**University of the Free State**
Faculty of Science and Agriculture
Department of Mathematical Statistics

# ON THE USE OF EXTREME VALUE THEORY IN ENERGY MARKETS

Promoter:
Prof. D de Waal,
Department of Mathematical Statistics,
University of the Free State

Thesis submitted for the PhD Degree in the Faculty of Science and Agriculture, Department of Mathematical Statistics, University of the Free State.

By: V Micali, Eskom

2004-2006
9 May 2007

# SUMMARY

The thesis intent is to provide a set of statistical methodologies in the field of Extreme Value Theory (EVT) with a particular application to energy losses, in Gigawatt-hours (GWh) experienced by electrical generating units (GU's).
Due to the complexity of the energy market, the thesis focuses on the volume loss only and does not expand into the price, cost or mixes thereof (although the strong relationship between volume and price is acknowledged by some initial work on the energy price [SMP] is provided in Appendix B)
Hence, occurrences of excessive unexpected energy losses incurred by these GU's formulate the problem.
Exploratory Data Analysis (EDA) structures the data and attempts at giving an indication on the categorisation of the excessive losses. The size of the GU failure is also investigated from an aggregated perspective to relate to the Generation System. Here the effect of concomitant variables (such as the Load Factor imposed by the market) is emphasised. Cluster Analysis (2-Way Joining) provided an initial categorising technique. EDA highlights the shortfall of a scientific approach to determine the answer to the question at when is a large loss sufficiently large that it affects the System. The usage of EVT shows that the GWh Losses tend to behave as a variable in the Fréchet domain of attraction. The Block Maxima (BM) and Peak-Over-Threshold (POT), the latter as semi and full parametric, methods are investigated. The POT methodologies are both applicable. Of particular interest is the Q-Q plots results on the semi-parametric POT method, which yielded results that fit

the data satisfactorily (pp 55-56). The Generalised Pareto Distribution (GPD) models well the tail of the GWh Losses above a threshold under the POT full parametric method. Different methodologies were explored in determining the parameters of the GPD. The method of 3-LM (linear combinations of Probability Weighted Moments) is used to arrive at initial estimates of the GPD parameters. A GPD is finally parameterised for the GWh Losses above 766 GWh. The Bayesian philosophy is also utilised in this thesis as it provides a predictive distribution of (high quantiles) the large GWh Losses. Results are found in this part of the thesis in so far that it utilises the ratio of the Mean Excess Function (the expectation of a loss above a certain threshold) over its probability of exceeding the threshold as an indicator to establish the minimum of this ratio. The technique was developed for the GPD by using the Fisher Information Matrix (FIM) and the Delta-Method. Prediction of high quantiles were done by using Markov Chain Monte Carlo (MCMC) and eliciting the GPD Maximal Data Information (MDI) prior. The last EVT methodology investigated in the thesis is the one that uses the Dirichlet process and the method of Negative Differential Entropy (NDE). The thesis also opened new areas of pertinent research.

Keywords:

Extreme Value Theory, Energy Markets, Gigawatt-hours Losses, Cluster Analysis (2-Way Joining), Q-Q Plots, Generalised Pareto Distribution, GPD Fisher Information Matrix, GPD Jeffreys' Prior.

# SAMEVATTING

Die doel van die skripsie is om ’n stel statistiese metodologië in die veld van Ekstreme Waarde Teorie te voorsien met ’n besonderse aanwending van verlore energie in Gigawatt-ure en wat ondervind is deur elektries-ontwikkelde eenhede.

As gevolg van die kompleksiteit van die energiemark, fokus die skripsie alleenlik op die volume verlies en nie op die pryskostes of die verhouding daarvan nie, alhoewel die sterk verhouding tussen volume en prys erken word aan die beginstadium van werk op die energieprys wat in Aanhangsel B voorsien word.

Hierna word verspreiding van buitensporige onverwagte verlore energie deur hierdie ontwikkelde eenhede blootgestel wat die probleem formuleer.

Verkennende data ontleding struktureer die data en pogings om ’n aanduiding te gee op die kategorisering van die oormatige verliese. Die grootte van die mislukte ontwikkelde eenheid is ook ondersoek vanuit ’n gesamentlike perspektief om die Opwekkingstelsel in verband te bring. Hier word die effek van gepaardgaande veranderlikes (soos wat die Gelaaide Faktor deur die mark voorgeskryf word) beklemtoon. Bondelontleding (2-Way Joining) het ’n aanvanklike kategorieserings tegniek voorsien. Verkennende data-ontleding lig die gebrek aan ’n wetenskaplike benadering uit om die antwoord op die vraag te bepaal wanneer ’n groot verlies grootgenoeg is om die Stelsel te beïnvloed.

Die gebuik van Ekstreemwaarde-teorie toon dat die GW-ure verliese geneig is om as ’n veranderlike in te tree in die Fréchet gebied.

Die Blok-Maksima en “Peak-Over Threshold” (POT) metodes, laasgenoemde as half en vol parametriese metodes, is ondersoek. Die POT metodologië is beide bruikbaar. Uit besonderse belangstelling lewer die Q-Q voorstellings van die half parametriese POT metode, goeie resultate. Die Veralgemeende Pareto-verdeling (GPD) modeleer die stert van GWh-verliese bokant a drempelwaarde onder POT goed. Verskillende metodologië was ondersoek deur die bepaling van parameters van die GPD. Die metode van 3-LM (lineêr kombinasies van die Waarskynlikheid-Geweegde-Momente-metode ) is gebruik as `n eerste skatting van die GPD parameters. ’n GPD is finaal geparameteriseerd vir die GW-ure verliese bo 766 GW-ure. Die Bayes-filisofie is ook gebruik in hierdie skripsie en voorsien ’n voorspellingsfunksie van (hoë kwantiele) van groot GW-ure verliese. Nuwe werk is in hierdie gedeelte van die skripsie gedoen in soverre dit die gebruik van die verhouding van die gemiddelde-oorskrydings-funksie relatief tot `n oorskrydingswaarskynlikheid as ’n aanwysing om die minimum van hierdie verhouding te vestig. Die tegniek was ontwikkel vir die GPD deur die Fisher-informasiematriks en die Delta-metode te gebruik. Voorspelling van hoë kwantiele is deur die gebruik van MCMC gedoen en die MDI prior vir die GPD is gebruik. Die laaste Ekstreemwaarde metodologie wat in die skripsie ondersoek is, is die een wat die Dirichlet proses en die metode van Negatiewe-Afgeleide-Entropie gebruik.

Die skripsie open ook nuwe areas vir gepaste navorsing.

Sleutelwoorde:
Ekstreemwaarde-teorie, Energiemarkte, Gigawature-verliese, Bondelanaliese (2-Way Joining), Q-Q voorstellings, Veralgemeende Paretoverdeling, GPD Fisher-informasiematriks, GPD Jeffreys' prior.

## AKNOWLEDGEMENTS

First, I thank God for giving me the strength and health to fulfil my duties and dreams. Secondly, I thank my wife, Nevana and children, Daniel, Kyara, Stuart-Lloyd and Steven, for their patience in sustaining their love for me throughout my studies and career.

My gratitude goes also to numerous persons for the reasons given next to their names.

Ehud Matya, my Managing Director, for approving my studies.

Brian Statham, my General Manager, for elevating my potential, for applying my work at the highest possible level in the organisation and recommending my research.

My gratitude goes also to Dr Terry Moss as my mentor in the executive, national and international arenas.

My colleagues, Kiren Maharaj, Robert Kydd, Gerhard Loedolff and Tony Moore for enabling me to expand my knowledge beyond the realm of Statistics.

A very special vote of gratitude goes to my supervisor Professor Danie de Waal, Department of Mathematical Statistics at the University of the Free State (UFS), for his relentless guidance and time devotion in the expert direction that culminated in this thesis.

I would also like to thank Professor Max Finkelstein (UFS), for valuable comments in the references and very interesting material necessary for the research support in some stages of the thesis.

My sincere appreciation to Dr Martin van Zyl (UFS), for some developments of this thesis.

To Ms E Mathee, I am grateful for the administrative support at the UFS

Last but not least, I would like to thank the examiners for toiling through this thesis.

# TABLE OF CONTENTS

**List of Figures**

Appendix A



**List of Tables**



Appendix A

# LIST OF ABBREVIATIONS

**EdF**: Electricité de France, the largest electricity utility in France.

**Eskom** or **EHL**: Eskom Holdings (Ltd), the largest electricity utility in the African continent. It supplies approximately 95% of the electric energy needs in South Africa.

**Extreme Value Theory (EVT)**: Theory that predicts the occurrence of rare events, outside the range of available data.

**GenCo**: Generating Company. Generally comprises of at least one power Station, and at most having no more than 33% of the total national Installed Capacity.

**Generating Unit (GU)**: here a generating unit is defined as the industrial unit spanning from the fuel provision to the electricity generator. A set of GU's make up a Power Plant or Power Station.

**Installed Capacity (IC)**: Nominal Capacity of a GU or total number of Megawatt installed in a Power Plant or System of Power Plants.

**MW**: Megawatts, one million Watts, the Watt being the unit of power.
**MWh**: Megawatt-hours, one million Watt-hours, the Watt-hour being the unit of energy [Gigawatt-hours

(**GWh**) = one billion Watt-hours; Terawatt-hours (**TWh**) = one trillion Watt-hours]

**V@R**: The Value at Risk is a single estimate of the amount by which an institution's position in a risk category could decline due to general market movements during a given holding period. Developed by JP Morgan and referenced in RiskMetrics (registered trademark of JP Morgan). Other indicators in the Energy Markets are:

**E@R**: Earnings at Risk

**CF@R**: Cash Flow at Risk

**P@R**: Profit at Risk, and coined in this thesis,

**Vol@R**: Volume at Risk, pronounced "volar"

**ROI**: Return on investment.

**Union of the Electricity Industry/Eurelectric (UEI/Eurelectric)**: International body based in Europe and HQ in Brussels (Belgium). After the creation of the European Union, it was formed by the amalgamation of UNIPEDE and EURELECTRIC.

**UEI/Eurelectric Nomenclature**

**EAF**: Energy Availability Factor, the generating unit availability after discounting the UCF with other losses. EAF% = UCF% - OCLF%
**EUF**: Energy Utilisation Factor represents the loading (in MW) imposed on a GU relative to that GU availability (EAF).
**LF**: Load Factor is the loading (in MW) imposed on a GU relative to its Installed Capacity during a specified period. LF% = EAF% x EUF%
**OCLF**: Other Capability Loss Factor, this is other energy losses incurred by a generating unit, outside the management control (e.g. flooding of the coal mines), expressed as the energy lost divided by the total potential energy (IC x t, where t is time in hours for the period).
**PCLF**: Planned Capability Loss Factor, energy losses of a GU due to maintenance, expressed as the energy lost divided by the total potential energy (IC x t, where t is time in hours for the period).
**UCF**: Unit Capability Factor, the capability of the generating unit after discounting planned (PCLF) and unplanned losses (UCLF). UCF% = 100% - UCLF% - PCLF%
**UCLF**: Unplanned Capability Loss Factor, this is the forced outage rate incurred by a generating unit. This is the forced outages incurred by a GU, expressed as the energy lost divided by the total potential energy (IC x t, where t is time in hours for the period).

# EAGLE VIEW of the APPROACH taken on this Thesis

- Observations
  - The occurrence of excessive "unexpected" energy losses incurred by GU formulates the problem.
  - The Energy Market platform and its participants forms a complex environment
  - Relationships and linkage exist between prices, fuel, volume underlyings

- Problem setting
  - Limitations: field too wide, hence focus on energy volume losses, not on other underlyings.
  - What is the meaning of "excessive"?
  - Can these losses be categorised?
  - Can a scientific method be used to determine at what level is an energy loss "excessive"?
  - What are the latest statistical techniques that may be used to determine the probabilities of these "excessive" losses?

- Hypothesis*****
  - Null Hypothesis: Utilisation of Extreme Value Theory (EVT) methods resolves the questions above
  - Alternate Hypothesis 1: EVT partially answers the questions above
  - Alternate Hypothesis 2: EVT does not answer the questions above

- Theory
    - Assumptions that the EVT techniques are sound and robust unless otherwise stated (by means of research)
    - This is an applied thesis; it utilises the academic theories already developed.
    - Research various EVT techniques (includes literature surveys)
    - Assure that replication of the experiment (within the sampled population) is possible when using these techniques.

- Experimentation
    - Data collection, formatting and filtering (MWh Losses per GU p.a.)
    - Exploratory Data Analysis (EDA) to understand data behaviour
    - Testing of Hypothesis on various EVT techniques
    - Prognosis

**** These Hypotheses are not to be confused with the Hypothesis Testing techniques in Inference; they are just part of the eagle-view process taken in the planning of the thesis***.

# PREFACE

There are wide arrays of new issues facing today's energy analysts since the recent changes in the energy industry. To be able to understand these issues, it is necessary to understand well the Energy Market and its dynamics, in terms of its **platform** and the involvement of its **participants**.

The energy industry **platform**'s restructuring and growing competition (internationally as well as in the Southern African market) elements are putting pressure on thorough analyses and prediction of market share, of market price, of fuel delivery and volume, of the competing generating companies' (GenCo's) costs. This type of market liberalization spawned different market domains that are inter-related: electricity, weather, coal, water, environment and ancillary services. The future challenge of this "new world" is their transformation from public service assets into commodities that can be traded in a similar manner as those in the capital and money markets. The impetus has sadly been slowed down by the unfortunate Enron (mankind's greed for money) saga in December 2001 and has undergone its own changing phases from speculative positions to hedging ones and to physical asset-based trading approaches. Volume and/or price exposures are some of the afflicting risks for these GenCo's. Limiting these exposures and trading to optimise profitability of the generators has become one of the GenCo's major goals.

To qualify the platform further, distinction is made in this thesis between a generating company (GenCo), a transmission company (TransCo) and a distribution company (DistCo). Whereby for some, e.g. in Eskom's case, as a Holdings company, it would see the three as integrated (Divisions). The platforms referred to above are rather more from a generic perspective than one particular to Eskom. In this thesis, the platform is from a GenCo viewpoint, in terms of its production of electricity and sale to various DistCo's or through a brokerage type company (e.g. Key Sales and Customers Division in Eskom), via a TransCo and possibly hedged over the counter or through a Power Exchange. When specifically addressing the Eskom case, the Generation perspective would be modelled as an integrated Company whereby the competing mode would be internal in terms of volume and cost-of-sale.

**Participants** in the energy markets can be characterised as operating between two diagrammatically opposite approaches: the trading-centric approach and the asset-centric approach.

The participant at the trading-centric extreme would tend to operate as in the money and capital markets. That is, to maximise their Mark-to-Market gains within the constraint of their trading limits. The GenCo's have inherent physical positions such as production and, hence, these positions turn into equivalent trading transactions and become part of the trading portfolio. In this case, the ability to deliver good sustainable results becomes generally the primary profit driver.

The asset-centric participant would tend to operate its assets with the objective to deliver a sustainable ROI. They support in sweating the physical operation as much as possible and back the GenCo's with effective hedging positions of volume risks (and at times regulated-price, by means of "claw-backs" contracts). Long term forward contracts, physical bi-laterals and the like, usually highlight their positions. These are at times boosted by option type derivatives which then induce them to operate in a similar fashion as the trading-centric participants. The asset availability of the core physical operation becomes its primary profit driver.
GenCo's participants, in this thesis, are essentially asset-centric.

Although recently the energy markets are taking far more prudent positions in shifting from being trading-centric towards being asset-centric (thanks to Enron's top management), this might be an extremist move from one side of the scale to the other. Indeed it could turn out a very hazardous affair. In Eskom's case the price is regulated therefore as a subset of an asset-centric participant, the primary drivers would then be the supply of primary-energy (fuel: coal, water, nuclear, kerosene, wind) and the volume produced to meet the demand.

The preceding paragraphs are given to illustrate that the business of producing electricity is quite complex in the interrelationship of offers, bids, price, fuel and volume available as well as the customers' demand for electricity.
After reflecting on different interesting aspects of extremes in the platforms given above, and before

hitting back on the subject of this thesis, I trust that the scene has been set to embark on the energy markets ***oceans***, an analogy given to illustrate the vastness and depth of these platforms.
Hence, within this analogy, the liberty is taken with another illustration, namely, one of a large Armada standing as a fleet of generating units (GU's) subjected to the <u>extreme</u> storms (events), exposed to <u>extreme</u> drops in wind – no wind (energy volume losses – zero prices) and struggling to make it into harbor with the cargo (i.e. being business viable).
Therefore, it is the very nature of these extremes that originated the title of the thesis:

"***On the Use of Extreme Value Theory in Energy Markets***".

Extreme Value Theory, or EVT, a field in itself in the Statistical Sciences, required initial research, and hence from an historical statistical perspective, it is important to take a step back before leaping "the fleet" forward.

In 1902, Prof. K Pearson, wrote a "Note on Francis Galton's Difference Problem" in Biometrika Vol. 1. on the problem of the range of samples. This seems to have engaged the interest of a young scientist working for the British Cotton Industry Research Association, a Mr. L.H.C. [35] Tippett, BSc London. But it wasn't to end there, it was merely the beginning as the "Roaring Twenties" seem to have produced more that just the "Charleston", namely: L. v. Bortkiewicz, "Variationabreite und mittlerer Fehler" Oct 1921; E. L. Dodd "The Greatest and Least Variate under General Laws of Error" Oct 1923, J. S-Nyman "Sur les valeurs théoriques de la plus grande des erreurs" 1924, J.O. Iawif "The Further Theory of Francis

Galton's Individual Difference Problem", 1925. In 1925, Tippett published an article titled "On the Extreme Individuals and the Range of Samples Taken from a Normal Population". There he mentions that the complete solution to the problem of the extreme values had not yet been resolved by Bortkiewicz, Dodd, Nyman and Iawif. In that paper, he publishes his attempts in closing the deficiencies.

It is my belief that the seed of Extreme Value Theory (EVT) originated with Tippett's publication in 1925, although later work with RA Fisher [15] published in 1928 is the most quoted reference (Fisher-Tippet Theorem); eventhough WE Fuller in 1914, on "Flood Flows" [16] and AA Griffith [18], in 1920 with "The phenomena of rupture and flow in solids" started to approach the EVT subject from an empirical position (see Chapter 3's Introduction).

Tippet remains, from the author's perspective, the "seed" originator of EVT in 1925.

Leaping the "fleet" forward: this thesis addresses EVT on the energy platforms with particular reference to the electricity market. Although the application of EVT is also applicable to energy prices, and some aspects are covered due to some interrelationship, this thesis primarily concentrates on the volume exposure within the generation of electricity platform as asset-centric participants. Due to the thesis being more of an applied nature rather than theoretical, [2] and [3] Beirlant is the favorite choice as a basis referential point.

Since EVT applies to extremes in the minimum and maximum domains, this thesis concentrates on the maximum aspect of EVT (unless otherwise stated).

As mentioned above, since the thesis is essentially structured from an applied perspective, most theories from respected sources are accepted within their assumptions. Hence the thesis reflects the application of these theories within the philosophy of extracting *raw data*, analyzing and modelling it to churn it from sterile data into *information* and packing it to distil the information into *knowledge* for decision making purposes, thereby adding value to the process.

The work in this thesis is structured in five Chapters and three Appendices containing the following elements.

***In Chapter 1, an introduction of the electricity generation environment is given. It portrays a statement of the problem in terms of the type of events occurring in the generation of electricity by, illustrating the complexity of the energy market in which these types of events occur. In it, the Extreme Value Theory methodology as a possible solution to model the behaviour is discussed.***

***Chapter 2 begins with the process of exploratory data analysis, the intricacy of the data is illustrated and a starting point is given by using a multivariate method (Cluster Analysis) to get an indication various category losses. Large values of losses are then "binned" in these categories to explore further the data behaviour. The number of events in terms of GUs failing is discussed since this form the assets of the Company and expectations of number of GUs failing in the various bins are given.***

***Also in this Chapter, the Size of Volume failure, in the (GWh Losses), is investigated and portrayed as an EVT problem from a GU and a System (aggregation of GUs) perspective.***
***It highlights the importance of a concomitant variable (the Load Factor imposed by the Market) as it exercises its influence on the volume losses from System perspective.***
***In summary, the Chapter reflects the channeling (i.e. introduces, after exploring the behaviour of the data) into the modelling proposed for various stages of the proposed solution (in Chapter 3), and the research needed for validation of the hypothesis, that is the* utilisation of Extreme Value Theory *(EVT) methods to resolve the questions posed (see the Eagle View, page vii).***

***<u>Chapter 3</u> is fully devoted to the application of EVT to the GWh Losses. It contains the application of EVT as described in Appendices A and C; it explores the application of the Block Maxima and the usage of the Gumbel Distribution. It then moves onto the Peak Over Threshold methodologies with particular reference to the Generalised Pareto Distribution as a model for the GWh Losses and explores if all the questions can be answered by this method.***

***<u>Chapter 4</u> is dedicated to the Bayesian philosophy of tackling the questions and sees how this method approached the unanswered questions. It gives a different perspective of fitting the GPD in terms of the GWh Losses. It also provides "new" methodologies in the Energy sphere to predict the large values of the losses (High Quantiles). One of these methods, the Dirichlet Entropy one, gives an***

***innovative scientific way of determining the level at which the GWh Losses form part of the high extremes.***

***The Conclusion is given in <u>Chapter 5</u>, wherein the value of EVT's research into the GWh energy losses and its results are summarised.***

***<u>Appendix A</u> contains some essential statistical aspects of EVT that are used in the main body of the thesis.***
***<u>Appendix B</u> provides some statistical background on prices returns in the Energy Markets and may be used for further research in this area***
***<u>Appendix C</u> holds all kinds of statistical work encountered in the research incurred for this thesis; it also considers some important work on EVT. This Appendix takes into account some additional relevant remarks and observations.***

# CHAPTER 1

# Introduction

## 1.1 The Energy Market Environment

Energy markets are increasingly being liberalised throughout the world. The introduction of competition, especially in the generating sector of the electricity supply industry has caused plant management to seek performance indicators that not only reflect technical excellence and commercial performance, but also the risk management aspect of the plant. A generating unit (here a unit is defined as the industrial unit spanning from the fuel provision to the electricity generator) could significantly improve its viability by managing its availability so as to be producing electricity when needed and at the right value. In other words, a unit's availability is worth more during certain hours than in others. Hence, of primary importance, is the technical availability of the energy generating plant. The failure of a generating unit (GU) in a GenCo affects this technical availability directly. The time frame by which this GU is out of production, can be classified in terms of its risk profile. In other words, it could be: as expected (tolerable), better than expected, exceptionally better than expected, or from the other side, worse than expected, much worse than expected, considered as a *Main/Major Event*, *Semi-Catastrophic* or *Catastrophic*.

Utilities consisting of large (>400 MW) generating units (numbering from 12 –as a GenCo to 100 –as EdF, say) are subject to events that range from major (availability loss of a unit for approx. 1 to 3 months, say) to catastrophic (> 9 months, say). The impact of these events in missing opportunities or being not competitive in this market highlights the importance of the work that needs to be done.

Exposure to any of these events will cause a substantial knock on revenue to the utilities. Hence it is also of vital importance for utilities to hedge this position vis-à-vis these events (e.g. revenue or volume hedges, associated with E@R, CF@R and Vol@R).

In the Eskom Generation context, the smallest GU is a gas-turbine type with an Installed Capacity (IC) of 57 MW, the largest fossil fuel GU at 669 MW and the largest nuclear GU at 900MW, with a total Generation IC being 36 208 MW. In 2005, there were a total of 84 GU's commercially available for generating, 64 of which are considered as the "Coal Fired" fleet.

Hence, although the probability is very small, the maximum possible energy loss for any particular hour, summated over the GU's, is 36 208 MWh, the maximum possible energy loss for a day is 868 992 MWh (or 868.992 GWh) or for a year (leap) approximately 318 TWh. What this means is that for a particular window of time, the distribution is bounded in its true sense. However, because these boundaries are highly improbable relatively to the extreme events exposed to, for all practical purposes, the probability density functions utilised are considered unbounded in this thesis. Nevertheless, these GU's raise questions about these extreme events, from a worse than expected situation to a catastrophic one.

More of this part is expanded in paragraph 1.2; however to illustrate the inter-relationship and for the sake of completeness, the aspect of revenue exposure explanation is also given below

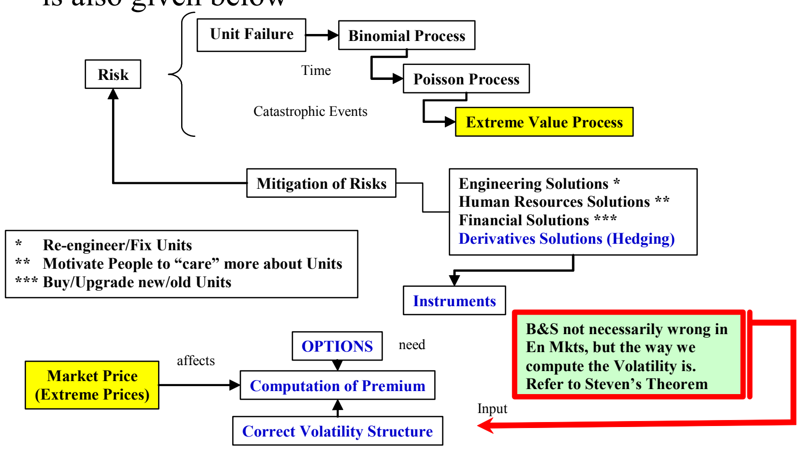



**Fig 1.1: Schematic Diagram of EVT with respect to GU's failures**

To exemplify how EVT maps onto the exposures of production (volume) and revenue, a diagram is produced and shown in Fig 1.1.
The top right hand side illustrates the risk of energy losses with its type of processes. The position of EVT is also indicated therein (in Yellow). It also shows possible avenues to mitigate this type of volume risk.

.

The bottom left hand side of Fig 1.1 shows the exposure to the financial losses due to extreme prices. In a price-response market, the customer would reject the purchase of energy, while in a captive market the GenCo's may reflect revenue losses (due to unavailability of the GU's and missing the opportunity when the prices are high). The GenCo's may also lose their licence from a National Regulating body (such as FERC in the USA or NERSA in RSA) if found irresponsible in "spiking" the prices. The GenCo's might be also responsible for inducing high credit risks, as their customers might go bankrupt (due to electricity bills that are exorbitantly high) and hence reducing the GenCo to a loss in revenue. These can be mitigated (as shown) by means of derivatives, such as Options. However, the recurring random price shocks would induce the premiums to be very high if using the Black & Scholes (B&S) model to price the Option. This does not mean that the structure of the B&S model is necessarily wrong for the energy markets, but that the methodology to compute the volatility for returns in the energy markets needs to be fundamentally reviewed (see Appendix B – REMARKS, at the end of the 3rd paragraph). Steven's Theorem [26] Micali , proposes that the Returns are Cauchy distributed under certain assumptions (see Appendix C paragraph 4.2). As the volatility increases to very high levels (to 40% as typical in the money markets, to 4000% in the energy markets), it makes this kind of property (Cauchy), in terms of the volatility definition, unreasonable for usage with the Black & Scholes (B&S) model to price Options in the energy markets. A similar discourse is also reflected in other works by [27] Moore, [1] Alexander, [5] Cootner and [24] Mandelbrot.

