",
                     "4200","4400"))
axis(2, las=1, at=0.05*20:g_range[2])
box()
```

Appendix B

Extreme mixture models R code

```
#-----------------------------------------------------------------------
#---------------------- Extreme mixture models --------------------
#-----------------------------------------------------------------------
attach(forecasts)
head(forecasts)
tail(forecasts)
win.graph()
plot(resid,xlab="Observation number",
     ylab="Residuals (megawatts)",col="blue")
#lines(smooth.spline(time(resid), resid, spar=1))
plot(density(resid))

library("evmix")
set.seed(20150924)
#par(mfrow = c(2, 1))
#x = resid

residuals=resid
xx = seq(-2500, 2500, 50)
y = dnorm(xx)
# Bulk model based tail fraction
fit = fgng(residuals)
```

```
hist(residuals, breaks = 100, freq = FALSE, xlim = c(-2500, 2500))
box()
lines(xx, y)
with(fit,lines(xx,lwd=2, lty = 2, dgng(xx, nmean, nsd, ul, sigmaul,
                                       xil, phiul,
                                       ur, sigmaur, xir, phiur),
               col="red"))
abline(v = c(fit$ul, fit$ur), lty = 2,col = "red")
fit

#----------------- Parameterised tail fraction ---------------------
fit2 = fgng(residuals, phiul = FALSE, phiur = FALSE)
with(fit2, lines(xx,lwd=2, lty = 1, dgng(xx, nmean, nsd, ul, sigmaul,
                                         xil, phiul,
                                         ur, sigmaur, xir, phiur),
                 col="blue"))
abline(v = c(fit2$ul, fit2$ur), col = "blue")
legend("topright", c("True Density","Bulk Tail Fraction",
                     "Parameterised Tail Fraction"),
       col=c("black", "red", "blue"), lty=1)
fit2

#------------ Diagnostic plots for assessing model fit---------------
evmix.diag(fit)
evmix.diag(fit2)
```

Appendix C

Simulation study R code

```
#-----------------------------------------------------------------------
#-------------------------- 1.SIMULATION STUDY ----------------------
#-----------------------------------------------------------------------
rm(list=ls())
library(R2OpenBUGS)
library(evir)

GPDmodel <- function(){#BUGS function (This will run in Bugs)
  for(i in 1:N) {
    y[i] ~ dgpar(0,sigma,evi)} #Your likelihood distribution
  sigma ~ dgamma(0.001,0.001)
  evi ~ dgamma(0.2,0.001)}
write.model(GPDmodel,'GPDmodel.txt')
data<-read.csv("forecasts.csv")
resids<-data$resid
resids<-resids[resids>0]#pick out the positive residuals
data<-sort(resids)
n<-length(data)
k<- length(data[data>=967]) # number of observations above the mean
threshold <- data[n-k]
abs.excesses  <- data[(n-k+1):n] - threshold
hill <- vector();for(j in 1:k){hill[j] <- mean(log(data[(n-j+1):n]))
```

```
                               - log(data[n-j]) }
InitialEVI  <- median(hill)
pars <- tryCatch({gpd(abs.excesses,0,method="ml")$par.est},
                 error=function(e){return(c(NA,NA))})
# BUGS run:
Fixed_data <- list(y=abs.excesses,N=k)
Initials <- function(){return (list(sigma=rchisq(1,1),
                                    evi=InitialEVI))}
GPDsims <- tryCatch({bugs(Fixed_data,Initials,'GPDmodel.txt',
                          parameters=c('evi','sigma'),
                          n.chains =5,n.iter=50000,n.burnin =2000,
                          n.thin =3,debug = FALSE)},
                    error=function(e){return("Crash")})

#---------------------------results----------------------------------
Est<-c(GPDsims$mean$evi,InitialEVI,pars[1])
names(Est)<-c("Bayes","Hill","MLE")
print(Est)

#--------------------------------------------------------------------
#------------------------ 2. EVI Simulations -----------------------
#--------------------------------------------------------------------
rm(list=ls())
library(R2OpenBUGS)
library(evir)

GPDmodel <- function(){#BUGS function (This will run in Bugs)
  for(i in 1:N) {
    y[i] ~ dgpar(0,sigma,evi)}
  sigma ~ dnorm(0.01,0.01)%_%T(0,)
  evi ~ dgamma(0.01,1)}
write.model(GPDmodel,'GPDmodel.txt')

ruNSamples <- function(sampleVars) {#Where all the magic happens
```

```
True_evi  <- 1/sampleVars[1]
n         <- sampleVars[2]
k         <- sampleVars[3]
data      <- sort((-log(runif(n)))^(-True_evi))
threshold <- data[n-k]
abs.excesses  <- data[(n-k+1):n] - threshold
hill <- vector();for(j in 1:k){hill[j] <- mean(log(data[(n-j+1):n]))
                                 - log(data[n-j]) }
InitialEVI  <- median(hill)
pars <- tryCatch({gpd(abs.excesses,0,method="ml")$par.ests},
                   error=function(e){return(c(NA,NA))})
  # BUGS run:
Fixed_data <- list(y=abs.excesses,N=k)
Initials <- function(){return (list(sigma=rchisq(1,1),
                                      evi=InitialEVI))}
GPDsims <- tryCatch({bugs(Fixed_data,Initials,'GPDmodel.txt'
                            ,parameters=c('evi','sigma'),
                            n.chains =5,n.iter=50000,n.burnin =2000,
                            n.thin =3,debug = TRUE)$mean$evi},
                      error=function(e){return("Crash")})
if(GPDsims!="Crash"){
    return(c(InitialEVI,GPDsims,pars[1]))
  }else{return(c(InitialEVI,NA,pars[1]))
  }
}
#-------------------------------------------------------------------
#------------ Running the Simulations in Parallel -----------------
#-------------------------------------------------------------------
NSamples <- 100 #number of simulations to do (or number of samples)
sampleVars <- matrix(c(2,1000,200),NSamples,3,byrow=T)
library(parallel)#Uses up all your computer cores...
Nodes<-detectCores()
cl <- makeCluster(Nodes)  #Initiate  your computer cores
clusterEvalQ(cl,{library(R2OpenBUGS)})
```

```
clusterEvalQ(cl,{library(evir)})
clusterExport(cl,c('ruNSamples','GPDmodel'))
clusterEvalQ(cl,{write.model(GPDmodel,'GPDmodel.txt')})
clusterExport(cl,"sampleVars")
cat("\n \n \n \n Simulating",NSamples,"samples of size",
    sampleVars[1,2],
    "from a Frechet distribution \n With EVI of 0.5 and
    estimating the EVI for k=",sampleVars[1,3],
    " observations above the threshold")
time<- system.time(results <- matrix(parRapply(cl,sampleVars,
                                               ruNSamples),3,
                                     NSamples))
stopCluster(cl)
Estims<-colMeans(t(results[1:3,]))
MSE <- colMeans((t(results[1:3,])-(1/matrix(sampleVars[,1],
                                            NSamples,3)))^2)
names(MSE)<-names(Estims)<- c("Median Hill","Bayes GPD","MLE GPD")
#========EVI======
print(Estims)
#========MSE======
print(MSE)
#========Time======
print(time)
```