The treatment of EVT for electricity prices, merits a separate thesis in its own right; here, the mere linkage to the volumetric losses is introduced from a statistical modelling approach (see Appendix B).

To limit the vastness of this platform, only the Coal Fired plant population is considered in this thesis. The Coal Fired stations form approximately 90% of the total Generation IC and unless specified, data is considered per GU over one-year window time periods. The data was collected in its raw form in terms of the records provided and filed in electronic format using Excel. It was filtered to a formatted spreadsheet to assess missing values or GU's not commissioned. The Excel working files names are given in Appendix C in the EDA paragraph.

As expressed above, this is where statistical modelling kicks-in and with appropriate statistical tools, scientific deduction (information) ought to be made from the data. Hence, [37] Tukey's Exploratory Data Analysis (EDA) finds its way in the beginning, even (or especially) in the Extreme Value Theory methodology used. EDA facilitates a researcher to use basic statistical techniques (such as scatter-plots, brushing, distribution fitting, contingency tables …) in order to have a preliminary knowledge of the data behaviour. After having a clear understanding of such, using EDA, Multivariate Exploratory techniques (in this case, Cluster Analysis) would enhance the insight of how to organise the energy losses into meaningful structures. For instance, it would enable a researcher to classify the energy losses into categories that range from expected, main/major events, semi-catastrophic, and catastrophic, say. Particular care was taken in the data collection and

compilation of the statistical data files by using Data Mining methodology; this is given in Appendix C.
Further insight is given from a statistical perspective on the number of GU's failing and the size of the loss (see Fig 1.1 above top RHS). This forms an evolutionary research after having performed EDA (Cluster Analysis) to explore further the Poisson behaviour (number of GU's failing per year) and the size of large MWH Losses incurred. This work is covered in Chapter 2. This follows into the natural progression of the usage of EVT in Chapter 3 with a more classical frequentist approach. Since EVT forms the kernel of the thesis, a basic overview is already provided in that chapter, while further details are supplied in Appendix A and C. Chapter 4 is a must, with its Bayesian approach, as it reveals the answers that would be left unresolved, in the author's opinion, in Chapter 3.
Chapter 5 concludes the thesis with a summary describing if the questions posed were answered and what opportunities were uncovered.

## 1.2 Basics of Extreme Value Theory

EVT, albeit a new technique, has been used extensively in the Actuarial and Financial Sciences. Its particular usefulness in the Risk Management field is becoming more and more evident as time goes on. It is expected to provide management with robust risk measures in decision taking.
The EVT technique essentially concentrates at the behaviour of the tails of a distribution. Before introducing the EVT technique, it is important to understand that there are numerous ways other than EVT to resolve the problem of long (fat) tails. These are: Student's T [39] Wilson,

Mixture of Normals [22] Hull et al, Generalised Error Distribution (GED) [29] Nelson , however, in this thesis, EVT is the technique used to deal with the problem of fat tails. Fat tails are defined here as tails of a greater thickness than normal. Even so, some distributions with fat tails exhibit longer tails than others (see Fig. 1.2). The effect of these "longer" tails is important in estimating the frequencies as correctly as possible, especially in the understanding the effects of low frequency (small occurrences) but high impact events. For instance, in terms of the relative frequency difference, an excess frequency can be, say, 2000 times larger than the normally thought of frequency. Therefore, the consistent deficiency of the relative frequency (long tail), calls for different modelling and hence, EVT [32] Reiss and the risk costs could be high.

Classic EVT can be classified in at least two groups. Both these groups divide the data into consecutive windows. These groups are: the Block Method, which focuses on the maxima within these windows and the Peak Over Threshold (POT) method, which focuses on the events that exceed a certain threshold; these are then modelled separately from the events below the threshold [13] Embrechts.

The Central Limit Theorem (CLT) plays an important role in seeding the basis of EVT. The CLT tries to find constants $a_n > 0$ and $b_n$, such that $(S_n - b_n)/ a_n$ tends to a (non-degenerate) standard normal as $n \to \infty$. In this case, $S_n = \sum (X_i)$, $b_n = n.E(X)$ and
$a_n = (n.Var(X))^{1/2}$.

In general, the Central Limit method can be applied to determine the distributions that may be obtained at the limits. However when the parental (underlying) distribution has very long tails, the method may yield distributions with infinite variances and, hence, non-normal limits for the estimate of the expectation [2] Beirlant. These factors are important in the probabilistic aspect of EVT when considering maxima or minima (instead of the expectation mentioned in the previous phrase), whereby one would replace $M_n = max\{X_1, X_2, \ldots, X_n)$ and replace the ***Central Limit*** method by the ***Extremal Limit*** method (see Appendix A for a more detailed rendition, Eq. A2.0).

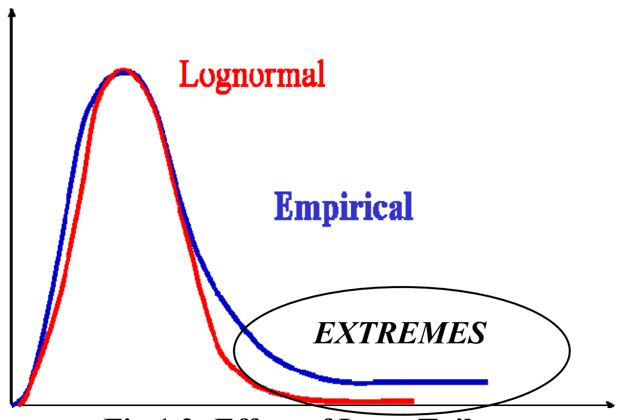



**Fig 1.2: Effect of Long Tails**

## 1.2.1 Maximum Domain of Attraction

A logical progression would now be to move onto the Maximum Domain of Attraction (MDA – see Appendix A for details, Eq A2.1) to link the parental (underlying) distribution to EVT. This is better illustrated by means of an example.

For instance, if one has a variable $X$ with a parental distribution that follows an exponential pdf (e.g. $X \equiv$ the mean time to fail for each year) what would be the extreme value distribution ***attracted*** to the pdf of $X$?

An answer could be construed in the following manner:
The pdf of $X$ is $f_X(x) = \lambda.exp(-\lambda x)$, for $\lambda > 0$ and $x \geq 0$.
Hence, its CDF (of $X$) is:
$F_X(x) = 1- exp(-\lambda x)$, as given (see Appendix A Eq. A1.0).
By taking $a_n = 1/\lambda$ and $b_n = \{Ln\ (n)\ /\ \lambda\}$ and substituting in A2.0 (in the Appendix), we get:
$P((M_n - b_n) / a_n \leq x) = P(\{\lambda M_n - Ln\ (n)\} \leq x)$
$= P(M_n \leq \{x + Ln\ (n)\}/ \lambda) = F^n(\{x + Ln\ (n)\}/ \lambda)$
$\rightarrow G_X(x)$, as $n \rightarrow \infty$ (see A2.0)
Now, $F_X(x) = 1- exp(-\lambda x)$
$\therefore\ F^n((\{x + Ln\ (n)\}/ \lambda) = [F_X(\{x + Ln\ (n)\}/ \lambda)]^n$
by substitution in the CDF above
$= [1- exp(-\lambda\{x + Ln\ (n)\}/ \lambda))\ ]^n$
$= [1- exp(-\{x + Ln\ (n)\})\ ]^n$
$\therefore\ P(M_n \leq \{x + Ln\ (n)\}/ \lambda)$
$= [1- exp(-\{x + Ln\ (n)\})\ ]^n$ Eq 1.0
$= [1- \{exp(-x).\ exp(-Ln\ (n))\}\ ]^n$
$= [1- \{exp(-x).\ 1/n\}\ ]^n$ Eq 1.1

But $exp(a) = [1 + a/n\ ]^n$ as $n \rightarrow \infty$,
hence let $a = -exp(-x)$ and substitute in Eq 1.1, we get:
$P(M_n \leq \{x + Ln\ (n)\}/ \lambda) = [1+ \{a/n\}\ ]^n = exp(a)$,
but $a = -exp(-x)$

$\therefore\ P(M_n \leq \{x + Ln\ (n)\}/\ \lambda) = G_X(x) = exp(-exp(-x))$
which is of a Gumbel form (see A2.5 in Appendix)

Hence, the Exponential distribution lies in the domain of attraction of a Gumbel extreme value distribution.

The following distributions lie in the **Gumbel domain** [2] Beirlant:
Benktander II, Weibull, Exponential, Gamma, Logistic, Normal, Log-Normal.

Following the same line of thought, for the other two families of the Generalised Extreme Value Distribution (GEV), see A2.2 in the Appendix, from [2] Beirlant, we get the following distributions in the **Weibull domain**:
Uniform, Beta, Reversed Burr, Extreme Value Weibull.

And in the **Fréchet domain**:
Pareto, Generalised Pareto, Burr (XII), Burr (III), Cauchy, F-distribution, Inverse Gamma, Log Gamma, Fréchet, T-distribution.

### 1.2.2 GEV Parametric Estimation

There are numerous methods in estimating the GEV distributions parameters; from the "Naïve" to the complex. The most important ones, especially in estimating the Extreme Value Index (EVI), are reflected below, (see Appendix A par. 2.1 and 2.3):

- Maximum Likelihood Estimation (MLE)
- Pickands Estimation (PE) method
- Hill Estimation (HE) method
- Regular Variation Approach (RVA)
- Dekkers-Einmahl-deHaan Estimation (DEHE) method
- Zipf Estimation (ZE) method
- Probability Weighted Moments (PWM) method
- L-Moments method
- Bayesian method

### 1.2.3 Entropy Method

The challenge in EVT is the determination of the probability density function from which the samples have been drawn. This determination is dependent on the EVT parameters which in turn depend on the threshold. However, once that aspect has been assessed, gauging the level(s) of threshold(s) becomes a new challenge. Here, one could use, say, the Mean Excess Function method or the Entropy method.

Entropy is more familiar in the world of thermodynamics, physics and chemistry, and although it has a certain mystique and cannot be directly measured, its occurrence can be inferred by changes in its variables. Maxwell, Boltzmann and Gibbs extended the work of thermodynamics in what today is termed Statistical Mechanics. In the latter, the macrostate variable is considered as an expression of a function of microstate variables, or expressed mathematically [33] Shannon:

$$\mathcal{H}(\boldsymbol{p}) = -\lambda \sum_{i}^{k} \left[ p_i \,.Log(p_i) \right] \qquad \text{Eq 1.2}$$

Where, $\boldsymbol{p} = \{p_1, p_2, \ldots, p_k\}$ is the probability density vector of element I (see Appendix C paragraph 2.8 for more details).

Equation 1.2 can be reduced to:

$$\mathcal{H}(\boldsymbol{p}) = \lambda\{Log(k)\} \qquad \text{Eq 1.3}$$

As an example: Let $\lambda = 1$, in these cases,

Case 1: a GU has equal probability of failing or not

Case 2: a GU fails with a probability of 0.03 and operates with a probability of 0.97

In Case 1, $\mathcal{H}(\boldsymbol{p}) = 1.000$ (using Eq 1.3), and in

Case 2, $\mathcal{H}(\boldsymbol{p}) = 0.194$ (using Eq 1.2)

The **larger the Entropy**, the **more unpredictable the outcome**, as illustrated in Fig. 1.3

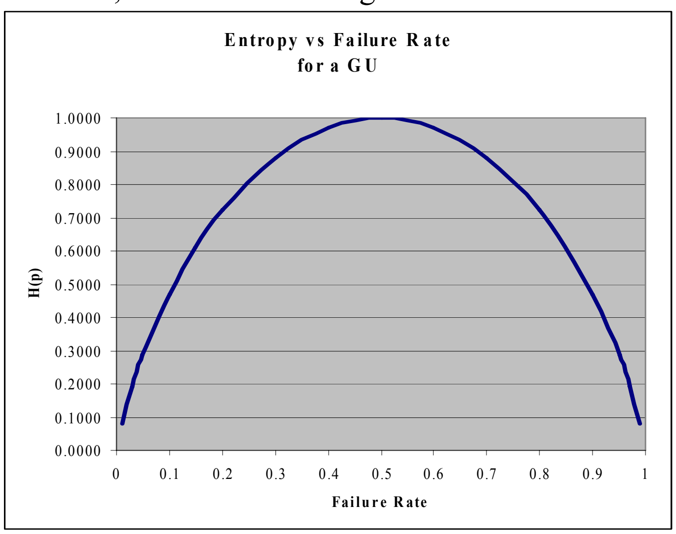



**Fig 1.3: Entropy of a GU failure rate**

In EVT, this type technique of using the statistical entropy with Dirichlet distributional properties is usually employed within the Peak Over Threshold (POT - see Appendix A) method. It uses different entropy values to estimate the number of points to be used in the tail of the distribution so as to assess the threshold level (the lower the entropy, the better the predictability of the threshold level).
More on this methodology is reflected in Chapter 4 paragraph 4.5.

# CHAPTER 2
# Exploratory Data Analysis (EDA)

Multivariate exploratory techniques are included in this section. EDA is closely related to the concept of Data Mining. EDA is used to identify particular behaviours in variables when there are no prior expectations of those behaviours. When using EDA techniques, many variables are scrutinized, in the search for coherent patterns. Scatter plots techniques are included in the relevant paragraphs

## 2.1 Data Structure

The data was formatted according to Appendix C – EDA paragraph
The Table below (2.1), reflects a Sub-sample of the data set acquired it also shows the Power Station (PS) abbreviation code (e.g. Arnot PS: AR) and the GU's within that Station, e.g. Arnot has 6 GU's as well as Duvha PS, while Hendrina PS has 10 GU's. The 1$^{st}$ Row indicates the No. of hours in a year, taking into account leap years, this was used for the computation of the energy in MWh from the Capacity (in MW) loss.
The 2$^{nd}$ Row gives the years that the sample of the population reflected.

| | | 8760 | 8760 | 8784 | 8760 | 8760 | 8760 | 8784 | 8760 | 8760 | 8760 | 8784 | 8760 | 8760 | 8760 | 8784 | 8760 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **STAT** | **UNIT** | **1990** | **1991** | **1992** | **1993** | **1994** | **1995** | **1996** | **1997** | **1998** | **1999** | **2000** | **2001** | **2002** | **2003** | **2004** | **2005** |
| AR | 1 | 251788.7 | 768085.6 | 680909.3 | 180675 | 37580.4 | 63019.44 | 54785.81 | 99732.6 | 44518.32 | 74871.72 | 42611.18 | 27173.52 | 102334.3 | 266531.8 | 40292.21 | 347763.2 |
| AR | 2 | 277227.7 | 475825.7 | 311032.7 | 245139.8 | 161884.8 | 288501.8 | 119137.4 | 104068.8 | 44229.24 | 23704.56 | 126384.2 | 46252.8 | 218255.4 | 65910.24 | 246970.9 | 758835 |
| AR | 3 | 424080.4 | 492014.2 | 350455.2 | 117366.5 | 80942.4 | 62730.36 | 7536.672 | 39314.88 | 130375.1 | 13586.76 | 6087.312 | 62441.28 | 23993.64 | 17055.72 | 144356.3 | 346896 |
| AR | 4 | 91927.44 | 228084.1 | 324366.8 | | | | | | 93661.92 | 18501.12 | 14493.6 | 41338.44 | 37869.48 | 122859 | 20291.04 | 89614.8 |
| AR | 5 | 225193.3 | 392859.7 | 82033.78 | | | | | | | 115632 | 352774.2 | 23415.48 | 33244.2 | 33533.28 | 155371.4 | 36135 |
| AR | 6 | 692635.7 | 379273 | 302046.6 | | | | | | | | | 53768.88 | 32087.88 | 51745.32 | 48698.5 | 15899.4 |
| DU | 1 | 679491.3 | 557092.2 | 241933.3 | 669921 | 277035 | 1591692 | 517201.9 | 44829.3 | 213568.8 | 529892.4 | 80307.72 | 23673.9 | 25688.7 | 33244.2 | 50002.92 | 32740.5 |
| DU | 2 | 370219.5 | 200472.6 | 551042.3 | 216591 | 80592 | 1188732 | 30809.88 | 15111 | 160176.6 | 337982.7 | 120714.1 | 20148 | 18636.9 | 4939786 | 739437.1 | 86636.4 |
| DU | 3 | 304738.5 | 503700 | 111117.6 | 206517 | 347553 | 15111 | 135361.4 | 379286.1 | 123406.5 | 219109.5 | 74246.76 | 45333 | 19644.3 | 13096.2 | 52528.32 | 81599.4 |
| DU | 4 | 951993 | 356619.6 | 348505.2 | 347553 | 831105 | 191406 | 796006.1 | 23673.9 | 34755.3 | 72029.1 | 393962.4 | 114843.6 | 318338.4 | 27199.8 | 117683.6 | 36770.1 |
| DU | 5 | 952496.7 | 676469.1 | 866212.2 | 1395249 | 85629 | 166221 | 155059.6 | 149095.2 | 244798.2 | 139021.2 | 156069.7 | 10074 | 57421.8 | 32236.8 | 109097.3 | 23170.2 |
| DU | 6 | 223642.8 | 789297.9 | 600540.1 | 266961 | 25185 | 1405323 | 1296035 | 115347.3 | 116354.7 | 124413.9 | 56568.96 | 75555 | 4533.3 | 182339.4 | 67680.72 | 34251.6 |
| HE | 1 | 269965.7 | 261643.7 | 782408.4 | 88712.52 | 98199.6 | 120336.1 | 42224.69 | 10319.28 | 1497.96 | 64245.84 | 48733.63 | 32955.12 | 80556.96 | 89877.6 | 24366.82 | 81722.04 |
| HE | 2 | 306083.2 | 539598.5 | 162222.9 | 174262.7 | 163111.2 | 123165.6 | 131013.4 | 79391.88 | 12982.32 | 132819.1 | 36717.12 | 65078.04 | 110849 | 189741.6 | 103141.7 | 48600.48 |
| HE | 3 | 250325.8 | 191572.4 | 486001.2 | 95370.12 | 149796 | 29959.2 | 136520.9 | 113012.8 | 128325.2 | 121501.2 | 53406.72 | 33288 | 168270.8 | 76562.4 | 81779.04 | 120502.6 |
| HE | 4 | 196565.6 | 162944.8 | 359660.9 | 91542 | 81555.6 | 123165.6 | 38219.18 | 1830.84 | 6657.6 | 500651.5 | 46063.3 | 5492.52 | 68240.4 | 94704.36 | 91625.9 | 123498.5 |
| HE | 5 | 133152 | 289938.5 | 366336.7 | 317234.6 | 226358.4 | 190407.4 | 92961.07 | 101861.3 | 58420.44 | 66076.68 | 167396.7 | 65244.48 | 125995.1 | 54425.88 | 46563.98 | 191905.3 |
| HE | 6 | 175760.6 | 249660 | 185087.7 | 163111.2 | 211378.8 | 51929.28 | 131347.2 | 81555.6 | 10652.16 | 17975.52 | 82279.73 | 26963.28 | 7822.68 | 21803.64 | 47565.36 | 87381 |
| HE | 7 | 162944.8 | 553246.6 | 236658.5 | 620321.9 | 133152 | 76229.52 | 409229 | 35784.6 | 63413.64 | 29127 | 216964.8 | 18474.84 | 18474.84 | 24300.24 | 28372.32 | 123165.6 |
| HE | 8 | 315736.7 | 44605.92 | 153711.2 | 83053.56 | 53260.8 | 64245.84 | 101472.8 | 122499.8 | 10485.72 | 25631.76 | 6008.256 | 70903.44 | 61582.8 | 122999.2 | 43559.86 | 73732.92 |
| HE | 9 | 201059.5 | 155455 | 88454.88 | 74232.24 | 76562.4 | 75730.2 | 48733.63 | 30125.64 | 27296.16 | 33121.56 | 18525.46 | 17309.76 | 65910.24 | 52095.72 | 18358.56 | 44273.04 |
| HE | 10 | 269133.5 | 75230.88 | 229148.2 | | 26630.4 | 38114.76 | 99303.12 | 58919.76 | 185072.5 | 392995.5 | 56226.38 | 70820.22 | 42297.66 | 31601.7 | 52976.3 | 41811.48 |
| KE | 1 | 718740.5 | 599324.2 | 638631.9 | 232105 | 347596.8 | 62231.04 | 79828.99 | 101475.8 | 69519.36 | 10652.16 | 65774.59 | 14016 | 37002.24 | 164828.2 | 69147.65 | 90263.04 |
| KE | 2 | | 437299.2 | 556554.2 | 31956.48 | 26910.72 | 149130.2 | 87699.46 | 63912.96 | 62791.68 | 2242.56 | 44974.08 | 602688 | 22425.6 | 56064 | 13492.22 | 76247.04 |
| KE | 3 | | | 201259 | 17379.84 | 22425.6 | 91944.96 | 153474 | 151372.8 | 56064 | 7288.32 | 11243.52 | 12894.72 | 15137.28 | 1360673 | 112997.4 | 113249.3 |
| KE | 4 | | | 469417 | 63912.96 | 28032 | 127265.3 | 145041.4 | 144084.5 | 27471.36 | 25228.8 | 283336.7 | 78489.6 | 1121.28 | 103157.8 | 82639.87 | 28592.64 |
| KE | 5 | | | | 46533.12 | 11212.8 | 37562.88 | 183269.4 | 103157.8 | 10091.52 | 120537.6 | 41601.02 | 3924.48 | 57185.28 | 159782.4 | 12930.05 | 173798.4 |
| KE | 6 | | | | | 22425.6 | 72322.56 | 56779.78 | 560.64 | 302745.6 | 31956.48 | 73645.06 | 21304.32 | 2242.56 | 250045.4 | 20238.34 | 24668.16 |
| KR | 1 | 563815.5 | 138561.3 | 132265.1 | 52844.7 | 815556 | 694470.9 | 35048.16 | 165191.7 | 157285.8 | 144802.8 | 71765.28 | 170601 | 70320.9 | 1090182 | 415571 | 186828.9 |
| KR | 2 | 760214.7 | 495159 | 640463.4 | 309578.4 | 557574 | 86548.8 | 35048.16 | 282531.9 | 28710.9 | 129823.2 | 0 | 80307.3 | 116924.1 | 177674.7 | 120165.1 | 155205.3 |
| KR | 3 | 755221.5 | 153957 | 186506.3 | 339953.7 | 615828 | 22053.3 | 78023.88 | 157701.9 | 98199.6 | 131071.5 | 180247.7 | 72401.4 | 20388.9 | 99031.8 | 155630.5 | 439817.7 |
| KR | 4 | 840105.9 | 915420 | 196937.3 | 119420.7 | 174762 | 64079.4 | 83865.24 | 64079.4 | 51596.4 | 86964.9 | 211123.4 | 76562.4 | 46603.2 | 815556 | 86368.68 | 148131.6 |
| KR | 5 | 734416.5 | 264223.5 | 388033.2 | 310410.6 | 449388 | 42026.1 | 210289 | 93622.5 | 182667.9 | 96951.3 | 58413.6 | 71985.3 | 151460.4 | 130655.4 | 969665.8 | 194318.7 |
| KR | 6 | 659518.5 | 79475.1 | 66758.4 | 175178.1 | 158118 | 42442.2 | 165227 | 148963.8 | 151460.4 | 74065.8 | 124754.8 | 43274.4 | 91125.9 | 116924.1 | 941710.7 | 256317.6 |

**Table 2.1: MWh Losses actuals 1990 to 2005 -Sub-sample**

The matrix contains the MWh Losses for every GU for every year. The "black" blocks, originally missing data, were investigated further. These yielded the following information. The large ones at the top & bottom, were not missing information, but legitimate beahaviour of the population viz. the GU's at Arnot PS were "mothballed" for 5 years and then gradually brought back into service. The GU's at Kendal PS were new GU's being commissioned gradually into the system. Hendrina GU 10 was truly a missing value for 1993 and unfortunately valid records were not available (this was the only true missing value – the others could be explained in a similar fashion as above). The data matrix of 16 yrs x 64 GU's = 1024 data points of which 929 could be used in univariate mode (approx 91%).

| STATION | STAT Code | UNIT | IC |
|---|---|---|---|
| AR | C1 | Ar1 | 330 |
| AR | C2 | Ar2 | 330 |
| AR | C3 | Ar3 | 330 |
| AR | C4 | Ar4 | 330 |
| AR | C5 | Ar5 | 330 |
| AR | C6 | Ar6 | 330 |
| DU | C7 | Du1 | 575 |
| DU | C8 | Du2 | 575 |
| DU | C9 | Du3 | 575 |
| DU | C10 | Du4 | 575 |
| DU | C11 | Du5 | 575 |
| DU | C12 | Du6 | 575 |
| HE | C13 | He1 | 190 |
| HE | C14 | He2 | 190 |
| HE | C15 | He3 | 190 |
| HE | C16 | He4 | 190 |
| HE | C17 | He5 | 190 |
| HE | C18 | He6 | 190 |
| HE | C19 | He7 | 190 |
| HE | C20 | He8 | 190 |
| HE | C21 | He9 | 190 |
| HE | C22 | He10 | 185 |
| KE | C23 | Ke1 | 640 |
| KE | C24 | Ke2 | 640 |
| KE | C25 | Ke3 | 640 |
| KE | C26 | Ke4 | 640 |
| KE | C27 | Ke5 | 640 |
| KE | C28 | Ke6 | 640 |
| KR | C29 | Kr1 | 475 |
| KR | C30 | Kr2 | 475 |
| KR | C31 | Kr3 | 475 |
| KR | C32 | Kr4 | 475 |
| KR | C33 | Kr5 | 475 |
| KR | C34 | Kr6 | 475 |
| LE | C35 | Le1 | 593 |
| LE | C36 | Le2 | 593 |
| LE | C37 | Le3 | 593 |
| LE | C38 | Le4 | 593 |
| LE | C39 | Le5 | 593 |
| LE | C40 | Le6 | 593 |
| MJ | C41 | Mj1 | 612 |
| MJ | C42 | Mj2 | 612 |
| MJ | C43 | Mj3 | 612 |
| MJ | C44 | Mj4 | 669 |
| MJ | C45 | Mj5 | 669 |
| MJ | C46 | Mj6 | 669 |
| MB | C47 | Mb1 | 615 |
| MB | C48 | Mb2 | 615 |
| MB | C49 | Mb3 | 615 |
| MB | C50 | Mb4 | 615 |
| MB | C51 | Mb5 | 615 |
| MB | C52 | Mb6 | 615 |
| ML | C53 | Ml1 | 575 |
| ML | C54 | Ml2 | 575 |
| ML | C55 | Ml3 | 575 |
| ML | C56 | Ml4 | 575 |
| ML | C57 | Ml5 | 575 |
| ML | C58 | Ml6 | 575 |
| TU | C59 | Tu1 | 585 |
| TU | C60 | Tu2 | 585 |
| TU | C61 | Tu3 | 585 |
| TU | C62 | Tu4 | 585 |
| TU | C63 | Tu5 | 585 |
| TU | C64 | Tu6 | 585 |

**Table 2.2: Stations, Coding, GU's & Installed Capacity**

Table 2.2 gives the coding used for each GU with its relevant information used in multivariate techniques, e.g. Stat Code C8 corresponds to GU 2 at Duvha PS with an Installed Capacity (IC) of 575 MW.

## 2.2 Cluster Analysis

Cluster Analysis (CA) helps in organising the MWh Losses of the GU's into meaningful categories. This multivariate method uses distances (dissimilarities) between GU's MWh Losses when forming the Clusters. The usual statistical way used to calculate the distances is the Euclidean method or essentially the geometric distance in a multidimensional space:

$$\delta(x;y) = [\Sigma_i \ (x_i - y_i)^2 \ ]^{0.5}$$

The Chebychev distance was the one chosen as one would want to detect the GU's MWh Losses if they were different on anyone of the years in question, in this case the distance is modelled as: $\delta(x;y) = Max\,|x_i - y_i|$

Once these distances have been computed a linkage rule is required to link the Clusters which are sufficiently similar. The most common one is called Single Linkage or Nearest Neighbour; in this case, the distance between the Clusters is determined by the distance of the closest GU's MWh Losses (Nearest Neighbours) in the different Clusters.

The rule of Complete Linkage (Furthest Neighbour) was used here whereby the distance between the Clusters is determined by the greatest distance between any two MWh Losses in the different Clusters. This technique is particularly useful if the MWh Losses would tend to bunch-up.

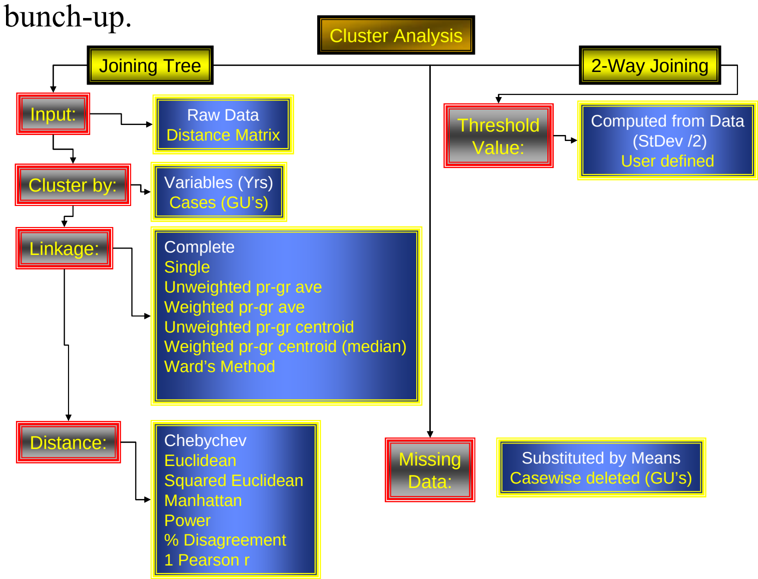



**Fig 2.1: Cluster Analysis-Diagrammatic parameters selection**

Case wise (Row) deletion could have been used to handle Missing Data. However, the substitution by Means methods was deployed in order not to reduce the data matrix size. Both methods were tested and no difference was detected at the nodes; there was a marginal change in the cluster selection at the lower end ("the better years").

First Step was to use the Joining Tree method and Fig 2.1 above illustrates the parameters selected (in White over the Blue background) for the analysis. The software used was STATISTICA. The outputs are shown in Figs 2.2 and 2.3 below as well as Table 2.3.

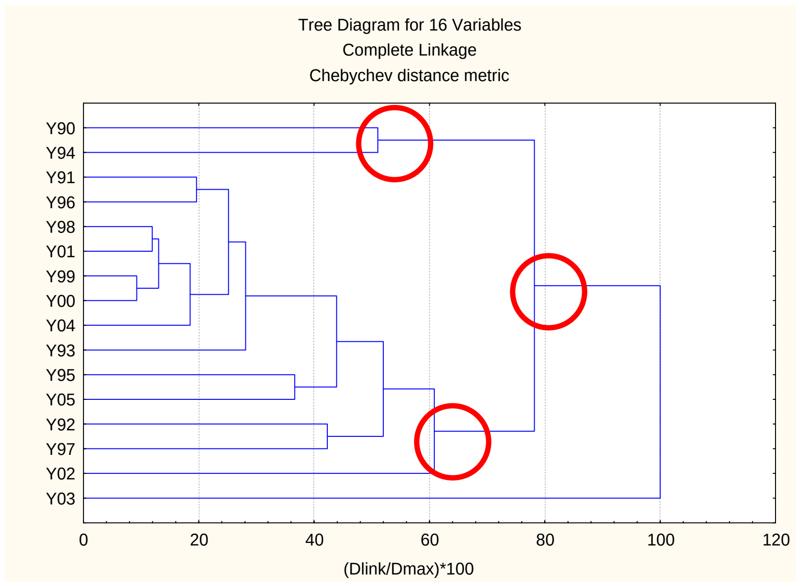



**Fig 2.2: Joining Tree Method of Clustering**

In the hierarchical tree of Fig 2.2 above three nodes (in red) can be identified. This tends to imply that Y03 (Year 2003) stands on its own i.t.o. the GU's MWh Losses. The next links are Y02 (2002) and both Y90 (1990) and Y94 (1994). This kind of categorisation invoked interest and a more detailed method of 2-Way Joining in Cluster Analysis was used (see Fig 2.1 above). In this analysis the same pattern was observed but more information was now available (see Fig 2.3 and Table 2.3 below)

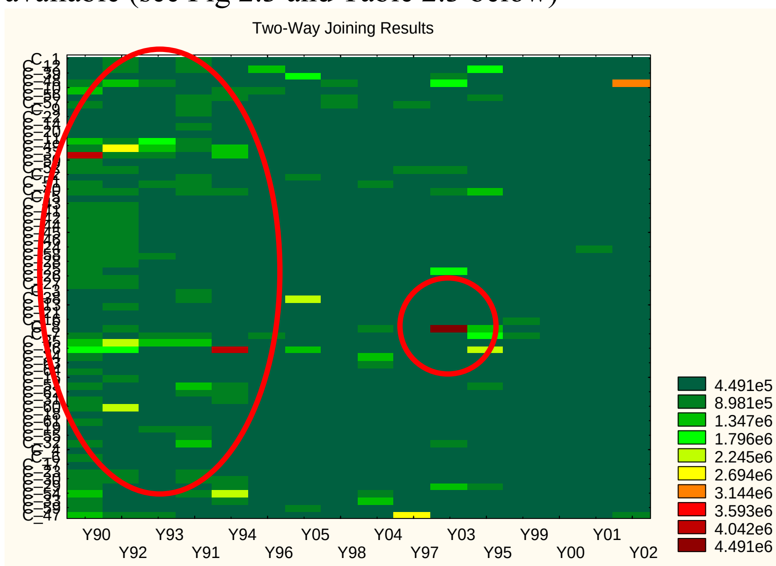



**Fig 2.3: 2-Way Joining Method of Clustering (Legend in MWh)**

From the re-ordered table (Table 2.3 below) and Fig 2.3 (highlights in red circles) above, it can be deduced that large MWh Losses incurred in those particular years (values highlighted in the Table); this was not evident from

Fig 2.2 that the nodal clustering was due to these very large values.
This was an enabling finding as it was now providing a basis for categorising these extreme events. After inspecting the data, there seemed to be a natural 900 000 MWh (or 900 GWh) spread amongst the categories.
The reasons for the high losses in 2003 were due to a catastrophic event due to fire at one 575 MW GU.
The GWh Losses were then classified into six categories ranging from Sub-critical to Catastrophic and these are illustrated in Fig 2.4 below.
It must be noted, from the legend in Fig 2.3, the following GWh categorical values:

**449; 899; 1348; 1797; 2246; 2695; 3144; 3593; 4042; 4491**

The values highlighted in red will bear significance later in Chapter 3 in the analysis of Extremes.

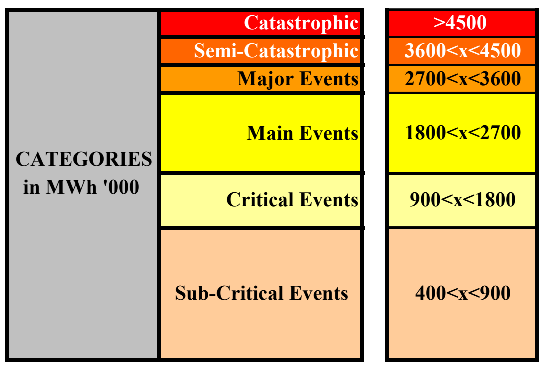



**Fig 2.4: Classification of Large MWh (‘000), i.e. GWh, Losses**

The results of the data classification are given in Table 2.4 below.

However caution needs to be exerted in the interpretation of the losses within the classification. For instance a GU out for a whole year at Hendrina Power Station would be classified differently from a GU out for the same period at Majuba Power Station as shown below.

| | HE | MJ |
|---|---|---|
| Normal Year | 1,620,600 | 5,860,440 |
| Leap Year | 1,625,040 | 5,876,496 |

This means that for a catastrophic event to occur at Hendrina (equivalent to 1 GU out at Majuba), 3 GU's would be out for a year and 1 GU for half a year. This is according to the statistical Classification; however from a managerial perspective, 1 GU out at Hendrina, although it would classify as Critical, it might actually be just as "Catastrophic", since Hendrina is one of the "cheap" and base-loaded Power Stations.

This may be rectified after the management/executive would decide on the severity of the impact of the loss and then assign weights to different GU's or/and Power Stations. Using the example above, if the management, say, felt that 1 GU at Hendrina lost for 1 year is equivalent to 1 GU loss at Majuba also for 1 year, whatever reason given, then the "Loss Score" would be Loss @ Hendrina x 3.5. These can then be classified in the table illustrated in Fig 2.4.

These weights (e.g. 3.5) can also be normalised for all the GU's so that the Loss Score for the System be determined.

| | Le3 | | | | Le2 | | | | | | Du2 | | | | | Mb2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Y90 | Y92 | Y93 | Y91 | Y94 | Y96 | Y05 | Y98 | Y04 | Y97 | Y03 | Y95 | Y99 | Y00 | Y01 | Y02 |
| C_1 | 251789 | 680909 | 180675 | 768086 | 37580 | 54786 | 347763 | 44518 | 40292 | 99733 | 266532 | 63019 | 74872 | 42611 | 27174 | 102334 |
| C_12 | 223643 | 600540 | 266961 | 789298 | 25185 | 1296035 | 34252 | 116355 | 67681 | 115347 | 182339 | 1405323 | 124414 | 56569 | 75555 | 4533 |
| C_39 | 161035 | 375042 | 25973 | 72726 | 93504 | 76571 | 1560482 | 166749 | 14064 | 161555 | 473235 | 98699 | 105971 | 38025 | 25973 | 23376 |
| C_48 | 600695 | 1027491 | 549515 | 59800 | 37712 | 113986 | 35018 | 589382 | 71849 | 295768 | 1592516 | 172397 | 111519 | 95078 | 85121 | 3030951 |
| C_10 | 951993 | 348505 | 347553 | 356620 | 831105 | 796006 | 36770 | 34755 | 117684 | 23674 | 27200 | 191406 | 72029 | 393962 | 114844 | 318338 |
| C_56 | 420086 | 411640 | 206517 | 797861 | 871401 | 144453 | 30222 | 651788 | 66165 | 76562 | 73540 | 520826 | 99229 | 9091 | 131969 | 65985 |
| C_57 | 710217 | 334868 | 171258 | 530900 | 312294 | 123240 | 330427 | 628114 | 118189 | 787283 | 82607 | 173273 | 100740 | 87379 | 40800 | 46340 |
| C_9 | 304739 | 111118 | 206517 | 503700 | 347553 | 135361 | 81599 | 123407 | 52528 | 379286 | 13096 | 15111 | 219110 | 74247 | 45333 | 19644 |
| C_22 | 269133 | 229148 | 268169 | 75231 | 26630 | 99303 | 41811 | 185073 | 52976 | 58920 | 31602 | 38115 | 392996 | 56226 | 70820 | 42298 |
| C_14 | 306083 | 162223 | 174263 | 539599 | 163111 | 131013 | 48600 | 12982 | 103142 | 79392 | 189742 | 123166 | 132819 | 36717 | 65078 | 110849 |
| C_20 | 315737 | 153711 | 83054 | 44606 | 53261 | 101473 | 73733 | 10486 | 43560 | 122500 | 122999 | 64246 | 25632 | 6008 | 70903 | 61583 |
| C_11 | 952497 | 866212 | 1395249 | 676469 | 85629 | 155060 | 23170 | 244798 | 109097 | 149095 | 32237 | 166221 | 139021 | 156070 | 10074 | 57422 |
| C_49 | 538740 | 2277011 | 1082867 | 634097 | 1093642 | 82653 | 20472 | 103438 | 103181 | 194485 | 108287 | 209031 | 174552 | 49700 | 36634 | 43099 |
| C_37 | 3724586 | 645905 | 467521 | 337654 | 1210361 | 105220 | 31688 | 123633 | 32295 | 83634 | 287266 | 249345 | 212462 | 58861 | 87271 | 4156 |
| C_50 | 525272 | 217707 | 48487 | 237584 | 185327 | 39436 | 19395 | 143305 | 61585 | 101822 | 278529 | 183172 | 34479 | 14046 | 84582 | 39328 |
| C_52 | 576091 | 675810 | 53874 | 414435 | 316240 | 69688 | 17778 | 89431 | 62665 | 662650 | 481095 | 149770 | 47948 | 72929 | 41483 | 102899 |
| C_2 | 277228 | 311033 | 245140 | 475826 | 161885 | 119137 | 758835 | 44229 | 246971 | 104069 | 65910 | 288502 | 23705 | 126384 | 46253 | 218255 |
| C_51 | 576091 | 398679 | 506416 | 572681 | 177784 | 82113 | 83505 | 82427 | 466747 | 94818 | 67343 | 28014 | 44177 | 22689 | 54413 | 69497 |
| C_40 | 576091 | 697994 | 192203 | 706477 | 664919 | 41150 | 45194 | 122075 | 303680 | 163113 | 594271 | 993223 | 519 | 80738 | 40519 | 18181 |
| C_5 | 225193 | 82034 | 268169 | 392860 | 373534 | 148121 | 36135 | 124795 | 155371 | 169872 | 33533 | 252572 | 115632 | 352774 | 23415 | 33244 |
| C_43 | 576091 | 481091 | 268169 | 414435 | 373534 | 148121 | 44497 | 124795 | 162887 | 169872 | 126522 | 252572 | 174773 | 188153 | 375815 | 78808 |
| C_41 | 576091 | 481091 | 268169 | 414435 | 373534 | 148121 | 232137 | 7506 | 101065 | 64333 | 36456 | 252572 | 6969 | 21503 | 536 | 115395 |
| C_42 | 576091 | 481091 | 268169 | 414435 | 373534 | 148121 | 33775 | 10722 | 120418 | 169872 | 42889 | 252572 | 2681 | 103753 | 13939 | 51467 |
| C_44 | 576091 | 481091 | 268169 | 414435 | 373534 | 148121 | 205115 | 124795 | 42311 | 169872 | 15237 | 252572 | 107011 | 14691 | 50986 | 49228 |
| C_45 | 576091 | 481091 | 268169 | 414435 | 373534 | 148121 | 60363 | 124795 | 32908 | 169872 | 24028 | 252572 | 107011 | 81998 | 2930 | 120139 |
| C_46 | 576091 | 481091 | 268169 | 414435 | 373534 | 148121 | 11721 | 124795 | 25269 | 169872 | 13479 | 252572 | 107011 | 81998 | 69515 | 104902 |
| C_24 | 576091 | 556554 | 31956 | 437299 | 26911 | 87699 | 76247 | 62792 | 13492 | 63913 | 56064 | 149130 | 2243 | 44974 | 602688 | 22426 |
| C_58 | 555077 | 487402 | 800883 | 390368 | 337479 | 171222 | 17126 | 89155 | 45457 | 128444 | 42311 | 193925 | 74548 | 48488 | 61451 | 43318 |
| C_28 | 576091 | 481091 | 268169 | 414435 | 22426 | 56780 | 24668 | 302746 | 20238 | 561 | 250045 | 72323 | 31956 | 73645 | 21304 | 2243 |
| C_25 | 576091 | 201259 | 17380 | 414435 | 22426 | 153474 | 113249 | 56064 | 112997 | 151373 | 1360673 | 91945 | 7288 | 11244 | 12895 | 15137 |
| C_26 | 576091 | 469417 | 63913 | 414435 | 28032 | 145041 | 28593 | 27471 | 82640 | 144084 | 103158 | 127265 | 25229 | 283337 | 78490 | 1121 |
| C_27 | 576091 | 481091 | 46533 | 414435 | 11213 | 183269 | 173798 | 10092 | 12930 | 103158 | 159782 | 37563 | 120538 | 41601 | 3924 | 57185 |
| C_3 | 424080 | 350455 | 117366 | 492014 | 80942 | 7537 | 346896 | 130375 | 144356 | 39315 | 17056 | 62730 | 13587 | 6087 | 62441 | 23994 |
| C_38 | 129867 | 156267 | 166230 | 472716 | 98699 | 1563 | 1883072 | 23376 | 521 | 34285 | 63375 | 79479 | 131425 | 39588 | 54544 | 27012 |
| C_13 | 269966 | 782408 | 88713 | 261644 | 98200 | 42225 | 81722 | 1498 | 24367 | 10319 | 89878 | 120336 | 64246 | 48734 | 32955 | 80557 |
| C_21 | 201060 | 88455 | 74232 | 155455 | 76562 | 48734 | 44273 | 27296 | 18359 | 30126 | 52096 | 75730 | 33122 | 18525 | 17310 | 65910 |
| C_16 | 196566 | 359661 | 91542 | 162945 | 81556 | 38219 | 123498 | 6658 | 91626 | 1831 | 94704 | 123166 | 500652 | 46063 | 5493 | 68240 |
| C_8 | 370220 | 551042 | 216591 | 200473 | 80592 | 30810 | 86636 | 160177 | 739437 | 15111 | 4939786 | 1188732 | 337983 | 120714 | 20148 | 18637 |
| C_7 | 679491 | 241933 | 669921 | 557092 | 277035 | 517202 | 32741 | 213569 | 50003 | 44829 | 33244 | 1591692 | 529892 | 80308 | 23674 | 25689 |
| C_35 | 1038936 | 1989804 | 909069 | 1106467 | 254539 | 212524 | 283110 | 50388 | 3125 | 77401 | 42596 | 45713 | 139737 | 40109 | 24934 | 50908 |
| C_36 | 1449316 | 1552256 | 168827 | 394796 | 3854453 | 146891 | 1133479 | 200515 | 367749 | 73764 | 18181 | 2183844 | 22337 | 67195 | 22337 | 3636 |
| C_34 | 659519 | 66758 | 175178 | 79475 | 158118 | 165227 | 256318 | 151460 | 941711 | 148964 | 116924 | 42442 | 74066 | 124755 | 43274 | 91126 |
| C_63 | 47659 | 154159 | 139902 | 25623 | 153738 | 25693 | 79944 | 177311 | 482518 | 65595 | 109666 | 19473 | 98392 | 28263 | 17424 | 48684 |
| C_64 | 576091 | 158784 | 133240 | 16399 | 97367 | 103287 | 12299 | 33310 | 172658 | 157325 | 76357 | 68670 | 74307 | 44706 | 427904 | 38947 |
| C_15 | 250326 | 486001 | 95370 | 191572 | 149796 | 136521 | 120503 | 128325 | 81779 | 113013 | 76562 | 29959 | 121501 | 53407 | 33288 | 168271 |
| C_53 | 789802 | 346990 | 45333 | 1071874 | 831105 | 105562 | 74548 | 99229 | 112128 | 45837 | 83111 | 663373 | 173273 | 30305 | 4030 | 7052 |
| C_62 | 330537 | 409036 | 51246 | 736405 | 450965 | 20555 | 74819 | 113254 | 154159 | 73282 | 88143 | 43559 | 234194 | 24665 | 76357 | 57396 |
| C_31 | 755222 | 186506 | 339954 | 153957 | 615828 | 78024 | 439818 | 98200 | 135631 | 157702 | 99032 | 22053 | 131072 | 180248 | 72401 | 20389 |
| C_60 | 757416 | 1862243 | 179361 | 142976 | 169112 | 155187 | 431491 | 95318 | 139771 | 34847 | 50221 | 124528 | 52271 | 9250 | 41509 | 122478 |
| C_18 | 175761 | 185088 | 163111 | 249660 | 211379 | 131347 | 87381 | 10652 | 47565 | 81556 | 21804 | 51929 | 17976 | 82280 | 26963 | 7823 |
| C_61 | 721031 | 59094 | 148613 | 220358 | 138364 | 141826 | 94293 | 31773 | 46762 | 12299 | 107104 | 33822 | 16399 | 43678 | 46634 | 32285 |
| C_19 | 162945 | 236659 | 620322 | 553247 | 133152 | 409229 | 123166 | 63414 | 28372 | 35785 | 24300 | 76230 | 29127 | 216965 | 18475 | 18475 |
| C_55 | 276531 | 191425 | 50370 | 635166 | 125925 | 109097 | 353094 | 10074 | 31315 | 164710 | 6548 | 27704 | 61955 | 164151 | 34252 | 106784 |
| C_32 | 840106 | 196937 | 119421 | 915420 | 174762 | 83865 | 148132 | 51596 | 86369 | 64079 | 815556 | 64079 | 86965 | 211123 | 76562 | 46603 |
| C_4 | 91927 | 324367 | 268169 | 228084 | 373534 | 148121 | 89615 | 93662 | 20291 | 169872 | 122859 | 252572 | 18501 | 14494 | 41338 | 37869 |
| C_6 | 692636 | 302047 | 268169 | 379273 | 373534 | 148121 | 15899 | 124795 | 48698 | 169872 | 51745 | 252572 | 107011 | 81998 | 53769 | 32088 |
| C_17 | 133152 | 366337 | 317235 | 289938 | 226358 | 92961 | 191905 | 58420 | 46564 | 101861 | 54426 | 190407 | 66077 | 167397 | 65244 | 125995 |
| C_23 | 718741 | 638632 | 232105 | 599324 | 347597 | 79829 | 90263 | 69519 | 69148 | 101476 | 164828 | 62231 | 10652 | 65775 | 14016 | 37002 |
| C_30 | 760215 | 640463 | 309578 | 495139 | 557574 | 35048 | 155205 | 28711 | 120165 | 282532 | 177675 | 86549 | 129823 | 0 | 80307 | 116924 |
| C_29 | 563816 | 132265 | 52845 | 138561 | 815556 | 35048 | 186829 | 157286 | 415571 | 165192 | 1090182 | 694471 | 144803 | 71765 | 170601 | 70321 |
| C_54 | 928823 | 220720 | 287109 | 740943 | 2231391 | 43437 | 128444 | 130962 | 548517 | 143051 | 385331 | 143051 | 151110 | 80813 | 138518 | 58429 |
| C_33 | 734417 | 388033 | 310411 | 264224 | 449388 | 210289 | 194319 | 182668 | 969666 | 93623 | 130655 | 42026 | 96951 | 58414 | 71985 | 151460 |
| C_59 | 467364 | 357135 | 160912 | 357697 | 261355 | 139771 | 462239 | 144001 | 326304 | 126065 | 250080 | 47146 | 111204 | 143882 | 109666 | 21011 |
| C_47 | 1297286 | 636915 | 614164 | 276374 | 199334 | 234994 | 143305 | 411597 | 328451 | 2607502 | 196640 | 81350 | 49025 | 56723 | 47948 | 529581 |

**Table 2.3: Re-ordered MWh Losses from 2-Way Method of Clustering**

.

| CATEGORIES in MWh '000 | | | 1990 | 1991 | 1992 | 1993 | 1994 |
|---|---|---|---|---|---|---|---|
| | UNIT Years | 930 | 49 | 53 | 56 | 53 | 55 |
| | Catastrophic | >4500 | | | | | |
| | Semi-Catastrophic | 3600<x<4500 | 3,724,586 | | | | 3,854,453 |
| | Major Events | 2700<x<3600 | | | | | |
| | Main Events | 1800<x<2700 | | | 2,277,010 | | 2,231,391 |
| | | | | | 1,989,804 | | |
| | | | | | 1,862,243 | | |
| | Critical Events | 900<x<1800 | 1,449,316 | | 1,552,256 | 1,395,249 | |
| | | | | | | | |
| | Sub-Critical Events | 400<x<900 | | 768,086 | 782,408 | 620,322 | |
| | | | | 539,598 | 486,001 | | |
| | | | | 553,247 | | | |
| | | | | | | | |
| | | | | | | | |

**1990 - 1994**

| CATEGORIES in MWh '000 | | | 1995 | 1996 | 1997 | 1998 | 1999 |
|---|---|---|---|---|---|---|---|
| | UNIT Years | 930 | 55 | 55 | 56 | 58 | 60 |
| | Catastrophic | >4500 | | | | | |
| | Semi-Catastrophic | 3600<x<4500 | | | | | |
| | Major Events | 2700<x<3600 | | | | | |
| | Main Events | 1800<x<2700 | 2,183,843 | | 2,607,502 | | |
| | | | | | | | |
| | | | | | | | |
| | Critical Events | 900<x<1800 | 1,591,692 | 1,296,035 | | | |
| | | | 1,405,323 | | | | |
| | Sub-Critical Events | 400<x<900 | | | | | 500,652 |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

**1995 – 1999**

| CATEGORIES in MWh '000 | | | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|---|
| | UNIT Years | 930 | 61 | 63 | 64 | 64 | 64 | 64 |
| | Catastrophic | >4500 | | | | 4,939,786 | | |
| | Semi-Catastrophic | 3600<x<4500 | | | | | | |
| | Major Events | 2700<x<3600 | | | 3,030,951 | | | |
| | Main Events | 1800<x<2700 | | | | | | 1,883,072 |
| | | | | | | | | |
| | | | | | | | | |
| | Critical Events | 900<x<1800 | | | | 1,592,515 | 969,666 | 1,560,482 |
| | | | | | | 1,090,182 | 941,711 | 1,133,479 |
| | Sub-Critical Events | 400<x<900 | | | | | 739,437 | 758,835 |
| | | | | | | | 548,517 | 462,239 |
| | | | | | | | 482,518 | 439,818 |
| | | | | | | | 466,747 | 431,491 |
| | | | | | | | 415,571 | |

**2000 – 2005**

**Table 2.4: Categorised Large MWh Losses 1990 - 2005**

## 2.3 The Number of GU’s failure and Loss Size Exposures

In the Introduction, the statement of the problem was presented and the GU’s losses (or Volume losses) were highlighted as the prominent issue. From this part of the Chapter onward, the topic of these exposures is analysed statistically for the Generation Division of EHL, in two parts: the number of GU’s failing and the size of the failure. The data used for the risk exposure spanned for every GU from 1990 to 2005. Prior to 1990, the forced outages (UCLF) for the Generation System averaged approx. 11%, oscillating between 8% and 14%. However, one ought to consider that in 1994 a management strategy (called the “90:7:3”) was put into effect to purposefully reduce the forced outages (UCLF) to a value of 3%, with realisation as soon as possible, but definitive before the turn of the millennium (i.e. by 1999). This target was considered accomplished in 1996 (see Fig. 2.5 below).
Therefore, given that this strategy is into effect, and observing its behaviour, the data from 1996 may be considered as stationary for other type of analyses (such as in Performance Management). This is particularly important for the analysis in paragraph 2.2.1, while for paragraph 2.2.2 the whole data set is used.

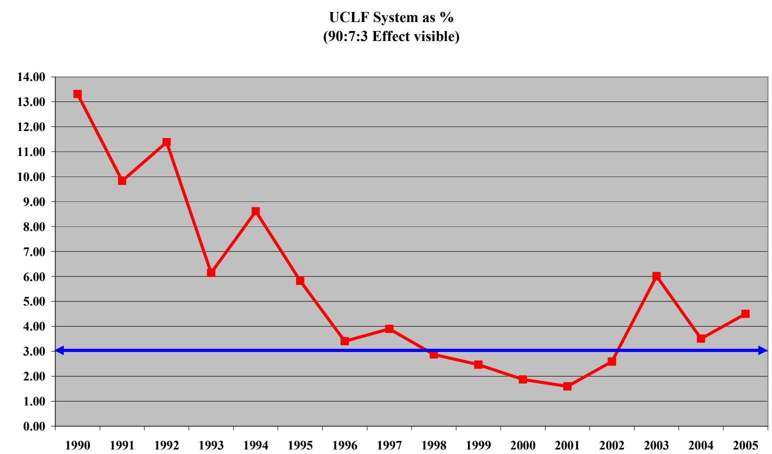



**Fig 2.5: UCLF actuals 1990 to 2005**

## 2.3.1 Number of Generating Units failing

Following up from Fig. 1.1 the GU failure (X) can be characterised from a Binomial distribution. Assume n identical GU of equal installed capacity (this kind of assumption can be valid for a Power Station: say a six-pack like Duvha PS, however Majuba has different IC's for its GU's – see Table 2.2, hence caution); each of these units has an expected forced outage rate (UCLF) $q$, then by characterising these GU with a Binomial distribution $B(n;q)$, one would get the probability, $P(X_i)$, of $i$ GU's failing being:

$$P(X_i) = C_i^n q^i (1-q)^{n-i}, \qquad \text{Eq 2.1}$$

where $C$ represents the Combination function of $i$ elements into $n$ within the $B(n;q)$ – see paragraph 2.2 (and equation C2.2 in Appendix C).

As seen in the equation above, a clear understanding of the forced outages on the system is necessary.
Another effect comes into play as well: due to the customers demand (or/and economic growth) the augmentation of installed capacity necessitated the number of GU's to increase (see Table 2.4 the Unit Years row, the No. of GU's per year increases from 1990 to 2005). This has an interesting effect on the modelling of the GU failures as a $B(n;q)$

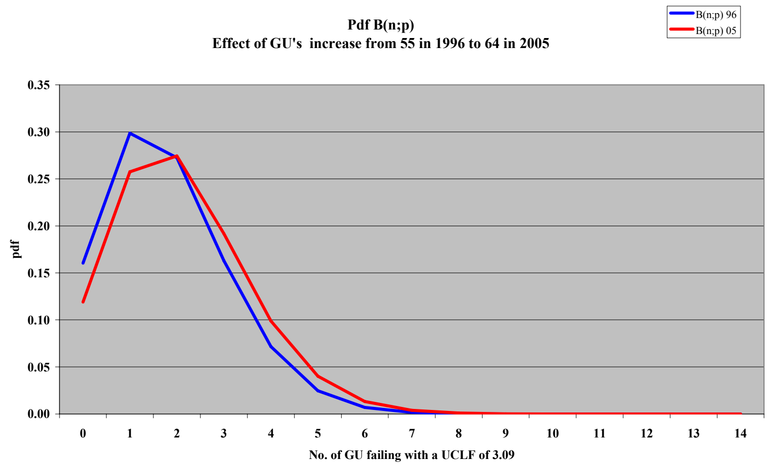



**Fig 2.6: Effect of GU's increase onto the *B*(*n;q*)**

Fig 2.6 indicates the effect of the increase of 9 GU's on the system (under the assumption of equal GU's – here we could be more accurate and convolute the Binomial for each installed capacity category and weight it accordingly to arrive at the Total System value). One can see that for 2 GU's the probability of failure is about the same but, for less than 2 GU's, the probability of failure is less for 64 GU's on the System than 55 GU's. However, the probability of failure of 3 units or more has increased from 1996 with the installation of more capacity.

Therefore to maintain the same characteristics for the system (with respect to the probabilities of failure), the system would have to improve the forced outage rate (UCLF). This is shown in Fig 2.7 below where an improvement of 0.49% in UCLF, rebalanced the probabilities to the 55 GU basis (shift from the "red" pdf to the "orange" pdf).

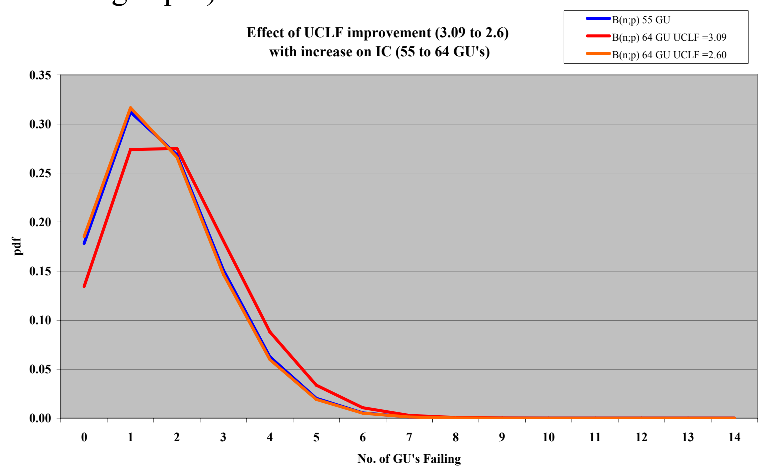



**Fig 2.7: Effect of UCLF improvement onto the *B*(*n;q*)**

| | | | |
|---|---|---|---|
| UCLF | 3.09% | 3.09% | 2.60% |
| GU's | 55 | 64 | 64 |
| P(x>2) | **0.24** | **0.32** | **0.24** |

**Table 2.5: Probability of more than 2 GU's failing**

As shown in Table 2.5, the system would be expected to operate at an improved UCLF (say 2.6%), given the increased capacity (to 64 GU's), so as not to worsen the probability of failure of more than 2 GU's (0.24).

However GU's tend to obey the laws of Reliability Theory, and as such when new they have high failure rates (also called "infant mortality" period or early period) then the GU tend to follow a period of stationarity (also called occasional period) and after that the failure rate tends to increase as the plant is getting old (called the degenerative period). This behaviour is often referred to as the "bath-tub" curve of the hazard rate.

Hence, bringing in new GU's not only increases the probability of GU's failing but it also increases the failure rate (due to the teething problems of the bath-tub), compounding the effect as it would be difficult to bring down the UCLF with new GU's coming in.

By the end of 1995, most of major GU's were now operational, hence one might ask whether the drop in UCLF was due to the 90:7;3 initiative, was it due to usual Reliability Theory behaviour, a bit of both and maybe some luck? From a personal perspective, it was both, and, having been involved in the 90:7:3 initiatives, the 90:7:3 with considerably more weight than the Reliability Theory one.

The Binomial process above can now be expanded further. Let us illustrate this expansion by the compilation of Table 2.6 below. The 930 value reflects the total number of Unit Years (i.e. the sum of however many GU's were in the fleet every year) over the 16 years period.

| | UNIT Years | 930 | No. of GU Cumulative | Rate in Category | E[x] |
|---|---|---|---|---|---|
| CATEGORIES in MWh '000 | Catastrophic | >4500 | 1 | 0.0010753 | 0.0688172 |
| | Semi-Catastrophic | 3600<x<4500 | 3 | 0.0032258 | 0.2064516 |
| | Major Events | 2700<x<3600 | 4 | 0.0043011 | 0.2752688 |
| | Main Events | x<2700 | 11 | 0.0118280 | 0.7569892 |
| | Critical Events | 900<x<1800 | 23 | 0.0247312 | 1.5827957 |
| | Sub-Critical Events | 400<x<900 | 39 | 0.0419355 | 2.6838710 |

Unitary rate (i.e. for any population N in a particular year). This means that for a year with a Tot fleet of 64 units, we would multiply this rate by 64 to get the E[x]

Expected No. of GU's failing in these Categories (or higher Categories) in any year.

**Table 2.6: Table of Expectations in Categories**

The next column gives No. of GU's for each Category accumulated from the previous one (e.g. in Semi-Catastrophic there are 2 Events + 1 Event in the Catastrophic Category; in the Major Events there is 1 Event + the previous 3 accumulated ones, etc…). The "Rate" can be interpreted as the "$q$" in Eq 2.1.

Since the Poisson model [$Psn(\lambda)$] (see Appendix C paragraph 2.3) is derived as the limit of the Binomial as $n \rightarrow \infty$ and $q \rightarrow 0$ and $nq = \lambda$ being constant, the $E[x] = n \,.\, q$ (see comments on the Rate in Table 2.6 above), this may be used to compute the expectations in each category (refer to Appendix C paragraph 2.3 for the calculations in the table).

The expectation in the last column of Table 2.6 above can be interpreted as follows: on average, one would expect, in a year, to have 2.7 GU's failing in the Sub-Critical category, 1.6 in the Critical, 0.8 GU in the Main Events category, and skipping the other categories but going straight to the Catastrophic, one would expect such an event once every 14.5 years (= 1 / 0.0688).

Furthermore, the Poisson Distribution may now be used to compute the probabilities in each of the categories given in the top table of Table 2.7

| GU's in Category failing | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Catastrophic | 0.933497304 | 0.0642407 | 0.002210432 | 5.07052E-05 | 8.72348E-07 | 1.20065E-08 | 1.37709E-10 |
| Semi-Catastrophic | 0.813465622 | 0.1679413 | 0.017335875 | 0.001193006 | 6.15745E-05 | 2.54243E-06 | 8.74815E-08 |
| Major Events | 0.759367965 | 0.2090303 | 0.028769765 | 0.002639806 | 0.000181664 | 1.00013E-05 | 4.58841E-07 |
| Main Events | 0.469076577 | 0.3550859 | 0.134398113 | 0.033912642 | 0.006417876 | 0.000971653 | 0.000122588 |
| Critical Events | 0.205400 | 0.3251063 | 0.257288449 | 0.135745017 | 0.053714157 | 0.017003707 | 0.004485566 |
| Sub-Critical Events | 0.068298261 | 0.1833037 | 0.245981768 | 0.220061108 | 0.147653905 | 0.079256806 | 0.035452507 |

| Expected Events in Categories | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Catastrophic | 15 | 1 | 0 | 0 | 0 | 0 | 0 |
| Semi-Catastrophic | 13 | 3 | 0 | 0 | 0 | 0 | 0 |
| Major Events | 12 | 3 | 0 | 0 | 0 | 0 | 0 |
| Main Events | 8 | 6 | 2 | 1 | 0 | 0 | 0 |
| Critical Events | 3 | 5 | 4 | 2 | 1 | 0 | 0 |
| Sub-Critical Events | 1 | 3 | 4 | 4 | 2 | 1 | 1 |

| Observed Events in Categories | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Catastrophic | 15 | 1 | 0 | 0 | 0 | 0 | 0 |
| Semi-Catastrophic | 14 | 2 | 0 | 0 | 0 | 0 | 0 |
| Major Events | 15 | 1 | 0 | 0 | 0 | 0 | 0 |
| Main Events | 11 | 4 | 0 | 1 | 0 | 0 | 0 |
| Critical Events | 8 | 4 | 4 | 0 | 0 | 0 | 0 |
| Sub-Critical Events | 10 | 2 | 1 | 1 | 1 | 1 | 0 |

**Table 2.7: Table of Expectations (top) and Contingency Tables**

These may be used further to compute the expected frequencies in each class. This is done as follows. For example, in the 16 years, we have **observed** 15 years with 0 catastrophic events and 1 year with a catastrophic one. We would have **expected** 0.933497*16 ≈ 15 and 0.0642*16 ≈ 1 events according to the Poisson model. The other frequencies for the rest of the classes in the middle and bottom tables of Table 2.7 are computed in a similar fashion.

Validation of the Poisson model for the Categories:

More Contingency tables are compiled for the various categories and the $\chi^2$ test used to validate the Poisson model.

GROUPED

| Semi-catastrophic | 0 | 1 |
|---|---|---|
| EXPECTED | 13 | 3 |
| OBSERVED | 14 | 2 |

$\chi^2$: 0.250

Prob signif. $\chi^2$: 0.6170

df= 1

GROUPED

| Major Events | 0 | 1 |
|---|---|---|
| EXPECTED | 12 | 3 |
| OBSERVED | 15 | 1 |

$\chi^2$: 2.312

Prob signif. $\chi^2$: 0.1284

df= 1

GROUPED

| Main Events | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| EXPECTED | 8 | 6 | 2 | 1 |
| OBSERVED | 11 | 4 | 0 | 1 |

$\chi^2$: 4.661

Prob signif. $\chi^2$: 0.1984

df= 3

GROUPED

| Critical Events | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| EXPECTED | 3 | 5 | 4 | 2 | 1 |
| OBSERVED | 8 | 4 | 4 | 0 | 0 |

$\chi^2$: 10.073

Prob signif. $\chi^2$: 0.0392

df= 4

GROUPED

| Sub-Critical Events | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| EXPECTED | 1 | 3 | 4 | 4 | 2 |
| OBSERVED | 10 | 2 | 1 | 1 | 1 |

$\chi^2$: 77.680

Prob signif. $\chi^2$: 0.0000

df= 4

**Table 2.8: Contingency Tables of Grouped Categories**

As the results were a bit sparse, Grouping from Sub-Critical Events to Semi-Catastrophic was performed. The “bins” were reduced [e.g. from 7 to 5 (i.e. spanned from 0 to 4 Events)]. Catastrophic had almost a perfect fit, so the table was excluded as redundant.

From Table 2.8, it is significant to note the following:

- The top table can be considered as a 2 x 2 Contingency table
- That the $\chi^2$ test indicated that the model fitted well in the categories from Main Events to Catastrophic
- It did not fit well for the other 2 categories, probably due to the interval limits for the two categories Critical and Sub-Critical events.

## 2.3.2 Size of Volume failure

In terms of the energy loss (power x time) in MWh, in theory it would be quite possible to determine that the total maximum loss is 318 TWh in a year (given that it is a leap year); however this kind of loss would be cataclysmic for South Africa which would be similar to the probability of a global nuclear warfare and the annihilation of a country. For all practical purposes, the asymptotic properties of the distributions are assumed in this thesis. From the other end, there were 2 very small losses (0.1 MWh) that were excluded from the analyses in Chapter 3, reducing the total GU-Years from 930 to 928.

Plotting the size of the losses that every GU incurred in any year is shown in Fig 2.8 below, indicating that over a period from 1990 to 2005 the max size of the loss was 4939.79 GWh, 1.6% of the total maximum potential loss. However, this kind of loss is considered catastrophic as it represents one large GU out of action for one whole year, resulting in losses of revenue approximating R 550 million (in 2003 Rands). If one adds up the repair to damage that would put a GU out for such a time, the costs would be over a billion Rand.

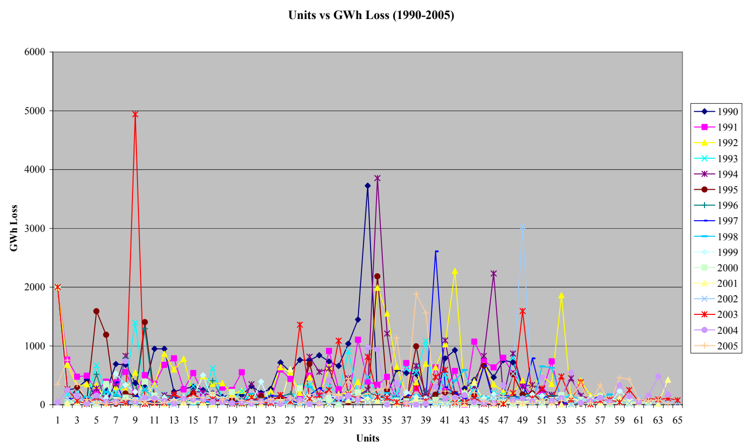



**Fig 2.8: GWh Losses incurred by GU's in particular years**

Let us assume that the MWh losses can be expressed as an essentially positive random variable (energy losses ≡ $X$) expressed by a particular distribution function having a pdf $f_X(x)$ and a CDF $F_X(x)$. Here it is also assumed that $F$ is continuously differentiable and that the following applies:

$$F_X(x) = \int_0^x f_X(x).dx$$

(see A1.0 and related issues in Appendix A)

Since empirical quantiles open the way to tail estimation, let $\hat{F}_X(x)$, the empirical CDF be derived by ordering $X_i$'s in such form: $X_1^* \le X_2^* \le \ldots \le X_n^*$ and associating each $X_i^*$ with its percentile $p = i/(n+1),\ i = 1,2,\ldots,n$ .

This $\hat{F}_X(x)$ is depicted below, in Fig 2.9. Here, the very long tail of the distribution is evident.

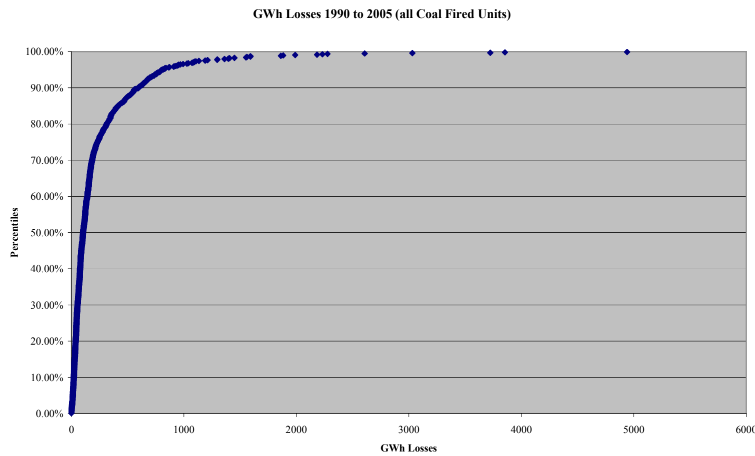



**Fig 2.9: CDF of GWh Losses**

Having looked at the individual GU's energy losses, and the maximum loss that an individual GU could incur in a year, one could also ask **what is the maximum sum of all the GU losses in a year**. Another relevant question could be **how the values of the maximum GU losses affect the value of the maximum sum of the GU losses**, over a threshold ***t*** ; in other words how do these individual losses affect the System. Referring to [13] Embrechts it is shown there that:

$$\lim_{t \longrightarrow \infty} \frac{P(X_1 + \ldots + X_m > t)}{P(\max(X_1, \ldots, X_m) > t)} = 1$$

, for every $\boldsymbol{m} \geq 2$ Eq 2.2

Hence the tails of the distributions, either of the sum or the individual GU's losses of the first ***m*** GU's losses that would exceed the threshold ***t***, are asymptotically of the same order. **<u>This indicates the strong influence that the maximum individual GU energy loss has on the system</u>**.

One could model further this aspect of total losses using the classical risk model:

$$U(\tau) = u + \alpha.\tau - S(\tau)$$

Where $u$ is the initial capacity, $\alpha$ is the available capacity @ time $\tau$ and $S(\tau)$ is the sum of the GU losses (i.e the Generation System Total Losses) until time $\tau$.

Then, $S(\tau) = \sum_{i=1}^{N(\tau)} X_i$ , where $N(\tau)$ is the number of GU's that have failed until time $\tau$ and is a homogeneous Poisson process independent of $X_i$ , the GWh Losses.

Taking this modelling aspect one further step ahead onto the Poisson process (see Fig 1.1). The Poisson approximation provides the basic essentials for the analysis of infrequent extremes from a sample. It is also the key to the Weak Limit Theory of upper order statistics as well as for the Weak Convergence of Point Processes, [13] Embrechts.

Usage of the Cramér-Lundberg model:

a] The GU's losses size process:

The sizes $Xi$ are positive iid rv's having common non-lattice $F$, finite mean

$\mu = E[X_i]$ and variance $\sigma^2 = var[X_i] < \infty$

b] Times at which the GU losses occur are random instances of time:

$0 < T_1 < T_2 < \ldots$ a.s. (almost sure [i.e. with probability of almost 1])

c] The arrival process

The number of GU's in the interval $[0,\tau]$ is denoted by

$N(\tau) = sup\{n \geq 1 \colon T_n \leq \tau\}$

d] The inter-arrival times of losses
$Y_1 = T_1,\ Y_k = T_k - T_{k-1}\ ,\ k = 2,3,\ldots$ are iid exponentially distributed with
finite mean $E[Y_1] = 1/\lambda$ .

A consequence of the definition above is that $N(\tau)$ in a homogeneous Poisson process with intensity $\lambda > 0$ and $N(\tau) > 0$ hence:

$$P(N(\tau) = k) = e^{-\lambda\tau} (\lambda\tau)^k \big/ k! \quad , \text{for } k = 0,1,2,\ldots$$

Here the loss size $X$ and inter-arrival time $Y$ of losses are independent of each other [13] Embrechts. See Appendix C (equation C2.4) for an extension on this aspect.

When looking at $S(\tau)$ from a distributional aspect and express it as:

$G_\tau(s) = P[S(\tau) \leq s]$, [3] Beirlant (p117) showed that, using the independence assumption between $\{N(\tau)\ ;\ \tau \geq 0\}$ and $\{X_i\ ;\ i \geq 1\}$ and a conditioning argument on $N(\tau)$:

$G_\tau(s) = \sum_{n=0}^{\infty} p_n(\tau).F^{*n}(\tau)$, where $*n$, refers to the distribution of the total loss,

$S_n(\tau) = X_1+X_2+\ldots+X_n$, the $n$-th convolution of $F$ with itself. Considerable research is being done in this compound distribution function, [3] Beirlant (p117) and pp 13, 37-49, 571-576 [13] Embrechts.

In this field Super- and Sub-exponential distributions play an important role, for instance if the largest of the GU losses (*max* [$Xi$]) influences the stochastic behaviour of $S_n(t)$, then the Sub-exponential distributional properties ought to be used.

The number of GU's incurring a loss, has a distribution considered "light" if:

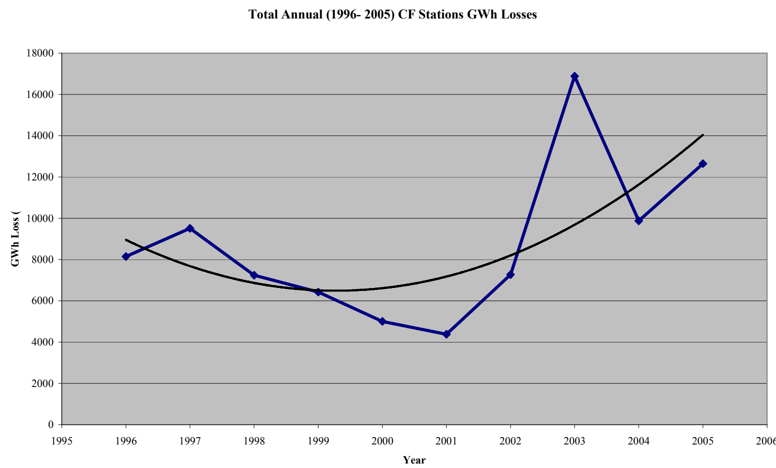
$$E[z^{N(\tau)}] = \sum_{n=0} p_n(\tau).z^n \quad < \infty \quad , \text{for some } z > 1$$

Hence a light "GUs-Losses-number" distribution has exponentially small probabilities. If we believe or can verify this statement, then according to [3] Beirlant, Cline's (1987) and (for more precise estimates) Willekens-Teugels' (1992) results can be used for the computation of quantiles of $S(\tau)$. Caution ought to be exercised since the inversion of Cline's model yields conservative results on the computation of high quantiles. Much more research is needed in this field and therefore the results are limited to the EDA for this topic.

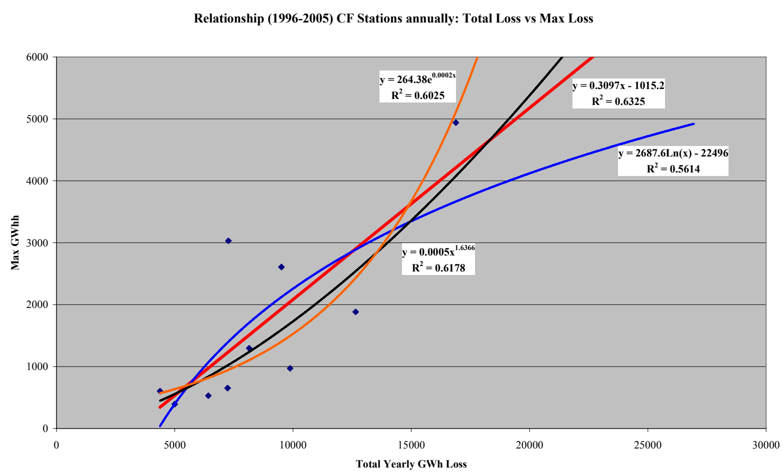
| Year | Max Loss | Total Loss |
|---|---|---|
| 1996 | 1296.0 | 8147 |
| 1997 | 2607.5 | 9513 |
| 1998 | 651.8 | 7238 |
| 1999 | 529.9 | 6421 |
| 2000 | 394.0 | 5002 |
| 2001 | 602.7 | 4379 |
| 2002 | 3031.0 | 7270 |
| 2003 | 4939.8 | 16885 |
| 2004 | 969.7 | 9876 |
| 2005 | 1883.1 | 12647 |

**Table 2.9: Maximum GU loss and Total System loss**

The subset 1996 to 2005 was used as this window is considered a more stationary period.




# Fig 2.10: Annual Total System GWh Losses




# Fig 2.11: Relationship: System GWh Losses vs GU Max Loss

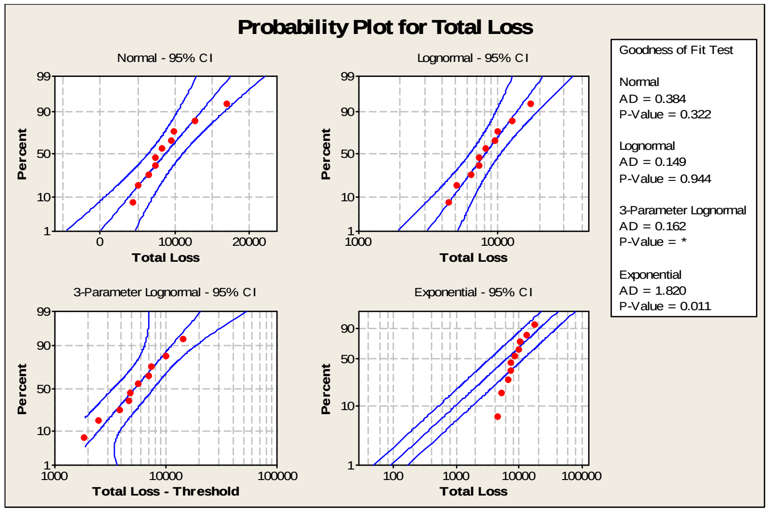



**Fig 2.12: Distributional Fits of System GWh Losses**

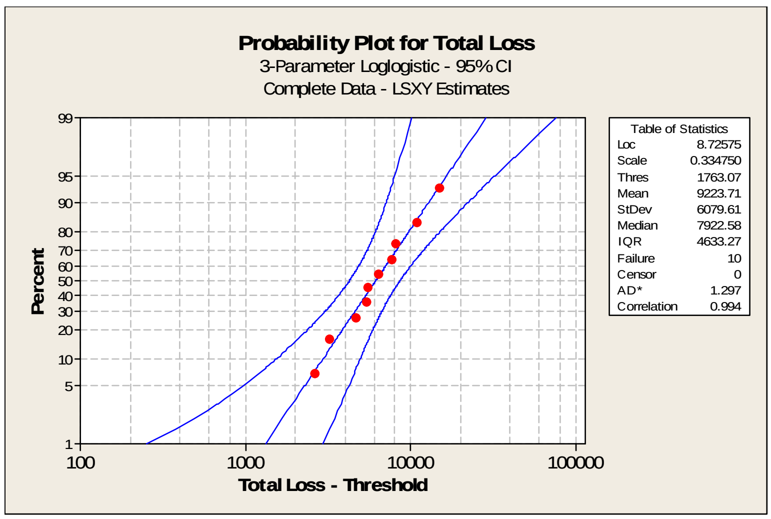



**Fig 2.13: LogLogistic Fit to System GWh Losses**

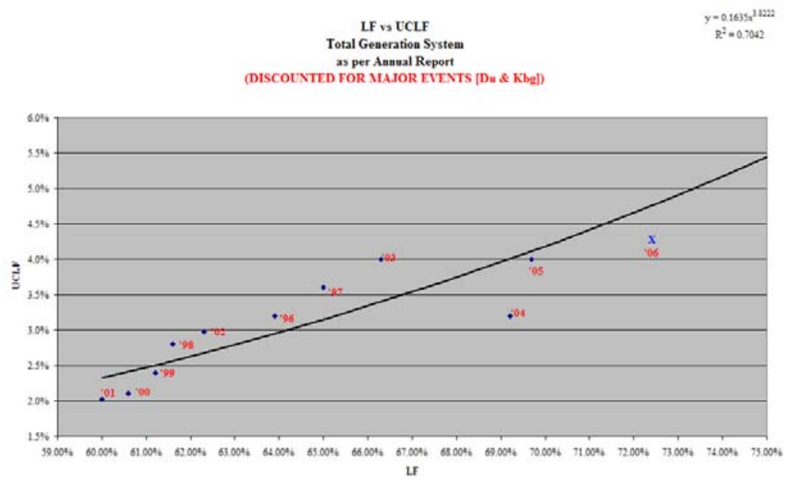



**Fig 2.14: Relationship: System GWh Losses UCLF and Load Factor**

The Total System Losses in GWh (TSL) for the period 1996 to 2005 are depicted in Fig 2.10. The TSL shows an increasing upward trend, or possibly a cyclic behaviour as time increases. Fig 2.11 shows the relationship between the GUs' max loss in a particular year associated with its corresponding TSL. Even though the Logarithmic model $y = b.Ln(x)+a$ has the lowest $R^2$ (blue line), from an engineering perspective it makes more sense for the following reason: the Max GWh loss (y-axis) is limited to the IC x Hours p.a. (say 5270.4 GWh for a 600 MW GU in a leap year), hence one would expect the relationship to be asymptotic towards that value; the limit of the TSL being approximately 318 000 GWh (x-axis, 10 time more to the right than the value illustrated).

If not trended, and distributional fits are performed on the TSL, the results are shown in Fig 2.12 and 2.13, with the

LogLogistic distributional fit being the "best". This indicated a "threshold" (shift parameter) of 1763.07 GWh. Although this analysis bodes well, it is somewhat limited as it is not leveraging the nature of the trend, cyclical or curvy-linear relationship, seemingly with time. Further work by the author indicates that another concomitant variable plays an important role, i.e. the Load Factor (LF). The LF is time related and a function of the energy demand imposed by the market on the IC. The more demand is imposed on the GUs' the higher the stress on the GUs' the higher the losses at System level (the model shown in Fig 2.14 is only valid at System level, though, and not at GU level: as the LF increases, at GU level, the GU fails and hence the LF at that GU is 0, however another GU "picks-up" the load, hence the LF at System level is seen as increasing with the corresponding loss at System level increasing). This relationship is illustrated in Fig 2.14 (the 2006 value is shown by **X**). Further studies were performed by the author on this issue to validate the relationship, with positive results. This also involved experimental work on the Cobb-Douglas production model: $e^{kY}.Y = A.X^{b}.Z^{c}.\varepsilon'$ or in linear form:

$kY + Ln(\mathrm{Y}) = a + b.Ln(X) + c.Ln(Z) + \varepsilon$,

where $Y$= UCLF, $X$= LF and $Z$ the lagged (5year) PCLF or, most likely, the cost of production; a simplified version is shown in Fig 2.14.

It would also be appropriate to utilise the property of the losses conditional to the load when estimating the TSL and its expected extremes. To do that, appropriate forecasting techniques ought to be used to forecast the energy demand with its appropriate credible set (or confidence region; the Bayesian "confidence interval"), say the High Posterior

Density (hpd) credible set. This could then be used to infer the TSL losses on the logarithmic model proposed above. A contour of TSL given a region of LF can then be constructed.
Delving into EVT, a recent method developed by [28] Nel et al using the Logistic Copula can be used to estimate the joint probability of the annual GU maximum loss and the annual TSL by considering the joint distribution of the of the GU and TSL marginals through the Logistic Copula. Should the annual values be too small of a sample to consider, the quarterly max GU and TSL values could be utilised; this might be of interest to determine the winter and summer risk exposure to maximum losses and the provision of reserve capacity in those periods.

To return to the focal point of this thesis, i.e. the GUs' extreme losses incurred in a year, the above topic as well as this Chapter is ended here but certainly gives a background for further research.

**CHAPTER 3**

# Analyses of Extremes in GWh Losses

## 3.1 Introduction

The risk associated with extreme events has and is intriguing management more and more. In the energy platform this aspect is no different and the implementation of risk models that allow for rare but damaging events is starting to become more intrinsic in the running of the electricity business under the umbrella of Risk Management.
It is envisaged that an energy Vol@R in the form of GWh and provision for reserve capacity, would become part of NERSA legislature for utilities in South Africa.
The approach in this thesis is in attempting to model these types of risks in such a way that the GUs' extreme losses are addressed. Eventually, this is foreseen to provide a measure for risk as mentioned above (Vol@R).

Although the research indicated that [16] Fuller in 1914 and [18] Griffith in 1920 played a role in describing extreme events, it is still the author's belief that the seed of EVT results from [35] Tippett.

Fig 1.2 portrayed a visual of the occurrence of extremes in a distribution. In paragraph 1.2 the basics and an outline of

EVT have been mentioned in terms of the history, Maximum Domains of Attraction (MDA), various methodologies to estimate the parameters of the distributions in the extreme values domain. It also gives a short introduction to using the method of "Entropy" to, say, assess an optimal level of the threshold given the distribution parameters estimates. The entropy concept is also used in EVT in stochastic modelling so that modelling is not done just heuristically and informally, but a more general approach is taken so that a simplistic, explicit and reproducible manner is used, [43] Zellner

The extremes of the GUs' losses in GWh for the period 1990 to 2005 in the form of the empirical CDF, are illustrated in Fig 2.9. From there the "long" tail of the losses behaviour is clearly visible. What is also visible is that, at the tail, the set consists of very few data points relative to the rest of the set.

## 3.2 Block Maxima

Even though in this part of the modelling, two approaches are used, the Block Maxima (BM) and the Peak-Over-Threshold (POT), the BM was explored but the POT approach was eventually adopted.
The idea was to model the extreme yearly GU GWh loss. The methodology consisted in taking the maximum GWh loss for each year over all the GUs', $Y_i = Max\ \{X_j\}$, for $i$=1,2,…,$n$-th year and $j$=1,2,…,$m$-th GU. Hence for year 2005, the max loss over 64 GUs' was $Y_{16}$ = 1883 GWh loss for the 16$^{th}$ year since 1990, Fig 3.1 below shows the data points for $Y_i$.

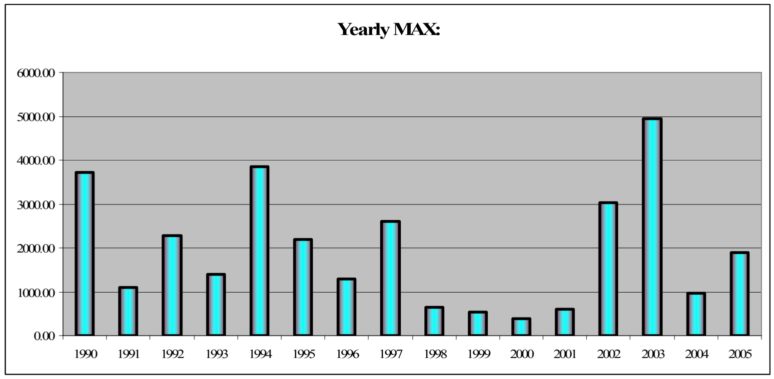



**Fig 3.1: Yearly Maximum GU GWh**

From EVT, the distribution of $Y_i = Max\ \{X_j\}$, as $m \to \infty$ converges to a GEV or more specifically to a *SGEV*:

$$G_\gamma(x) = \begin{cases} exp\ \{ -(1 + \gamma.[(x-\mu)/\sigma]^{-1/\gamma} \} & \text{, if } \gamma \neq 0 \\ exp\ \{ -exp\{-(x-\mu)/\sigma\} \} & \text{, if } \gamma = 0 \end{cases} \quad \gamma \in \Re \quad \text{Eq 3.1}$$

where $\gamma$ is the ***shape*** parameter, $\mu$ the ***location*** parameter and $\sigma$ the ***scale*** parameter.

By using the PWM method the parameters were estimated as:

| | |
|---|---|
| **Shape** | **0** |
| **Scale** | **995.04** |
| **Location** | **1358.55** |

With these estimates, Eq 3.1 reduces to:

$G_\gamma(x) = exp\{ -exp\{-(x-\mu)/\sigma\} \}$ which is of the Gumbel type (see Appendix Eq. A2.5)

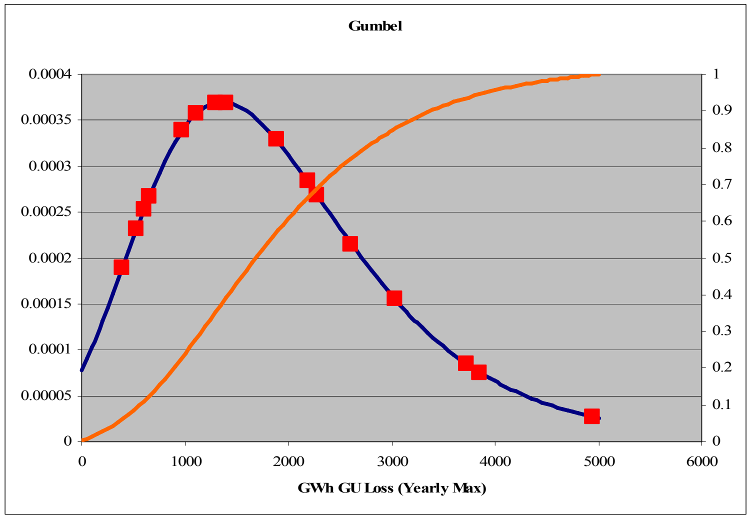



**Fig 3.2: Yearly Max Distribution: Gumbel**

The Gumbel pdf is given by:

$g_\gamma(x) = (1/\sigma).\ exp\ \{-(x-\mu)/\sigma\}.\ exp\ \{-exp\ \{-(x-\mu)/\sigma\}\}$ Eq 3.2

for $x \in \Re$ (see Eq. A2.8)

Using the estimates above, the pdf illustrated in Fig 3.2 is obtained. The actual yearly maxima are shown by the red blocks when entered into the pdf equation above.

The interpretation of Fig 3.2 is quite useful.

Over a similar window period the expectation is to see values of 1359 GWh losses more often than other values as a yearly maximum. Half of the time, over a similar window period, the system is expected to incur yearly maxima greater than 1700 GWh.

What is also of interest is the return period.

One could be interested what value could be exceeded in any particular year with a chance of one in twenty (i.e. $p$=1/20 = 0.05). By using:

$Q_{Y;p} = \mu - \sigma Ln\,(-Ln(1\text{-}p))$ Eq 3.3

(see Appendix A, paragraph 2.3.1), and the estimates above, we get $Q_{Y;0.05}$ = 4314 GWh.

Let us assume that the management needs to know what is the yearly maximum GWh loss that we expect to exceed every 10 years, i.e. $P(Y > {}_{10}y) = 0.1$

$\therefore$ from Eq 3.3 above, ${}_{10}y$ = 3598 GWh; i.e. the company should expect the maximal annual GWh loss to exceed 3598 GWh every 10 years.

| p0 | p | Qy;p | |
|---|---|---|---|
| 0.85 | 0.150 | 3166.5 | |
| 0.875 | 0.125 | 3362.0 | |
| 0.9 | 0.100 | 3597.8 | |
| 0.925 | 0.075 | 3897.4 | |
| 0.95 | 0.050 | 4314.0 | |
| 0.975 | 0.025 | 5016.6 | |
| 0.980 | 0.020 | 5241.1 | |
| 0.985 | 0.015 | 5529.9 | |
| 0.990 | 0.010 | 5935.9 | |
| 0.9375 | **0.0625** | 4085.45 | **1/16** |
| 0.97302 | 0.02698 | **4939.789** | **1/37** |

**Table 3.1: Quantiles, Block-Maxima method**

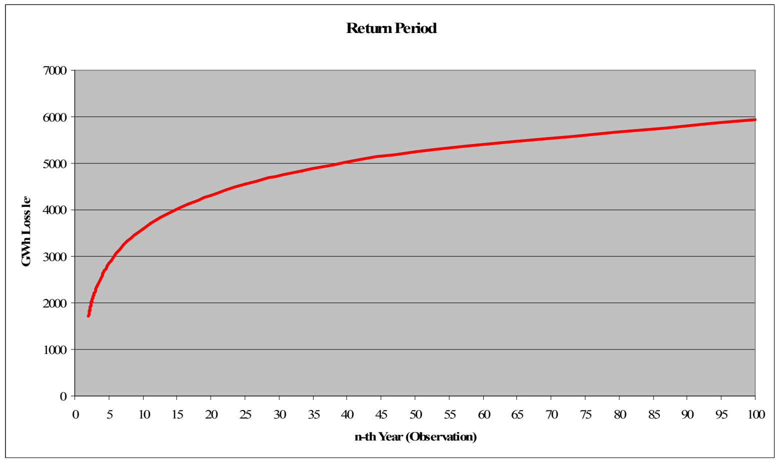



**Fig 3.3: Return period of GWh Losses – Yearly Maxima**

The GWh loss levels can be expressed in the form of Fig 3.3 above. This may be interpreted as follows: by entering the expected year one reads off the value on the y-axis (the values are computed by means of Eq. 3.3). For instance, the management can expect every 20 years to exceed the value of 4314 GWh loss. These quantiles refer to the Stress Loss, for instance, $p_0$ = 0.95 yields the 20 year return level of a GU GWh loss It may also be considered as a kind of unconditional quantile estimate for the unknown underlying distribution of F (which is in the MDA of the Gumbel, see Appendix A – Classical EVT).
However, the quantiles in Table 3.1 are based on limited information and inferences are made upon the parental distribution based on 16 values out of 928. If one had to then be even more prescriptive and invoke stationarity, then even less observations (see Table 2.9 in Chapter 2, the data spans from 1996 instead of 1990)) can be used (i.e. that the block maxima are invariant under time shifts – this

was relaxed when doing the analysis). Comparing Table 3.1 results from the fitted pdf, the expectation of the quantile for 1/16 years is 4085.5 GWh, however within this period, a value of 4939.8 GWh was observed; by computing the inverse, for a value of 4939.8 GWh we get a return period of 0.02698 or 1/37 years, over than twice the span!

Clearly, this method seems not to be yielding the expected results. This might be that for the next 21 years we won't observe such a high value, although unlikely; it might also be that there is not enough history. Or this might well be that, even with a limited dataset, by making use of the relationship of the losses within the categories, one could obtain better results.

For instance, assume that $X$ is the random vector of the GUs' losses falling in different categories, such as those illustrated in Table 2.4 in Chapter 2.

Then, $\mathbf{X} = (X_1, X_2,\dots,X_d)$ for each of the $d$ different categories, assume that the $X$'s have a joint distribution $F_X(x_1, x_2,\dots,x_d) = P(X_1 \le x_1, X_2 \le x_2,\dots, X_d \le x_d)$ and assume that the individual losses have continuous marginals $F_i(x) = P\{X_i \le x\}$, then by [25] McNeil p111,

$$F_X(x_1, x_2,\dots,x_d) = C\,[F_1(x_1), F_2(x_2),\dots, F_d(x_d)] \qquad \text{Eq 3.4}$$

where $C$ is the Copula of $F$ and measures the dependence structure of $\mathbf{X}$. For the GEV in the Block Maxima application these are referred to as MEV (Multivariate Extreme Value) Copulas.

In the BM methodology used, the Gumbel Copula would be given by Eq 3.5 below:

$$C_{\beta}^{Gu}(v_1, v_2) = e^{[-[\{-Ln(v_1)^{1/\beta}\}+\{-Ln(v_2)^{1/\beta}\}]^{\beta}]}\ ,\ 0 < \beta < 1 \qquad \text{Eq 3.5}$$

The Copula would show the tail dependence, i.e. the tendency of extreme values to occur together in the different categories.
Let $d$=2 (i.e. only two categories of losses):
If $\beta$ =1, $C_{\beta}^{Gu}(v_1, v_2) = v_1.v_2$, which models the independence of $(X_1, X_2)'$
For $\beta$ = ]0;1[ , then the Copula models the dependence between $X_1$ and $X_2$ .
At $\beta$ = 0, the process would have perfect dependence between $X_1$ and $X_2$ i.e. $X_1 = g(X_2)$.
In Chapter 2 Table 2.4 exhibited 6 categories. As it may be seen, the modelling of 6 dimensions in amalgamating Eq 3.5 into Eq 3.4 becomes very complex. This is a typical problem when d becomes large in the usage of Copulas in the BM methodology. This problem is usually referred to as the "curse of dimensionality"; a usual solution is to collapse the model from a Multivariate into a Univariate one.

NB: In the parameter estimation for the BM pdf above, the PWM method was used. The reason for this is that the value for $\gamma$ was not known. Seen that the GEV applies to the BM method, if one were to choose MLE to estimate the parameters, one ought to exercise caution as a condition for using MLE is that the data must be free of the parameters we are trying to estimate, this is usually referred to as regularity. The GEV is known to be "non-regular" for $\gamma \neq 0$.

Although the BM method provided an insight in the extremes and surely more research for the future in using

the Copulas, unfortunately, with all the data that was collected, it did not yield the needed information adequately (see Table 3.1 and comments in p.49 [top]).
This enticed further research into the Peak-Over-Threshold method as a different approach to analyse the GWh losses.

## 3.3 Peak Over Threshold (POT)

As described in the previous paragraph, the BM approach yielded possibilities, such as stress-loss assessment, usage of Copulas, but somehow lacked effective returns out of the analyses (see Table 3.1 above whereby that max loss should have predicted 1/16 and not 1/37). Somewhat lately, a more modern approach has been one of using large observations which exceed a relatively high threshold. This method is known as the Peak Over Threshold (POT) method. Because of the usually limited data sets, the POT method is generally considered, at present, the most effective applied approach to EVT. There are two further types of methods within POT: The semi-parametric (POTsp) and the full-parametric (POTfp). The POTsp concentrates around the Hill, Pickands, Zipf, Beirlant-Q estimators while the POTfp, is based on the GPD properties.
There is a further methodology that is gathering momentum and that is the one that applies itself to both the BM and POT methodologies: the Bayesian approach to EVT, Chapter 4 is devoted to this methodology.
In this Chapter, the analysis concentrates on the classical approach, and utilises the POTsp to narrow down the analysis; the POTfp is the avenue taken for the GUs GWh losses parametric estimation.

### 3.3.1 Semi-parametric method (POTsp)

As in the classics, it is said that a picture explains a thousand words; this approach follows that kind of philosophy. The Quantiles properties are leveraged to yield required results. For instance, Quantiles (of the sample and distribution) are linearly related if the drawn sample is derived from the hypothesised distribution. By utilising this property a power instrument called the Q-Q plot can be used to derive the goodness-of-fit of the GWh Losses (see Appendix C paragraph 2.5 on derivation of Quantiles from particular distributions).

Since we are dealing with extremes the first attempt is at plotting the classical ones that map onto the MDA: the Weibull, Gumbel and Fréchet distributions.

This was performed and the results are illustrated in Fig 3.4 to Fig 3.7 here below.

The Exponential Q-Q plot of Fig 3.6 was given as illustrative of a specific part of the Gumbel domain (note the kink between the12$^{th}$ and 11$^{th}$ largest observations).

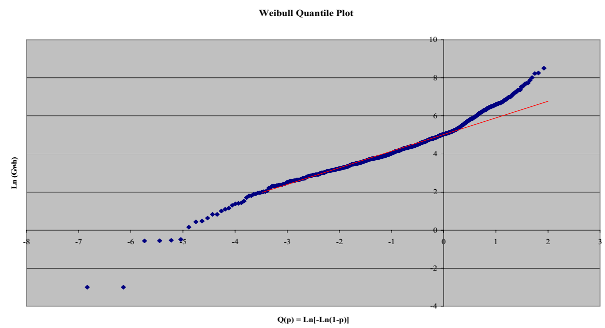



Fig 3.4: Weibull Q-Q Plot of the GWh Losses

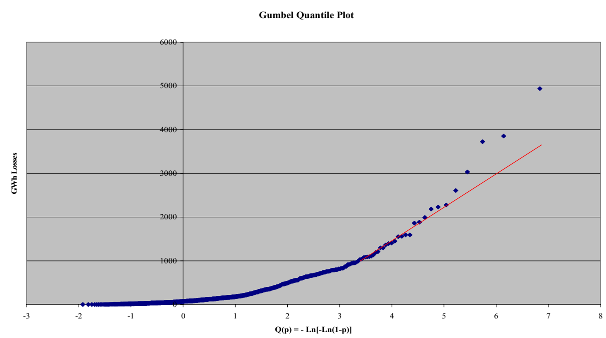



Fig 3.5: Gumbel Q-Q Plot of the GWh Losses

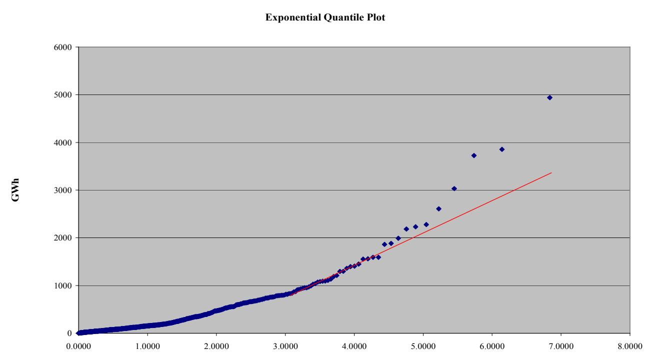



Fig 3.6: Exponential Q-Q Plot of the GWh Losses

Visually, Fig 3.4 to Fig 3.6 exhibit a significant departure at the tail (using the red straight line as a reference). Since the objective here is to narrow down to which population the GWh Losses belong to, the Gumbel and Weibull families already show that the tails from these families do not represent well the behaviour of the GWh losses. There remains one other class and that is the Fréchet class of distributions. When exploring this avenue, the Q-Q plot shown in Fig 3.7 was obtained. It is visually clear that this type of distribution seems more plausible than the others, as the linear tail trend fit well in the tail and well into the empirical on the left hand side.

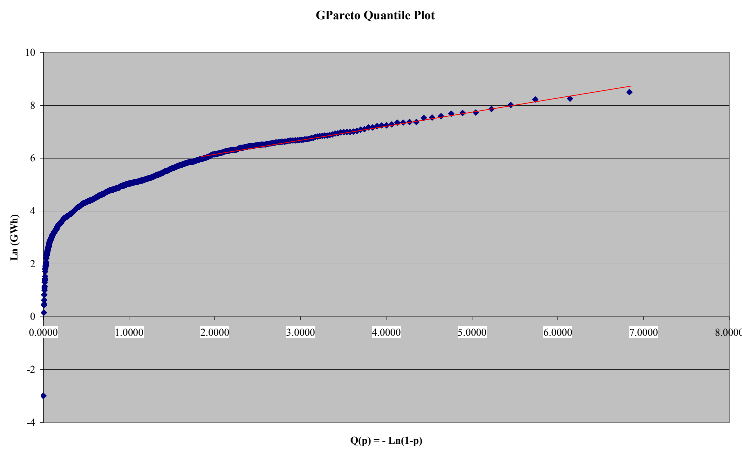
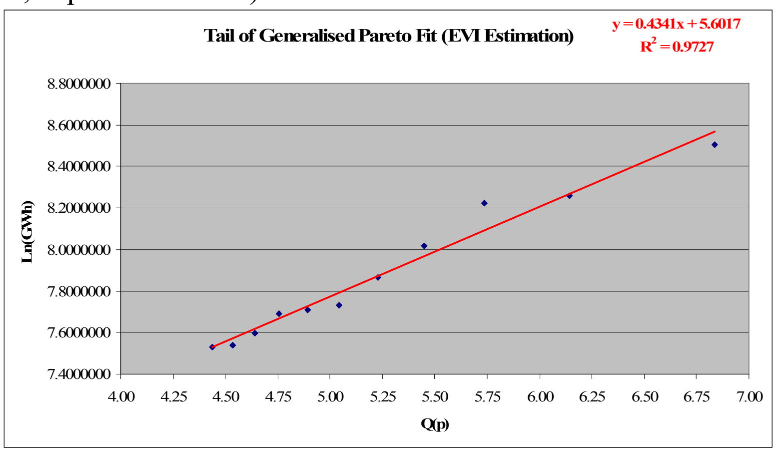



**Fig 3.7: Generalised Pareto Q-Q Plot of the GWh Losses**

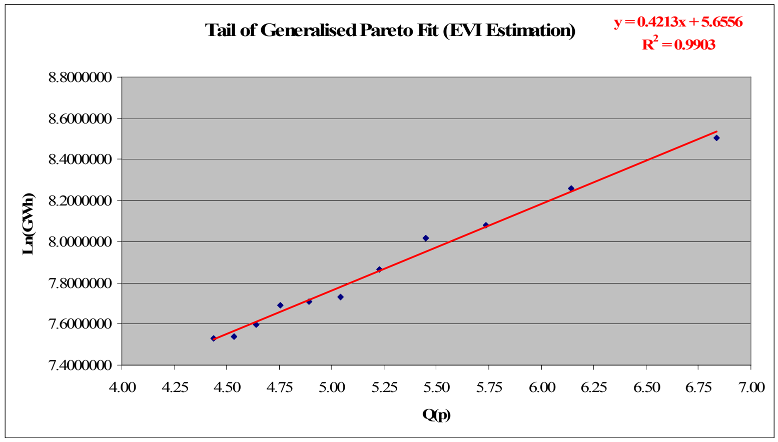
The Q-Q plot of the GPD also indicates that there seems to be a kink between the 12th and 11th largest observations. Zooming in at this aspect of the tail and utilising the 11 largest observations in the Quantile domain, and applying a linear model fit to obtain the parameters, Fig 3.8 was obtained. The EVI (see Appendix A paragraph 2.1) can be estimated from the slope of this fit namely 0.4341 with a 95% CI [0.3793 ; 0.4889].

This seems to give a reasonable result and prompted further investigation of this tail: the fit in Fig 3.8, indicated a "blip" in the 3rd largest observation. This loss was incurred at a GU3 of a particular power station in '90, if a "background loss" of approx. 500 GWh was removed, then a much better fit was obtained which is illustrated in Fig 3.9 yielding to an EVI estimate of 0.4213 with a 95% CI [0.3898 ; 0.4528]. With an EVI > 0 , the process indicates that it not does not derive from a Gumbel Class (Appendix A, Eq A2.3 to A2.5).




| *k =11* | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| **Intercept** | 5.6017 | 0.1282 | 43.698782 | 0.000000 | 5.3117 | 5.8916 |
| **X -Slope** | **0.4341** | 0.0242 | 17.924033 | 0.000000 | **0.3793** | **0.4889** |

**Fig 3.8: GPD Q-Q Plot of the 11th Largest Observations**




| *k =11 (Adj.)* | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| **Intercept** | 5.6556 | 0.0737 | 76.736524 | 0.000000 | 5.4889 | 5.8223 |
| **X -Slope** | **0.4213** | 0.0139 | 30.256506 | 0.000000 | **0.3898** | **0.4528** |

**Fig 3.9: GPD Q-Q Plot of the 11th Largest Observations (Adjusted)**

[3] Beirlant has shown that if the behaviour can be expressed "when a non-decreasing continuous function g exists between *Y* and *X*, such as $Y = g(X)$, then the same relation exists between the corresponding quantiles of the respective distributions", i.e.
$Q_Y(p) = g(Q_X(p))$ , $\forall p \in (0;1)$, extensive work was also done by [12] Dierckx in this regard. Hence, having carried out the analyses that yielded Fig 3.4 to 3.9 (see [3] Beirlant pp 81-87, particularly top of p85), the Q-Q method emphasised the characterization of the behaviour of the GWh Losses at annual GU level, which can therefore be described by a Generalised Pareto Distribution.
**Mean Excess Function (MEF) Diagnostic** (see Appendix A paragraph 2.4)

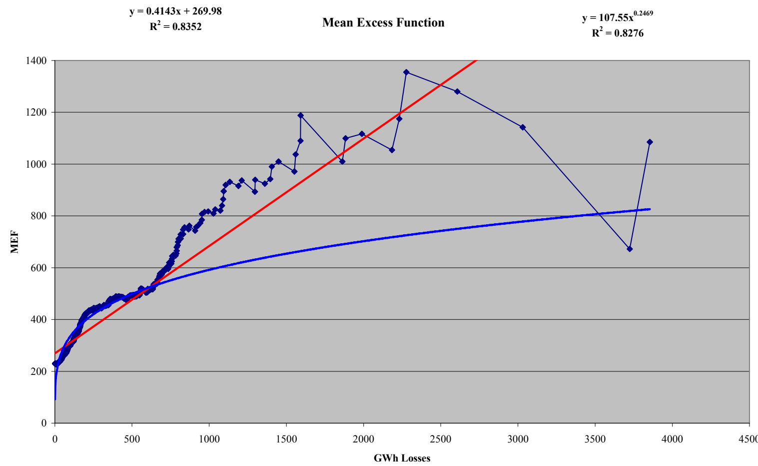



**Fig 3.10: MEF vs GWh Losses**

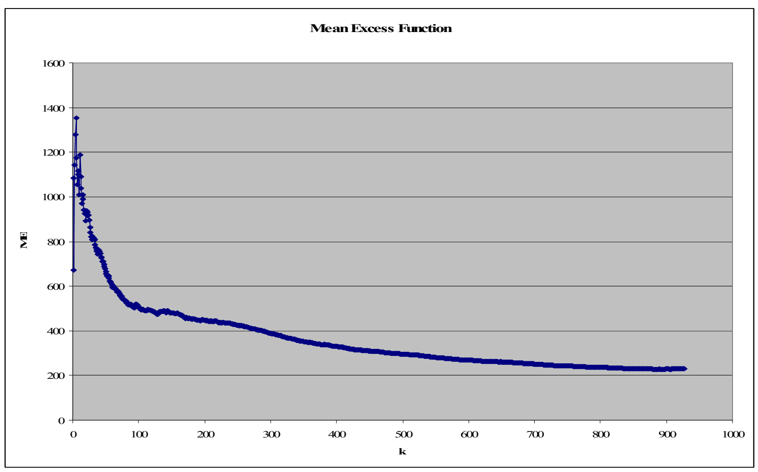



**Fig 3.11: MEF vs k**

In Fig 3.10, there is again substantiation that the GWh Losses derive from a distribution in the Fréchet class; the linear trend is more prominent than the curvy-linear one. [3] Beirlant, p39 Fig I15, shows that the linear behaviour belongs more to the Pareto type than the Gumbel or Weibull types (see also [3] Beirlant Fig I.18 p42, in comparing with Fig 3.10 and Fig 3.11).

Fig 3.11 shows the MEF versus k, the indicator that represents the largest observation, e.g. the largest observation would be the first, therefore $k$=1, the 2nd largest, $k$=2, etc…, this means that the set of $k$ would span from1 to $n$-1, where $n$ is the total number of observations. Then

$$\text{MEF} = \hat{e}(k) = \frac{1}{k}\sum_{i=1}^{k} y_{n-i+1} \quad - y_{n-k} \qquad \text{Eq 3.6}$$

where $y$ represents the ordered GWh losses

This instrument can be used in management circles when the mean GWh loss of say 20 GUs' might incur at the extremes. By using Fig 3.11 (or more specifically, computing the MEF thereof) the Mean Excess loss at $k$ = 20 is 893 GWh.

Statement: The MEF is related to the Conditional Tail Expectation (CTE).

Let the goal be in estimating the CTE $\mu_p$ then,

$\mu_p = E\,[Y|Y>y_p]$, where $y_p$ is the upper $p^{th}$ quantile and not very small (e.g. 5%). Let $(k/\text{n}) \approx p$ and define

$D_k = \frac{1}{k}\sum_{i=n-k+1}^{n} Y_i$ , then for large $n$,

$\sqrt{k}\,(\,\text{D}_k - \mu_\text{p}\,) \sim N\,[\,0\,;\ \sigma_p^2 + p(y_p - \mu_p)^2\,]$, where $\sigma_p^2 = var\,[Y\,|Y > y_p]$.

The values of $y_p$ , $\mu_p$ and $\sigma_p^2$ can be estimated from the sample by using the sample quantile $Y_{n-k}$ , the computed $D_k$ and $(S_D)^2$ , respectively; where $(S_D)^2$ is the sample variance of the top-k order statistic.

The CTE and MEF are then related by:

$E\,[Y\,|Y > \boldsymbol{c}] \;=\; \boldsymbol{c} + \hat{e}(\boldsymbol{c})$.

The MEF extends to the Hill estimator (this estimator is biased, p56 [3] Beirlant).

On the Q-Q plots of Fig 3.7 to Fig 3.9 it was mentioned that the slope is expected to be linear if the variable is behaving in the Fréchet class; further, this slope is defined as the EVI (see Appendix A paragraph 2.1) or the inverse of the *Pareto Index* (if the distribution is of the Pareto type).

The Hill estimator is given by:

$$H_{k;n} = \frac{1}{k}\sum_{i=1}^{k} Ln\,(y_{n-i+1}) \quad - Ln\,(y_{n-k}) \qquad \text{Eq 3.7}$$

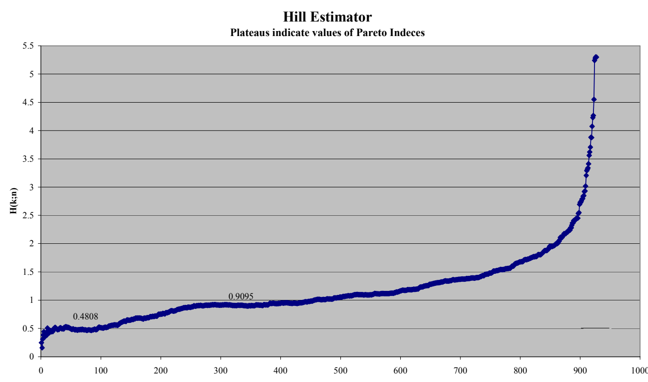



**Fig 3.12: Hill Estimator vs k**

As shown in Fig 3.12, the Hill estimator exhibits two plateaus, one at 0.4808 and another at 0.9095. If we compare this to the results obtained in Fig 3.8 (i.e. without the outlier adjustment), then the value for the EVI of 0.4808 is within the 95% CI, and this seems a plausible value.
A more thorough analysis by inspection yielded a value for the EVI of 0.47508.

**The Pickands estimator result is given in Appendix C, paragraph 2.6.**

Fig 3.12 is very similar to the one obtained by [3] Beirlant Fig I20 p 44 of a Burr distribution. Beirlant goes further in assessing an EVI by means of weighting the Hill estimator to arrive at an optimal k (pp 59-64).
This method yields an EVI of 0.4684 at k =61, which is comparable to the one obtained in Fig 3.8; its CI's can be obtained by ([3] Beirlant p64):

$$\left[\frac{\hat{\gamma}}{1+\frac{z_{\alpha/2}}{\sqrt{k}}};\frac{\hat{\gamma}}{1-\frac{z_{\alpha/2}}{\sqrt{k}}}\right],\ \therefore\ \text{at 95\% two-sided,}$$

$$\hat{\gamma} = [0.3744\ \ ; 0.6253]$$

Although this interval is wider from the upper side, than the one obtained from the Quantile regression ($\gamma$ is adjusted for the outlier), the results remain relatively comparable.
From Fig 3.9's parameters the following model for the GWh losses was derived:

In a year:

GU-GWh Losses = $\boldsymbol{t} \cdot e^{-\gamma .\, Ln\,(1\text{-}p)}$ , $0 \leq p < 1$ Eq 3.8

Here the EVI may be interpreted as the probabilistic rate at which all GUs would tend to incur the losses and this rate would induce the losses to vary, non-linearly according to $p$ (at least from $p \geq 93\%$), see Fig 3.13 below.

In Eq 3.8, t may be interpreted as the expected "background" losses or a possible threshold. For $p = 0$, Eq 3.8 reduces to $\boldsymbol{t}$ = 285.89 GWh.

The results of Eq 3.8 can essentially be summarised in the table below:

| p | 1-p | Q(p) =-Ln(1-p) | Ln(GWh) | GWh |
|---|---|---|---|---|
| 0.95 | **5.0%** | 2.9957 | 6.9177 | **1010.02** |
| 0.955 | **4.5%** | 3.1011 | 6.9621 | **1055.86** |
| 0.96 | **4.0%** | 3.2189 | 7.0117 | **1109.58** |
| 0.965 | **3.5%** | 3.3524 | 7.0680 | **1173.79** |
| 0.97 | **3.0%** | 3.5066 | 7.1329 | **1252.55** |
| 0.975 | **2.5%** | 3.6889 | 7.2098 | **1352.55** |
| 0.98 | **2.0%** | 3.9120 | 7.3038 | **1485.88** |
| 0.985 | **1.5%** | 4.1997 | 7.4250 | **1677.34** |
| 0.99 | **1.0%** | 4.6052 | 7.5958 | **1989.79** |
| 0.995 | **0.5%** | 5.2983 | 7.8878 | **2664.60** |
| 0.996 | **0.4%** | 5.5215 | 7.9818 | **2927.25** |
| 0.997 | **0.3%** | 5.8091 | 8.1030 | **3304.43** |
| 0.998 | **0.2%** | 6.2146 | 8.2738 | **3919.98** |
| 0.999 | **0.1%** | 6.9078 | 8.5659 | **5249.38** |

**Table 3.2: Quantiles, POTsp method**

Practically, from Table 3.2, one may deduce that with a fleet of 64 units, the company expects to exceed 1010 GWh once in twenty years. Another way of interpreting the table above, is that since the expectation is in GU-Years, is that, 1 / [(1-$p$). 64] = return period, i.e. for 1-$p$ = 0.1%, one would expect that a loss by a GU of 5249 GWh be equalled or exceeded every 15.63 Years (or 15 years and 8 months). Similarly, from another angle, with a p = 99.8%, one would expect that a loss by a GU of 3920 GWh be equalled or exceeded every 7.8 Years (or 7 years and 10 months). If the fleet was to increase to, say 70 GUs, then with p = 99.8%, one would expect that a loss by a GU of 3920 GWh be equalled or exceeded every 7.14 Years (or 7 years and 2 months). Also, the expectation to have the absolute maximal GWh Loss of 5935 GWh, is 0.07% or with a fleet of 64 GUs, once in 20.9 Years (or 20 years and 11 months). A comparison may be made with Fig 2.9, whereby from table 3.2 we get that there is a 5% chance that a value of 1010 GWh would be exceeded.

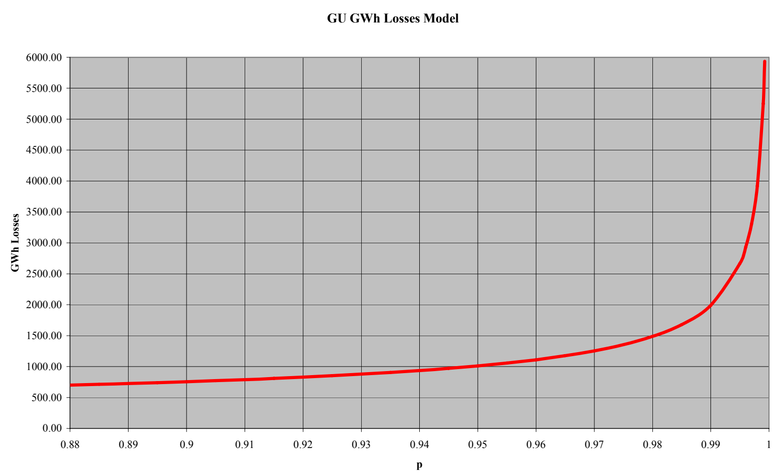



**Fig 3.13: GU-GWh Losses Tail Model**

The model may be inversed so as to derive the expected $p$'s :

$p = 1 - e^{-[Ln\,(\text{GWh})-Ln\,(\boldsymbol{t})]\,/\,\gamma}$ , $\gamma \neq 0$ Eq 3.9

Using Eq 3.9 and keeping $\boldsymbol{t}$ constant (even though its 95% CI is [241.99 ; 337.76]), but letting $\gamma$ be within its 95% CI, Fig 3.14 was obtained.

Scrutinising these results one may see that the empirical values do not fit well with a $\boldsymbol{t}$ = 285.89 GWh (the "black line" in Fig 3.14 tends to be outside the confidence bands).

A choice criterion might be to choose $\boldsymbol{t}$ =1296 GWh; that is the value at which the empirical crosses the level of significance, but the odds that this value would be the most likely $\boldsymbol{t}$, are risky.

Another choice criterion might be a $\boldsymbol{t}$ = 1591 GWh; this being the value at which the empirical is close to the expectation (the "blue line" in Fig 3.15), here $k$ = 13 largest values.

Or one may try to get even closer to expectation by using $\boldsymbol{t}$ = 1862 GWh while here $k$ = 11 largest values. Although this is not surprising, as the Q-Q plot was fitted with $k$ =11 (see Fig 3.6 to 3.9 – kink level); what is surprising is that the fit extends up to $k$ =21 with a value of $\boldsymbol{t}$ =1296 GWh.

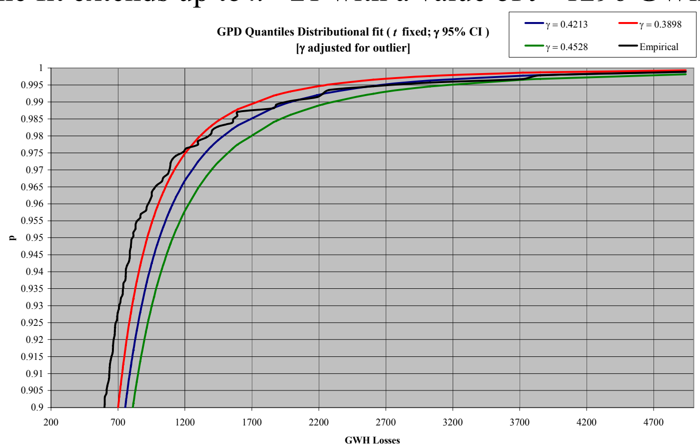



**Fig 3.14: GU-GWh Losses GPD Fit**

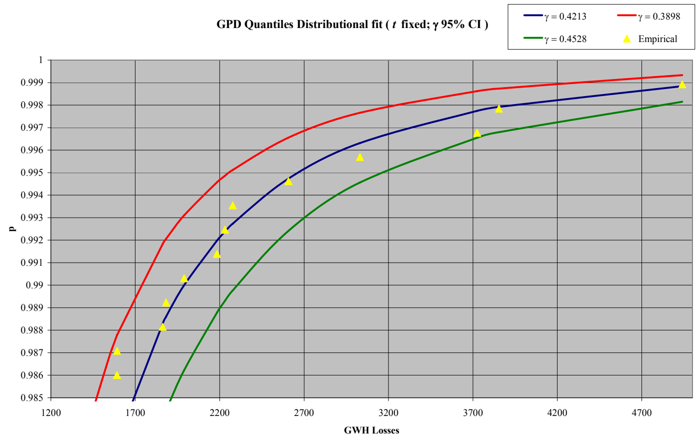



**Fig 3.15: GU-GWh Losses GPD Fit (zoomed)**

Fig 3.15 zooms-in to illustrate the fit up to $k$ =13 with a $\boldsymbol{t}$ = 1591 GWh

## 3.3.2 Full-parametric method (POTfp)

The POTsp approach yielded very good results, and as it might have been perceived, raised a few more questions. The most fundamental one from the previous paragraph was "what should the value of $\boldsymbol{t}$ be?", "how should it be chosen?" "can it be estimated effectively?

It is already known (from paragraph 3.3.1, above) that the GWh losses follow a Fréchet type of distribution. Furthering the thought, from the works of [31] Pickands, the study pinpointed onto making use of the GPD properties (see Appendix A paragraphs 2.2 and 2.3.2) to model the GWh Losses behaviour in more detail.

Let $X$ represent the ordered GWh losses and let $\boldsymbol{t}$ be the threshold, if $Y_i = X_i - \boldsymbol{t}$ ,

for $X_i > \boldsymbol{t}$, then for $\boldsymbol{t}$ – large, $Y \sim$ GPD, whereby:

$$\text{GPD}_{\gamma,\sigma}(y) = \begin{cases} 1-(1+\gamma . y/\sigma)^{-1/\gamma} & ,\text{ if } \gamma \neq 0 \text{ , } \gamma . y > -\sigma \text{ , } y > 0 \\ 1-exp\{-y/\sigma\} & ,\text{ if } \gamma = 0 \text{ , } y \geq 0 \end{cases}$$

Eq 3.10

When $\gamma > 0$, the GPD is heavy tailed, this means that the distribution does not have a complete set of moments (while the Normal, by contrast, has a complete set). Hence with the shape $\gamma > 0$, the moments are infinite for $k \geq 1/\gamma$, when $\gamma = 0.5$ the GPD has an infinite variance and when $\gamma = 0.25$ the GPD has an infinite 4th moment. This does not mean that there is no justification in a real market for infinite moments (like the 2nd and the 4th).

Usually the choice of $\hat{t}$ should be such that it's high enough so that the asymptotic properties for the GPD can be satisfied and low enough so that there is sufficient data to estimate the parameters.

Given $k$ exceedances above $\boldsymbol{t}$, two methods were used to estimate the location ($\hat{t}$), the shape ($\hat{\gamma}$) and scale ($\hat{\sigma}$) parameters: the L-Moments method and the Bayesian method (see paragraph 1.2.2 in the 1st Chapter). The Bayesian method is reflected in Chapter 4.

The estimation of percentiles may be computed as follows

Let $z_p$ = upper $p^{th}$ percentile of $\text{GPD}_{\gamma,\sigma}(x-\boldsymbol{t})$, the estimate of $z_p$ is:

$$\hat{x}_p = t + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{P(X > t)}{p}\right)^{\hat{\gamma}} - 1\right] \equiv$$

$$\hat{x}_q = t + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{n(1-q)}{N_t}\right)^{-\hat{\gamma}} - 1\right] \qquad \text{Eq 3.11}$$

where $P(X > \boldsymbol{t}) > p$ and $P(X > \boldsymbol{t})$ is estimated by the sample proportion exceeding $\boldsymbol{t}$.

That is $P(X > \boldsymbol{t}) = N_t / n$ and $p = (1 - q)$ and inverting by making the exponent $\gamma$ negative.

Let us say for instance that we would like to position our threshold at the 98.75th percentile, this means that the sample proportion $[(n - N_t) / n] = 0.9875$, since $n = 928$, $N_t = 928.(1\text{-}0.9875) = 11.6$, i.e. we ought to have 11 or 12 observations above our threshold. In Eq 3.11, say with $\boldsymbol{t} = 1725$ (at the 98.75th percentile), then:

$\hat{x}_{0.999} = 1725 + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{928(1-0.999)}{11}\right)^{-\hat{\gamma}} - 1\right]$, we still need to estimate $\hat{\gamma}$ and $\hat{\sigma}$

The reason of estimating the parameters of the tail using properties of the GPD under a threshold condition is usually because the data is very sparse at the tail end.

### 3.3.2.1 GPD Parameters Estimation

The usual preferred method is the MLE, however in this thesis a different approach is used. The first is the one of L-Moments (LM) and the second one, which is given in Chapter 4 is the Bayesian method.

Whenever using these methods, the cautionary statement (underlined) made in the paragraph above must be kept in mind.

The LM method (see Appendix A paragraph 2.3, aspect **IV**) resulted as an excellent method, particularly for the assessment of the threshold. The PWM is the precursor to the LM but with significant advantages, notably the ability to summarise a statistical distribution in a more meaningful way: they are linear functions of the sample values, they are virtually unbiased and have relatively low sample variance. LM ratio estimators have also a small bias and variance especially in comparison with the classical coefficients of skewness and kurtosis. LM estimators are relatively insensitive to outliers

The LM method has shown its power in the case of heterogeneous data sets; the quantile estimation for a Pareto distribution gives the smallest bias and RMSE when compared with MLE, LS and classical Moments [40] van Gelder.

Assuming a threshold $\boldsymbol{t}$, and the GWh Losses represented by $X$ with a distribution function $F_X(x)$, consider $Y = X - \boldsymbol{t}$ , [31] Pickands has shown that asymptotically, the distribution of $Y$ (i.e. the $X$'s conditional on $\boldsymbol{t}$), follows a GPD, given in Eq 3.10, above.

This distribution is ***unbounded*** (i.e. $0 < y < \infty$), if $\gamma \geq 0$ and ***bounded*** (i.e. $0 < y < \{\sigma/\gamma\}$), if $\gamma < 0$ (Eq 3.10).

Eq 3.10 has a unique threshold property, i.e. if $X$~GPD then $Y$, the conditional distribution of the excesses, also follows a GPD with the same $\gamma$ in $X$'s GPD.

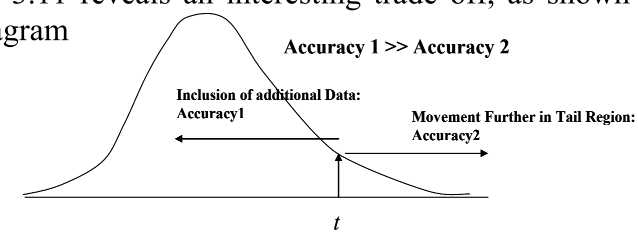
Computations of quantiles of exceedances from Eq 3.11 of a value $x_r$, based on a $r$-year period, corresponds to a period $N_{\boldsymbol{t}}\ .\ r\ /\ n$, where $\xi = \hat{N}_t$ is the Mean Exceedance rate per year. From Table 2.4 in Chapter 2, Table 3.3 below was obtained,

| | No. Exceedances over a threshold of t = 400 |
|---|---|
| 2005 | 7 |
| 2004 | 7 |
| 2003 | 3 |
| 2002 | 1 |
| 2001 | 0 |
| 2000 | 0 |
| 1999 | 1 |
| 1998 | 0 |
| 1997 | 1 |
| 1996 | 1 |
| 1995 | 3 |
| 1994 | 2 |
| 1993 | 2 |
| 1992 | 6 |
| 1991 | 3 |
| 1990 | 2 |
| Total | 39 |
| Tot Years | 16 |
| Mean Exceedance Rate | 2.44 |

**Table 3.3: Mean Exceedance Rate**

The table above shows that, should threshold of 400 GWh be used then, $\xi$, the Mean Exceedance rate is 2.44 GUs p.a. That is one would expect 2.44 GUs each year to exceed the threshold of 400 GWh (derived from a total of 39 GUs over a period of 16 years). The recurrence period would be $1/\xi = 0.41$, i.e. one would expect one GU every (0.41x12) approx. 5 months, to exceed a loss of 400 GWh. Again this aspect shows the importance of not only of estimating the shape and scale but also of the threshold.

Eq 3.11 reveals an interesting trade off, as shown in this diagram




Hence an optimal threshold might exist that would result in a minimal trade off of accuracies

[7] Davison et al, have also shown that the distribution of the maximum excesses,

$Z = Max\ \{Y_i\}$, i= 1, 2…, k follows a GEV, provided that the exceedances over the threshold are generated from a Poisson process.

Returning to L-Moments, the method of 3-LM was used so as to get an indication of the threshold estimate. The 3-LM was initially applied to the full dataset and yielded to following result (refer to Appendix A paragraph 2.3, aspect **IV)**:

| r | 1 | 2 | 3 |
|---|---|---|---|
| b(r) | *186.1983* | *161.1779* | *144.0855* |
| b(0) | 229.6999 | | |

| | | |
|---|---|---|
| | λ1 | 229.6999 |
| | λ2 | 142.6968 |
| | λ3 | 79.57717 |
| | λ4 | 51.0534 |
| L-CV | τ | 0.621231 |
| | τ3 | 0.557666 |
| | τ4 | 0.357775 |

| | | |
|---|---|---|
| Shape | γ | 0.432056 |
| Scale | σ | 127.0721 |
| Location | t | 5.959328 |

**Table 3.4: 3-LM Parameter Estimates - Full Dataset**

The indicator for the Threshold is 6 GWh as expected from the Full dataset; comparing it to the one obtained by the Two-Way Joining method in Cluster Analysis: 449.1 GWh (Fig 2.3 the lowest value in the Legend) it indicates that further selection is required. The shape parameter, with the full dataset, shows that the full distribution, if of the Pareto type, would be unbounded ($\gamma > 0$).

As suggested by [30] Pandey et al, when the 3-LM is applied on a $Y = X - \boldsymbol{t}$, type of variable, it is very efficient under the GPD domain (but still exercise caution as mentioned above).

Under this premise, the moments were computed and the results are tabulated here below.

Threshold values were selected according to various criteria, for instance, the value of 484 was chosen as this seemed the expected background average GWh loss, resulting primarily from the work in Chapter 4.

| | | Shape | Scale | Location | | |
|---|---|---|---|---|---|---|
| | t | γ | σ | u | k | MEF |
| | 150 | 0.30693345 | 253.959047 | -14.532722 | 351 | 359.99 |
| | 375 | 0.34645987 | 288.806697 | 31.8118215 | 155 | 458.78 |
| | 403.4 | 0.36770832 | 280.830467 | 39.9049166 | 143 | 467.35 |
| | 449.1 | 0.38927422 | 272.856838 | 30.3970655 | 132 | 466.15 |
| | 484 | 0.40434481 | 271.215555 | 32.7506348 | 120 | 477.55 |
| | 878 | 0.36912826 | 429.504954 | 74.3137793 | 39 | 724.29 |
| | 1076 | 0.48827845 | 419.759019 | -3.7451523 | 27 | 816.71 |
| Q-Q Plot | 1296 | 0.40722789 | 436.672533 | 114.265318 | 21 | 815.16 |
| Q-Q Plot | 1591 | 0.44453959 | 414.14702 | 261.011381 | 13 | 954.48 |
| | 1796 | 0.53903737 | 311.70399 | 308.219593 | 11 | 1036.63 |
| Q-Q Plot | 1862 | 0.53903737 | 311.70399 | 242.219593 | 11 | 959.45 |
| | 1917 | 0.58509384 | 268.114843 | 418.941368 | 9 | 1236.99 |
| | 2252 | 0.67061101 | 185.041956 | 591.941207 | 6 | 1766.92 |

**Table 3.5: 3-LM Parameter Estimates–Various Thresholds**

The value of 403.4 GWh was obtained visually from the full distribution Q-Q plot of Fig 3.7 in Chapter 3 (giving an intercept of 6 and calculating $e^6$ = 403.4). The values of 449.1 and 1796 were obtained from the lowest threshold in the two-way joining Cluster Analysis results of Fig 2.3. The value of 1076 was arbitrarily chosen (almost as a mid-point), while the value of 2252 was taken from the results of the entropy method (see Chapter 4 and Appendix C) at $k$ = 6. The values of 1296, 1591 and 1862 were as a direct result from the POTsp methodology.

Using Table 3.5, the following graphs were produced:

Fig 3.16 a) indicates the $\gamma$ increases after $\boldsymbol{t}$ > 878 GWh, with a marked kink at 1862 GWh, before rising.

The kink is also evident in Fig 3.16 b) with a decrease in MEF at $\boldsymbol{t}$ = 1862 GWh

The location parameter also drops at this level. These kinds of effects seem to be indicative that the parameters yielded from using $\boldsymbol{t}$ = 1862 GWh would be the ones to use as a preferred choice, i.e.

| Shape | Scale | Location |
|---|---|---|
| γ | σ | u |
| 0.53903737 | 311.70399 | 242.219593 |

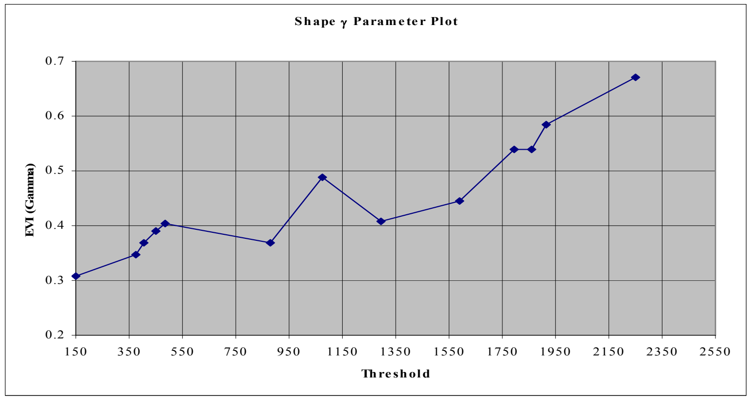



**Fig 3.16 a) : EVI, γ at various threshold levels**

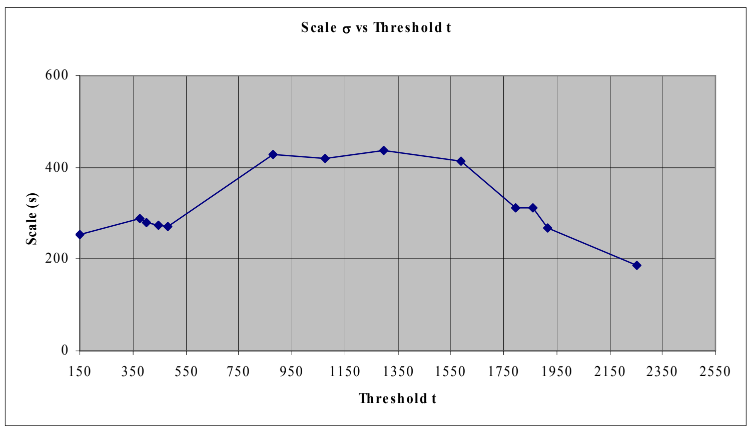


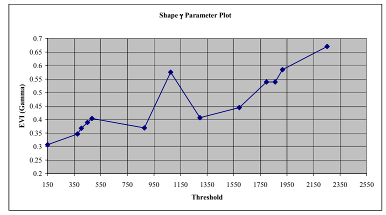

**Fig 3.16 b): EVI, σ at various threshold levels**

### 3.3.2.2 GWh Quantile Model

| | | Shape | Scale | Location | t = 1862 |
|---|---|---|---|---|---|
| | t | γ | σ | u | k |
| | 150 | 0.30693345 | 253.959047 | -1726.5327 | 351 |
| | 375 | 0.34645987 | 288.806697 | -1455.1882 | 155 |
| | 403.4 | 0.36770832 | 280.830467 | -1418.6951 | 143 |
| | 449.1 | 0.38927422 | 272.856838 | -1382.5029 | 132 |
| | 484 | 0.40434481 | 271.215555 | -1345.2494 | 120 |
| | 878 | 0.36912826 | 429.504954 | -909.68622 | 39 |
| | 1076 | 0.57558759 | 292.836049 | -659.43715 | 27 |
| Q-Q Plot | 1296 | 0.40722789 | 436.672533 | -451.73468 | 21 |
| Q-Q Plot | 1591 | 0.44453959 | 414.14702 | -9.9886194 | 13 |
| | 1796 | 0.53903737 | 311.70399 | 242.219593 | 11 |
| Q-Q Plot | 1862 | 0.53903737 | 311.70399 | 242.219593 | 11 |
| | 1917 | 0.58509384 | 268.114843 | 473.941368 | 9 |
| | 2252 | 0.67061101 | 185.041956 | 981.941207 | 6 |

| Shape | Scale | Location |
|---|---|---|
| γ | σ | u |
| 0.53903737 | 311.70399 | 242.219593 |

| Shape | Scale | Location |
|---|---|---|
| γ | σ | u |
| 0.44453959 | 414.14702 | -9.9886194 |



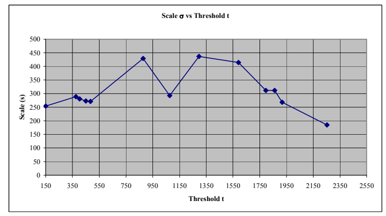




**Fig 3.17: EVI, $\sigma$ at various threshold levels – 3LM (Known Threshold)**

Fig 3.17 indicates that the parameter estimates, when the thresholds are known, differ substantially, but are still in concordance with results obtained from the POTsp. The shift occurs in that thresholds below 1000 GWh do not seem plausible, and that parameters associated with a threshold of 1591 GWh (also a value from the Q-Q plot) seem more plausible. Therefore the following values were adopted:

| Shape | Scale | Location |
|---|---|---|
| γ | σ | u |
| 0.44453959 | 414.14702 | -9.9886194 |

Now, one may make usage of Eq 3.11:

$$\hat{x}_q = t + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{n(1-q)}{N_t}\right)^{-\hat{\gamma}} - 1\right]$$

where $P(X > \boldsymbol{t}) > p$ and $P(X > \boldsymbol{t})$ is estimated by the sample proportion exceeding $\boldsymbol{t}$.
Unfortunately, the parameters above yielded a poor fit. The parameters were solved by a squared error (keeping the threshold constant and adding the last percentile) minimizing algorithm that yielded the following new parameters:

| | | | | |
|---|---|---|---|---|
| **n** | **928** | | | |
| | | **Shape** | **Scale** | **Location** |
| | **Nt** | **γ** | **σ** | **u** |
| | **13** | **0.3964544** | **756.88837** | **1591** |
| | | **0.4445396** | **414.14702** | **1591** |

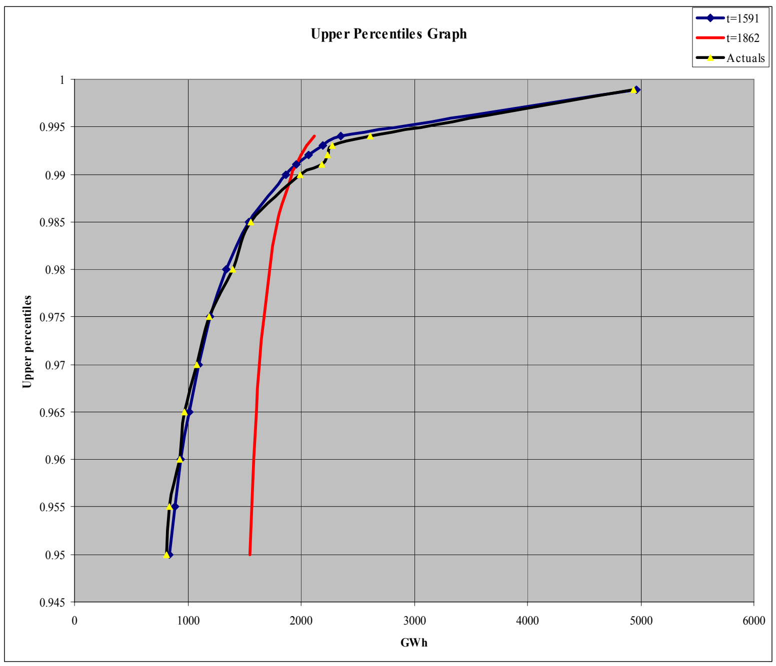



**Fig 3.18: Upper Percentiles; re-parameterisation $N_t = 13$**

| Nt | Shape | Scale | Location |
|---|---|---|---|
| | γ | σ | u |
| 11 | 0.2801847 | 908.44831 | 1841.8416 |
| | 0.5390374 | 311.70399 | 1862 |

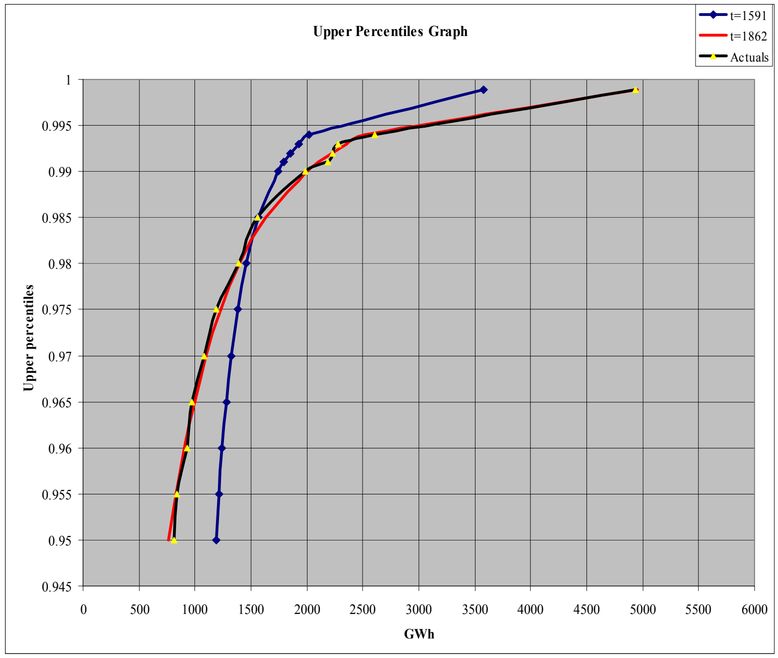



**Fig 3.19: Upper Percentiles; re-parameterisation $N_t = 11$**

For $N_t$ = 13 (giving a min of 756.8 GWh), the fit for the model with a threshold of 1591 when re-parameterised is shown in Fig 3.18 (blue line) while the model with a threshold of 1862 was kept with the original parameters (red line), for comparative reasons. The actual, empirical percentiles are given by the black and yellow curve. The necessity for adjusting the parameters is evident. Using the same algorithm, the model with the 1862 threshold did not perform as well. By contrast when reducing $N_t$ to 11, it (the red line) performed much better (Fig 3.19) than the model with the 1591 threshold; the latter was then fixed with its original parameters for comparative reasons.

From Eq 3.10, the final adopted model for the GWH Loss quantiles is:

$$\hat{x}_q = 1841.8 + \frac{908.5}{0.2802}\left[\left(\frac{928(1-q)}{11}\right)^{-0.2802} - 1\right]$$

Eq 3.12

for GWh Losses above 765.8 GWh.

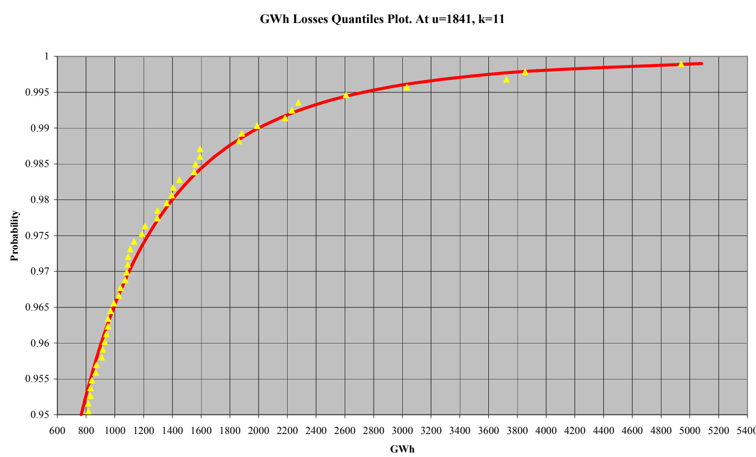



**Fig 3.20: GWh Losses Quantiles**

The Eq 3.12 can be expressed in graphical form as shown in Fig 3.20 and may be used as follows: for a GWh Loss greater than 2700 GWh, reading off the graph (or using Table 3.6 below for more exact values), there is a
(1 – 0.995) = 0.005 chance of occurrence; now, for a fleet of 64 GUs one would expect to see a GU with such a loss, or greater, once every 3.125 years (1/[0.005x64]). On the average, this seems consistent with the observations in Table 2.4.

| $\gamma$ | $\sigma$ | u |
|---|---|---|
| 0.28018471 | 908.448309 | 1841.84157 |

| Percentile | GWh Loss | Percentile | GWh Loss | Percentile | GWh Loss |
|---|---|---|---|---|---|
| 0.95 | 765.76 | 0.967 | 1033.22 | 0.984 | 1580.47 |
| 0.951 | 778.05 | 0.968 | 1054.29 | 0.985 | 1634.87 |
| 0.952 | 790.67 | 0.969 | 1076.23 | 0.986 | 1694.11 |
| 0.953 | 803.64 | 0.97 | 1099.08 | 0.987 | 1759.04 |
| 0.954 | 816.96 | 0.971 | 1122.94 | 0.988 | 1830.70 |
| 0.955 | 830.66 | 0.972 | 1147.87 | 0.989 | 1910.44 |
| 0.956 | 844.75 | 0.973 | 1173.97 | 0.99 | 2000.05 |
| 0.957 | 859.26 | 0.974 | 1201.34 | 0.991 | 2101.93 |
| 0.958 | 874.21 | 0.975 | 1230.09 | 0.992 | 2219.44 |
| 0.959 | 889.62 | 0.976 | 1260.35 | 0.993 | 2357.44 |
| 0.96 | 905.51 | 0.977 | 1292.27 | 0.994 | 2523.30 |
| 0.961 | 921.93 | 0.978 | 1326.01 | 0.995 | 2728.95 |
| 0.962 | 938.90 | 0.979 | 1361.79 | 0.996 | 2995.37 |
| 0.963 | 956.44 | 0.98 | 1399.81 | 0.997 | 3364.37 |
| 0.964 | 974.60 | 0.981 | 1440.34 | 0.998 | 3937.62 |
| 0.965 | 993.42 | 0.982 | 1483.70 | 0.999 | 5081.85 |
| 0.966 | 1012.95 | 0.983 | 1530.26 | | |

**Table 3.6: Percentiles for GWh Losses–Final Adopted Model**

These satisfactory findings conclude the work in this Chapter.

## CHAPTER 4

# Bayesian Approach to the GPD Fit

## 4.1 Introduction

The workings in this Chapter utilise extensively de Waal's contribution in [2] Beirlant, Chapter 11.
After using LM method to estimate the GPD parameters, such as the EVI and estimating the threshold ***t***, one ought to consider that the EVI would have a confidence interval (in the classical frequentist terminology) and hence the results obtained in Chapter 3, paragraph 3.3.2.2, would not be sufficient: a confidence band or region ought to be determined.
However, with reference to the classical sense, in obtaining a $(1-\alpha)\%$ confidence interval, one might be solving a different issue, since, from the author's philosophical perspective, it is difficult to conceptualise repeatability when events are supposed to be rare; in this classical sense the "IF the experiment could be repeated" is a big "if", and in the author's belief, an assumption on the prior is far more plausible than a big "IF".
One might need to answer what is the probability distribution of the worst GWh losses that would be observed in 20xx. This refers to a predictive distribution of an <u>unobserved variable</u>. Hence a predictive probability

would have to be constructed for 20xx conditional on the event of, say, a "new" catastrophic loss would happen. It is not evident that this issue may be solved by a classical approach; however by using a Bayesian approach, one may define the issue as follows.

Let the GWh Loss be represented by the dataset $X = (x_1, x_2, \ldots, x_n)$ and be modelled by the conditional predictive distribution $f_X\,[x|(\gamma,\sigma)]$, essentially this is what was done in Chapter 3 and $f_X\,[x|(\gamma,\sigma)]$is also referred to as the Likelihood of $(\gamma,\sigma)$.

Let $\boldsymbol{\pi}\,[(\gamma,\sigma)\,|\,X]$ be the posterior density of the parameters $(\gamma,\sigma)$ given the past data $X$.

Let $\boldsymbol{\pi}\,(\gamma,\sigma)$ be the prior denoting the density of the parameters.

Now, the Likelihood of $(\gamma,\sigma)$,

$f_X\,[x|(\gamma,\sigma)] = \prod_{i=1}^{n} f_X\,[x_i|(\gamma,\sigma)]$, assuming that the GWh losses are independent. According to Bayes' Theorem:

$$\boldsymbol{\pi}\,[(\gamma;\sigma)\,|\,X] = \frac{f_X(x\,|\,\gamma;\sigma)\;\pi(\gamma;\sigma)}{\int_{\Omega} f_X(x\,|\,\gamma;\sigma)\;\pi(\gamma;\sigma)\;d(\gamma;\sigma)}, \qquad \text{Eq 4.1}$$

where $\Omega$ is the parameters' space

Now the denominator:

$$\int_{\Omega} f_X(x\,|\,\gamma;\sigma)\;\pi(\gamma;\sigma)\;d(\gamma;\sigma) = \text{constant}$$

Therefore Eq 4.1 can be expressed as (the symbol $\propto$ means "proportional to"):

$$\boldsymbol{\pi}\,[(\gamma;\sigma)\,|\,X] \propto \boldsymbol{\pi}\,(\gamma;\sigma)\ \text{x}\ f_X\,[x|(\gamma;\sigma)] \qquad \text{Eq 4.2}$$

i.e. **Posterior Density $\propto$ Prior Density x Likelihood**

Estimates of $\gamma, \sigma$ can be obtained from the mode, median or mean of the Posterior.

The credibility around these estimates can be described by the Posterior itself in terms of the "***highest posterior density*** (hpd) ***region***", according to a certain probability $(1 - \alpha)$, being the hpd region which will contain $100(1 - \alpha)\%$ of the values, with no need to fall back into asymptotic theory (as in the classical way when obtaining confidence intervals).

Another attractive aspect is the ease of prediction: if $X_{n+1}$ is a future prediction with density $f_X\,[x_{n+1} \mid (\gamma,\sigma)]$, then the predictive distribution, over the parameter space $\Omega$, is given by:

$$f_X[x_{n+1} \mid X] = \int_{\Omega} f_X\,[x_{n+1}|(\gamma;\sigma)]\ .\ \boldsymbol{\pi}\,[(\gamma;\sigma) \mid X]\ \ d(\gamma;\sigma)\ ,$$

Hence this predictive density reflects its uncertainties through the future observations' uncertainties and the ones of the Posterior.

This is excellent logic in approaching the challenge, measures ahead than the classical approach, however the million dollar question hinges around the Prior: which prior one chooses á-priorí? When calling onto a prior density by some means of information or supposition, the process is called elicitation.

For the estimation of the $\gamma$ and $\sigma$ parameters and in the elicitation of the prior, the variance-covariance matrix of $(\hat{\gamma}, \hat{\sigma})$ is approximated by using the sample Fisher-Information Matrix (FIM), which in turn is used to define Jeffreys' prior, $J\,(\gamma,\sigma)$.

$J\,(\gamma;\sigma) \propto \sqrt{|\,I(\gamma,\sigma)\,|}$ , where $I(\gamma;\sigma)$ is the FIM with

$I\ (\theta) = \mathrm{E}\ \left\{ \dfrac{\partial^2 \ln \mathrm{f_X}(\mathrm{X}\,|(\theta)}{\partial^2\theta} \right\}$, where $\theta$ is the distribution's parameters vector.

Jeffreys' prior is considered the standard starting rule for an objective Bayesian analysis.

Another prior is the Maximal Data Information (MDI) prior by [43] Zellner, designed to provide maximal average data information on the $\gamma$ and $\sigma$ parameters, and is defined as:

$\boldsymbol{\pi}(\gamma;\sigma) \propto exp\ \{\ E[f_X\ (X|(\gamma;\sigma))]\}$

In this chapter we will address the fit of the Log-Normal and the GPD to the GWh losses using MCMC. The prior on the two parameters that we will consider will be an objective prior, namely the MDI prior by Zellner (1977) (see bottom of page 447). A big advantage of the Bayesian approach is that confidence intervals on the parameters are easily to establish and also the prediction of high quantiles with confidence limits are easy to compute.

Let us begin with a preamble resulting from the works of [23] Kedem et al.

Let $X$ represent the GWh Losses and let $X(\omega)$ be the r.v. with a distribution that indicates a GU not failing with probability $1 - p$ and within this event, it admits a continuum of GWh Losses. Therefore, if $G_X(x)$ denotes the distribution function of $X$, then it is of the mixed type resulting in the following:

$G_X(x) = (1 - p)\ H_X(x) + p\ F_X(x)$,

where $H_X(x) = 1$ and $F_X(x) = 0$ if $x = 0$ and

$H_X(x) = 0$ and $F_X(x) = F_X(x)$ if $x > 0$
with $H_X(x)$ being discrete and $F_X(x)$ being continuous and $p$ representing the probability of a GU failing.
To establish a linear relationship between X's expectation ***and*** the expected value of an integrable function of $X$, we proceed as follows:
Let $\varphi(X)$ be an arbitrarily integral function of X. Then:

$$E[\varphi(X)] = (1 - p)\,\varphi(0) + p\,E[\varphi(X) \mid X > 0] \qquad \text{Eq 4.3}$$

In particular for $\varphi(X) = X$, Eq 4.3 results in:

$$E[X] = p\,E[X \mid X > 0] \qquad \text{Eq 4.4}$$

from Eq 4.4, $p = E[X] / E[X \mid X > 0]$ Eq 4.5
substituting $p$ in Eq 4.3 :

$$E[\varphi(X)] = (1 - \frac{E[X]}{E[X \mid X > 0]})\,\varphi(0) + \frac{E[X]}{E[X \mid X > 0]} E[\varphi(X) \mid X > 0]$$

Solving for $E[X]$, we get:

$$E[X] = \frac{E[X \mid X > 0]}{E[\varphi(X) \mid X > 0] - \varphi(0)} \left\{ E[\varphi(X)] - \varphi(0) \right\}$$

$$E[X] = \beta_\varphi \left\{ E[\varphi(X)] - \varphi(0) \right\} \qquad \text{Eq 4.6}$$

whereby $\beta_\varphi$ now becomes the slope of the linear relationship between $E[X]$ and $E[\varphi(X)]$; this slope only depends on the continuous part of the distribution of X and $\varphi(0)$, hence the slope is independent of the probability of a GU failing, it depends only on the GWh Losses' distribution i.e. $F_X(x)$, assuming that it exists, and if $\varphi(0) = 0$, which means that then there is no intercept in Eq 4.6.

Let us now fix a time $\tau$ and let $X_\tau(a)$ be a r.v. that yields a GWh Loss at time $\tau$. Again, $X_\tau(a)$ would have a mixed distribution governed by $p_\tau$ the probability of a GU failing at time $\tau$ and its losses behaving as $f_{X\tau}(x_\tau)$.

Equation Eq 4.6 can now be modified as:

$$E[X_\tau] = \beta_\varphi(\tau)\{E[\varphi(X_\tau)] - \varphi(0)\}, \qquad \text{Eq 4.7}$$

where, $\beta_\varphi(\tau) = \dfrac{E[X_\tau \mid X_\tau > 0]}{E[\varphi(X_\tau) \mid X_\tau > 0] - \varphi(0)}$ Eq 4.8

But now, $\beta_\varphi(\tau)$ depends on $\tau$. By making use of the simplifying homogeneity assumption, we assume that $\beta_\varphi(\tau)$ is independent of $\tau$. The homogeneity assumption would state that the continuous part of the distribution of the GWh Losses is homogeneous in the time dimension: i.e. $\forall\tau$, $f_{X\tau}(x_\tau) = f_X(x)$. No similar assumption was made on $p_\tau$.

$$p_\tau \Rightarrow \beta_\varphi(\tau) = \beta_\varphi. \qquad \text{Eq 4.9}$$

However, there is an underlying assumption that the failure rates (linked to $p_\tau$) are not homogeneous with time, i.e. $p_\tau$ might be smaller than $p_{\tau+\kappa}$.

Let $\overline{X}_\tau$ and $\overline{\varphi(X_\tau)}$ represent the sample averages obtained at time $\tau$, and let $\boldsymbol{t}$ represent a threshold such that:

$$\varphi(x) = 1 \quad , x > t \qquad \text{Eq 4.10}$$
$$\varphi(x) = 0 \quad , x \le t$$

then, $\varphi(0) = 0 \quad , E[\varphi(X_\tau)] = P(X_\tau > t)$

From Eq 4.7, Eq 4.8 and Eq 4.9

$$E[X_\tau] = \beta_\varphi(t)\, P(X_\tau > t) \qquad \text{Eq 4.11}$$

and under homogeneity conditions,

$$\beta_\varphi(t) = \frac{E[X \mid X > 0]}{P(X > t) \mid X > 0)} \qquad \text{Eq 4.12}$$

This means that now $\beta_\varphi(t)$ depends only on $f_X(x)$ and $\boldsymbol{t}$ and not on time $\tau$.

Let $I[X_\tau > t]$ be the indicator function of the event $X_\tau > t$, then $\langle I[X_\tau > t] \rangle$ is denoted as the average number of GUs failing per period $\tau$ above a critical level $\boldsymbol{t}$ of the GWh Losses (this can be estimated by counting the cells with an event from table 2.4 in Chapter 2, after choosing a certain critical level $\boldsymbol{t}$ and dividing it by the ave. No. of total GUs p.a.).

Since $\overline{X}_\tau$ is the approximation of $E[X_\tau]$ and $\langle I[X_\tau > t] \rangle$ is the approximation of $P(X_\tau > t)$, then from Eq 4.11

$\overline{X}_\tau \cong \beta_\varphi(t) \langle I[X_\tau > t] \rangle$, where $\beta_\varphi(t)$ is a constant for each critical level $\boldsymbol{t}$.

The linear relationship in Eq 4.12 indicates that one may expect a high correlation between the average GWh Losses and the average number of GUs failing, then the question arises "but at what level of $\boldsymbol{t}$?".

Eq 4.12 starts to offer a direction in the provision of an answer. If $\boldsymbol{t}$ can be chosen so as to minimise the variance of the MLE of $\beta_\varphi(t)$ under an appropriate choice of $f_X(x)$ then the answer to the question above would be provided. As noted in [23] Kedem et al, since the study centers on Eq 4.10, the notation on the slope changes to focus more on the vector parameter $\theta$ stemming from the choice of $f_X(x)$.

In this case $\beta_\varphi(t) \equiv \beta_\theta(t)$. Eq 4.13

The process would then be to fix a threshold $\boldsymbol{t}$ and then derive the MLE of $\beta_\theta(t)$ under the choice of $f_X(x)$. Note the change in the notation from $\varphi$ to $\theta$. From the MLE theory:

$$\sqrt{n}\,(\hat{\theta} - \theta) \xrightarrow{L} N[\mathbf{0}; \mathbf{I}^{-1}(\theta)] \qquad \text{Eq 4.14}$$

where $\mathbf{I}(\theta)$ is the Fisher Information Matrix (see Appendix C, paragraph 3.1 and paragraph 3.2 for the Delta-Method) and assessed with the choice of $f_X(x)$.

The asymptotic distribution of $\beta_{\hat{\theta}}(t)$ is obtained by invoking the Delta-Method and is expressed by:

$$\sqrt{n}\ (\beta_{\hat{\theta}}(t) - \beta_{\theta}(t)) \xrightarrow{L} N[\mathbf{0}; \nu_{\theta}(t)] \qquad , n \to \infty \qquad \text{Eq 4.15}$$

where for a 2-parameter pdf (i.e. $\theta = (\theta_1\ ;\ \theta_2)$ see Eq C3.26 Appendix C),

$$\nu_{\theta}(t) = \begin{pmatrix} \dfrac{\partial \beta}{\partial \theta_1} & \dfrac{\partial \beta}{\partial \theta_2} \end{pmatrix} \mathbf{I}^{\mathbf{-1}}(\theta) \begin{pmatrix} \dfrac{\partial \beta}{\partial \theta_1} \\ \dfrac{\partial \beta}{\partial \theta_2} \end{pmatrix} \qquad \text{Eq 4.16}$$

# 4.2 The Fit of the Tail – Log Normal and GPD

## 4.2.1 Log Normal

Let us assume that as a choice of pdf for the GWh Losses, the Log-Normal distribution is selected (this is not an Extreme Value distribution). As expressed in the paragraph above Eq 4.3, the GWh Losses represented by X are essentially positive; hence the assumption of Log-Normal behaviour in this variable is not far-fetched; the example here is used to illustrate the position of the threshold if using this assumption (see Fig 1.2).

In this case then: $\theta = \begin{pmatrix} \theta_1 = \mu \\ \theta_2 = \sigma \end{pmatrix}$ and

$$f_X(x;\theta) = \frac{1}{\sqrt{2\pi}\,\sigma\, x} e^{-\frac{(Ln(x)-\mu)^2}{2\sigma^2}} \quad , x > 0 \,, -\infty < \mu < \infty \,, \sigma > 0$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Ln(X_i) \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\{Ln(X_i) - \hat{\mu}\}^2$$

$$E[X \mid X > 0] = e^{(\mu + \frac{\sigma^2}{2})} \quad \text{and} \quad P(X > t) = 1 - \Phi\,[\frac{Ln(t) - \mu}{\sigma}]$$

and $\phi\,[\frac{Ln(t)-\mu}{\sigma}]$ is the density of the Standard Normal distribution. While $\Phi\,[\frac{Ln(t)-\mu}{\sigma}]$ is the distribution function of the Standard Normal.

In essence, as mentioned in the paragraph above, the derivation of an "optimal" ***t***, that gauges the average GWh Losses and the average number of GUs failing, can now be assessed. To enable us to do that we need to derive the MLE $\beta_{\hat{\theta}}(t)$ from the GWh Losses data set. From Eq 4.12 and Eq 4.13 above:

$$\beta_\theta(t) = \frac{E[X \mid X > 0]}{P(X > t) \mid X > 0)} = \frac{e^{(\mu + \frac{\sigma^2}{2})}}{1 - \Phi\,[\frac{Ln(t) - \mu}{\sigma}]} \qquad \text{Eq 4.17}$$

or, as estimates from the dataset:

$$\beta_{\hat{\theta}}(t) = \frac{E[X \mid X > 0]}{P(X > t) \mid X > 0)} = \frac{e^{(\hat{\mu} + \frac{\hat{\sigma}^2}{2})}}{1 - \Phi\,[\frac{Ln(t) - \hat{\mu}}{\hat{\sigma}}]}$$

The matrix $\mathbf{I}^{-1}(\theta)$ derived from inverting Eq C3.6 in Appendix C yields:

$$\mathbf{I}^{-1}(\theta) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{bmatrix}$$

Then from the Delta Method (see Appendix C & paragraph 3.2) and Eq 4.15, Eq 4.16:

$$\nu_{\theta}(t) = \begin{pmatrix} \frac{\partial \beta}{\partial \mu} & \frac{\partial \beta}{\partial \sigma} \end{pmatrix} \mathbf{I}^{-1}(\theta) \begin{pmatrix} \frac{\partial \beta}{\partial \mu} \\ \frac{\partial \beta}{\partial \sigma} \end{pmatrix}$$

$$= \sigma^2 \left[ \left( \frac{\partial \beta}{\partial \mu} \right)^2 + \frac{1}{2} \left( \frac{\partial \beta}{\partial \sigma} \right)^2 \right] \qquad \text{Eq 4.18}$$

from Eq 4.17:

$$\frac{\partial \beta_{\theta}(t)}{\partial \mu} =$$

$$\frac{e^{(2\mu + \sigma^2)\frac{1}{2}}}{\left( 1 - \Phi\,[\frac{Ln(t) - \mu}{\sigma}] \right)^2} \cdot \left\{ 1 - \Phi\,[\frac{Ln(t) - \mu}{\sigma}] - \frac{1}{\sigma} \phi[\frac{Ln(t) - \mu}{\sigma}] \right\}$$

$$\therefore \left(\frac{\partial \beta_\theta(t)}{\partial \mu}\right)^2 =$$

$$\left(\frac{e^{(2\mu+\sigma^2)\frac{1}{2}}}{\left(1-\Phi\,[\frac{Ln(t)-\mu}{\sigma}]\right)^2}\right)^2 \cdot \left\{1-\Phi\,[\frac{Ln(t)-\mu}{\sigma}]-\frac{1}{\sigma}\phi[\frac{Ln(t)-\mu}{\sigma}]\right\}^2$$

Eq 4.19

Similarly,

$$\frac{\partial \beta_\theta(t)}{\partial \sigma} =$$

$$\frac{e^{(2\mu+\sigma^2)\frac{1}{2}}}{\left(1-\Phi\,[\frac{Ln(t)-\mu}{\sigma}]\right)^2} \cdot \left\{1-\Phi\,[\frac{Ln(t)-\mu}{\sigma}]-\frac{1}{\sigma}\,\{\frac{Ln(t)-\mu}{\sigma}\}\,\phi[\frac{Ln(t)-\mu}{\sigma}]\right\}$$

$$\therefore \left(\frac{\partial \beta_\theta(t)}{\partial \sigma}\right)^2 =$$

$$\left(\frac{e^{(2\mu+\sigma^2)\frac{1}{2}}}{\left(1-\Phi\,[\frac{Ln(t)-\mu}{\sigma}]\right)^2}\right)^2 \cdot \left\{1-\Phi\,[\frac{Ln(t)-\mu}{\sigma}]-\frac{1}{\sigma}\,\{\frac{Ln(t)-\mu}{\sigma}\}\,\phi[\frac{Ln(t)-\mu}{\sigma}]\right\}^2$$

Eq 4.20

Hence, now $\nu_\theta(t)$ may be computed from Eq 4.18, by using Eq 4.19 and Eq 4.20

In [23] Kedem et al, the authors have show that the minimum of $\nu_\theta(t)$ exists, hence from the computations' results, Fig 4.1 was obtained.

The graph is similar to the one of [23] Kedem et al. It reflects that the minimum is at t = 484 Gwh Loss.

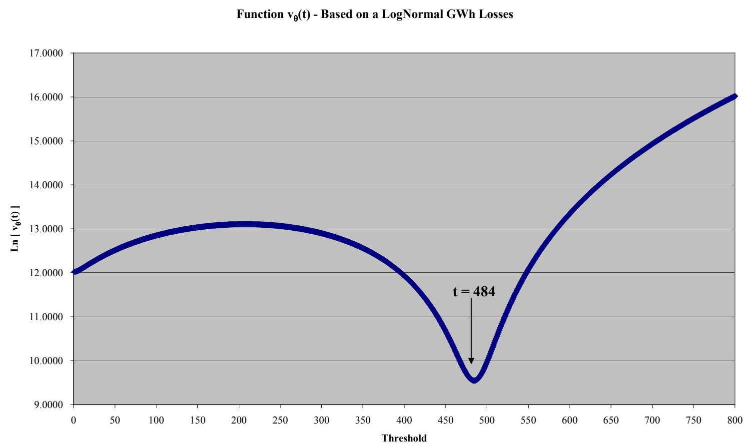



**Fig 4.1: Log of variance vs Threshold:** **min @ 484 GWh Loss**

| | |
|---|---|
| $\mu$ = | 4.6640 |
| $\sigma$ = | 1.2725 |
| **E[*X*] =** | 238.34 |
| **P[*X*>*t*] =** | 0.1163 |
| **Min Ln $\nu$t =** | 9.55 |
| **Min $\nu$t =** | 14,004 |
| $\mathbf{t_{min}} =$ | 484 |
| $\beta_\theta$**(t) =** | 2049 |

This value of 484 GWh answers the question posed in Chapter 2, just above and in relation to Eq 2.2, i.e. on how the individual GU's Losses affect the System. It seems that according to this method and results, the individual GUs' losing more than 484 GWh in a year affect the System in that year. This has a bearing on fig 3.2 of the System Yearly Maxima (Gumbel), in Chapter 3.

The GWh losses at or below 484 GWh may be considered as "background-noise" losses.

As a subjective interpretation, the analysis appears to indicate that the slope "stretches" to a value of 2049 GWh (see Eq 4.17). This value could also indicate that the tail deviates from the Log-Normal after the 9th largest observation.

### 4.2.2 Generalised Pareto Distribution (GPD)

In a similar way as in paragraph 4.2.1, let us now assume that after using an initial threshold of 484 GWh we wish to fit a distribution to the tail. We can use the method by [23] Kedem et al.

In this case then: $\theta = \begin{pmatrix} \theta_1 = \gamma \\ \theta_2 = \sigma \end{pmatrix}$ and $y = t - u > 0$ for $\gamma > 0$

.

$$F_Y(y;\boldsymbol{\theta}) = 1 - [1 + \frac{\boldsymbol{\gamma}(t-u)}{\boldsymbol{\sigma}}]^{-\frac{1}{\gamma}} \quad , y > u \ , \ \boldsymbol{\sigma} > 0$$

Since

$$P(Y \le t \mid y > u) = F_Y(y;\boldsymbol{\theta}) \qquad P(Y > t \mid y > u) = 1 - F_Y(y;\boldsymbol{\theta})$$

$$\therefore P(Y > t \mid y > u) = [1 + \frac{\boldsymbol{\gamma}(t-u)}{\boldsymbol{\sigma}}]^{-\frac{1}{\gamma}} = \mathcal{P} \qquad \text{Eq 4.21}$$

$$Ln\,\mathcal{P} = -\frac{1}{\gamma}Ln\,\{1+\frac{\gamma\,(t-u)}{\sigma}\}$$

$$\Rightarrow \frac{\partial}{\partial\mathcal{P}}Ln\,\mathcal{P} = \frac{1}{\mathcal{P}} \quad \Rightarrow \partial\mathcal{P} = \partial Ln\,\mathcal{P}.\mathcal{P}$$

$$\frac{\partial}{\partial\gamma}\mathcal{P} = \frac{\partial}{\partial\gamma}Ln\,\mathcal{P}.\mathcal{P} = \frac{\partial}{\partial\gamma}[-\frac{1}{\gamma}Ln\,\{1+\frac{\gamma\,(t-u)}{\sigma}\}]\,.\,[1+\frac{\gamma\,(t-u)}{\sigma}]^{-\frac{1}{\gamma}}$$

$$\therefore \frac{\partial}{\partial\gamma}\mathcal{P} = [1+\frac{\gamma\,(t-u)}{\sigma}]^{-\frac{1}{\gamma}}.\left[-\frac{1}{\gamma}\frac{(t-u)}{\sigma}\left[1+\frac{\gamma\,(t-u)}{\sigma}\right]^{-1}+\frac{1}{\gamma^2}Ln\,\{1+\frac{\gamma\,(t-u)}{\sigma}\}\right]$$

Eq 4.22

$$E[Y>0\,|\,y>u] = \xi = \frac{\sigma+\gamma\,u}{1-\gamma}$$ Eq 4.23

(see Table A1.1 in Appendix A)

$$\therefore \frac{\partial\xi}{\partial\gamma} = \frac{\partial}{\partial\gamma}\left(\frac{\sigma+\gamma\,u}{1-\gamma}\right) = \frac{\sigma+u}{(1-\gamma)^2}$$ Eq 4.24

By using equations Eq 4.21 to Eq 4.24 and the quotient rule of differentiation on $\beta_\theta(t) = \dfrac{E[Y>0\,|\,y>u]}{P(Y>t)\,|\,y>u)}$ (see Eq 4.17 above) to get the partial derivative $\dfrac{\partial\beta_\theta(t)}{\partial\gamma}$, we obtain:

$$\frac{\partial\beta_\theta(t)}{\partial\gamma} = \left(1+\gamma\frac{t-u}{\sigma}\right)^{\frac{1}{\gamma}}\left[\frac{\sigma+u}{(1-\gamma)^2}-\frac{\sigma+\gamma u}{1-\gamma}\left\{\frac{1}{\gamma^2}Ln\{1+\gamma\frac{t-u}{\sigma}\}-\frac{1}{\gamma}\left(\frac{t-u}{\sigma}\right)\left(1+\gamma\frac{t-u}{\sigma}\right)\right\}\right]$$

Eq 4.25

and

$$\frac{\partial \boldsymbol{\beta}_{\boldsymbol{\theta}}(t)}{\partial \boldsymbol{\sigma}} = \left(1+\boldsymbol{\gamma}\frac{t-u}{\boldsymbol{\sigma}}\right)^{\frac{1}{\gamma}} \left[(1-\boldsymbol{\gamma})^{-1} - \frac{\boldsymbol{\sigma}+\boldsymbol{\gamma}\,u}{1-\boldsymbol{\gamma}}\left\{\left(\frac{t-u}{\boldsymbol{\sigma}^2}\right)\left(1+\boldsymbol{\gamma}\frac{t-u}{\boldsymbol{\sigma}}\right)^{-1}\right\}\right]$$

Eq 4.26

The FIM for the GPD is (See Appendix C Eq 3.19)

$$[\mathrm{I}(\gamma;\sigma)]=\begin{bmatrix} \dfrac{2}{(1-\gamma)(1-2\gamma)} & -\dfrac{1}{\sigma(1-\gamma)(1-2\gamma)} \\ -\dfrac{1}{\sigma(1-\gamma)(1-2\gamma)} & \dfrac{1}{\sigma^2(1-2\gamma)} \end{bmatrix} \quad \text{Eq 4.27}$$

$$[\mathrm{I}(\sigma;\gamma)]=\begin{bmatrix} \dfrac{1}{\sigma^2(1-2\gamma)} & -\dfrac{1}{\sigma(1-\gamma)(1-2\gamma)} \\ -\dfrac{1}{\sigma(1-\gamma)(1-2\gamma)} & \dfrac{2}{(1-\gamma)(1-2\gamma)} \end{bmatrix} \quad \text{Eq 4.28}$$

The matrix $\mathbf{I}^{\mathbf{-1}}(\theta)$ derived from inverting matrix Eq 4.27 above yields:

$\mathbf{I}^{\mathbf{-1}}(\gamma;\sigma)$

$$= \begin{bmatrix} 2\,\sigma^2(1-\gamma) & -\sigma(1-\gamma) \\ -\sigma(1-\gamma) & (1-\gamma)^2 \end{bmatrix}$$

$$= \sigma^2(1-\gamma)^2 \begin{bmatrix} \dfrac{2}{(1-\gamma)} & -\dfrac{1}{\sigma(1-\gamma)} \\ -\dfrac{1}{\sigma(1-\gamma)} & \dfrac{1}{\sigma^2} \end{bmatrix}$$

Then from the Delta Method (see Appendix C & paragraph 3.2) and Eq 4.15, Eq 4.16:

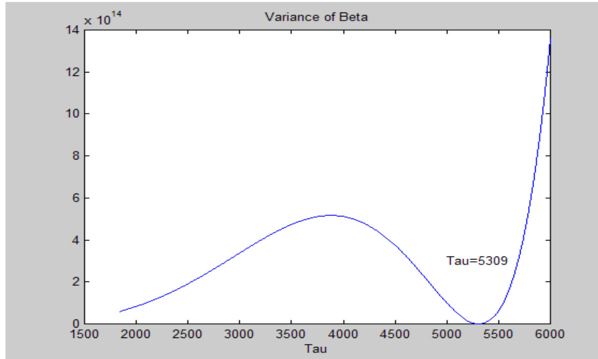
$$\nu_\theta(t) = \begin{pmatrix} \dfrac{\partial \beta}{\partial \gamma} & \dfrac{\partial \beta}{\partial \sigma} \end{pmatrix} \mathbf{I}^{-1}(\theta) \begin{pmatrix} \dfrac{\partial \beta}{\partial \gamma} \\ \dfrac{\partial \beta}{\partial \sigma} \end{pmatrix} \qquad \text{Eq 4.29}$$

Corresponding to the variance of ***t***.

Care ought to be taken when choosing the FIM on the sequence of the parameters, otherwise the threshold associated with the minimum variance level, might be underestimated. For instance if using Eq 4.28 the equation from the Delta Method becomes

$$\nu_\theta(t) = \begin{pmatrix} \dfrac{\partial \beta}{\partial \sigma} & \dfrac{\partial \beta}{\partial \gamma} \end{pmatrix} \mathbf{I}^{-1}(\theta) \begin{pmatrix} \dfrac{\partial \beta}{\partial \sigma} \\ \dfrac{\partial \beta}{\partial \gamma} \end{pmatrix} \qquad \text{Eq 4.30}$$

Now if we consider taking the parameters for the GPD given in Table 3.6 (Chapter 3) and utilise the coding given in Fig 4.2 below, we obtain the graph shown in the same Fig 4.2; the results produce a threshold of 5309 GWh close to the maximum absolute expected loss of a 669 MW GU (5876.5 GWh, see p. 30).




```
%gpdtail2.m Mean of the GPD related to the tail probability
sig=908.5;xi=.2802;
T=[];
for u=1841
x=u:1:6000;
S=(1+xi*(x-u)/sig).^(-1/xi);
I=[1/sig^2/(1+2*xi) -1/sig/(1+2*xi)/(1+xi)
    -1/sig/(1+2*xi)/(1+xi) 2/(1+2*xi)/(1+xi)];
V=[];
for tau=u:1:6000
w=(sig+xi*u)/(1-xi);
v=(1+xi*(tau-u)/sig)^(-1/xi);
dwdg=((1-xi)*u+sig+xi*u)/(1-xi)^2;
dvdg=(1+xi*(tau-u)/sig)^(-1/xi)*(log(1+xi*(tau-
u)/sig)/xi^2-(1+xi*(tau-u)/sig)^(-1)*(tau-u)/sig/xi);
dwds=1/(1-xi);
dvds=(1+xi*(tau-u)/sig)^(-1-1/xi)*(tau-u)/sig^2;
c1=(v*dwdg-w*dvdg)/v^2;
c2=(v*dwds-w*dvds)/v^2;
varb=[c1 c2]*inv(I)*[c1 c2]';
V=[V varb];
end
tau=u:1:6000;
[d1 d2]=min(V);
P=(1+(tau(d2)-u)*xi/sig)^(-1/xi);
mef=(sig+xi*u)/(1-xi);
beta=mef/P;
T=[T; u tau(d2) beta V(d2)];
end
plot(tau,V)
title('Variance of Beta')
text(5000,3*10^14,['Tau=',num2str(T(:,2))])
xlabel('Tau'); text
```

**Fig 4.2 Graph and Coding for the GPD–Min Variance $\beta$**

The fit of the tail of a distribution onto a GPD depends on selecting a threshold ***t***. A simple method of selecting the threshold is by inspecting the Pareto quantile plot (Fig 3.7) where the graph tends to be linear. A possible candidate seems to be $e^6 = 403.43$. The threshold from the lognormal fit in paragraph 4.2.1 gave a threshold of ***t*** = 484 where the variance of the estimate of the regression of the mean on the tail probability is at minimum. This is close to visual selected threshold (6.1821 vs 6).

When dealing with the tail of the distribution of a r.v. behaviour over a threshold, the GPD is usually the model of choice. The minimum variance value, derived from the GPD using [23] Kedem methodology, namely ***t*** = 2126 is chosen with $\gamma = 0.2802$ and $\sigma = 512.3$

The joint posterior density of the two parameters $\gamma$ and $\sigma$ is given by:

$$\pi(\gamma,\sigma \mid \text{data}) \propto \text{ Likelihood x prior}$$

where the Likelihood and MDI prior are respectively given by [2] Beirlant et al p 447:

$$L(\gamma,\sigma \mid \text{data}) = \prod_{i=1}^{n} \left(\frac{1}{\sigma}\right)^{n} \left(1 + \frac{\gamma\, y_i}{\sigma}\right)^{-\frac{1}{\gamma}-1} \text{ and}$$

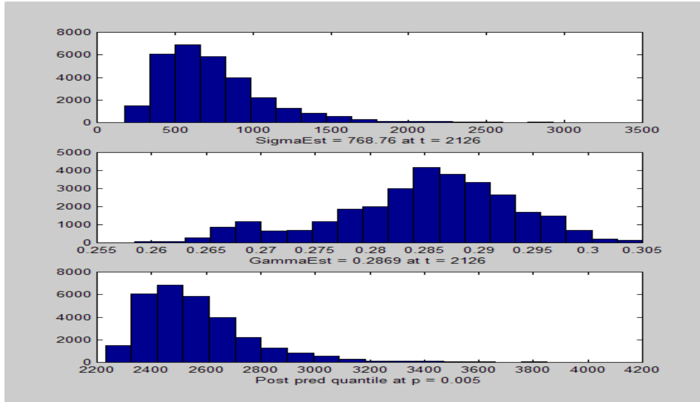
$$p^*(\gamma;\sigma) = \frac{1}{\sigma} e^{-\gamma}$$

A subjective prior can be elicitated from experience, but we will stick to the objective prior. Other priors such as Jeffreys prior (see Eq C3.20 in Appendix C) is a popular prior, but it is restricted to $\gamma < 0.5$ and tends to pull the estimate of $\gamma$ towards zero, hence the choice of the MDI prior.

Starting the Markov process with initial values $\gamma = 0.2802$ and $\sigma = 908.5$, the process was run for 30000 iterations. The convergence was found to be good. The final posterior distributions are given in Fig 4.3 below, which includes the code. The estimates from the simulations are obtained as $\hat{\gamma} = 0.2869$ $\hat{\sigma} = 768.76$.

The 95% hpd region for the GWh Losses based on a GPD with the parameters estimates above and on a threshold of 2126 GWh is: [2343.3 ; 3090.5]

The methodology used and the coding supplied in Fig 4.3 provide the technique in the paragraph following, i.e. the prediction of High Quantiles through Markov Chain Monte Carlo (MCMC).




```
%gpdvince1.m Fit a GPD to GU energy loss data GHWloss1.m

GWHloss1
t=2126;r=find(L2>t);
n=length(r);ntot= length(L2);
x=sort(L2(r))-t;alpha=512.3;xi=.2802;
I=0;X=[];
for i=1:30000
  lpost=-(1+1/xi)*sum(log(1+xi*x/alpha))-(n+1)*log(alpha)-
xi;
alphas=exp(.04*randn(1,1)+log(alpha));xis=.0004*randn(1,1)+
xi;
  lposts=-(1+1/xis)*sum(log(1+xis*x/alphas))-
(n+1)*log(alphas)-xis;
   if rand(1,1)<exp(lposts-lpost)
   alpha=alphas;xi=xis;
   I=I+1;
   end
   X=[X;[alpha xi]];
end
alpha=mean(X(5000:30000,1))
xi=mean(X(5000:30000,2));n;ntot
p=0.005;m=n/(ntot.*p)
subplot(3,1,1); hist(X(:,1),20)
xlabel('Sigma');subplot(3,1,2)
hist(X(:,2),20);xlabel('Gamma')
U=X(:,1).*(m.^X(:,2)-1)./X(:,2)+t;
subplot(3,1,3)
hist(U,20);xlabel('Post pred quantile at p=.005')
G=zeros(length(x),1);subplot(111)
```

**Fig 4.3 Graphs and Coding for the GPD – Posterior**

# 4.3 Predicting High Quantiles through MCMC

The prediction of high quantiles forms a very necessary part of an extreme value analysis.
Assume that one would like to make inferences about future values of high quantiles. The predictive distribution is given below (see paragraph 4.1 above)

$$f_X(x_{n+1} \mid X) = \int_{\Omega} f_X(x_{n+1} \mid \boldsymbol{\gamma};\boldsymbol{\sigma})\ \boldsymbol{\pi}(\boldsymbol{\gamma};\boldsymbol{\sigma} \mid X)\ d(\boldsymbol{\gamma};\boldsymbol{\sigma}) \qquad \text{Eq 4.31}$$

This shows that the predictive density consists of the integration of the likelihood (of a single x, i.e. $x_{n+1}$, multiplied by the posterior.
From paragraph 4.2 above,

$$L(\gamma,\sigma \mid \text{data}) = \prod_{i=1}^{n} \left(\frac{1}{\boldsymbol{\sigma}}\right)^{n} \left(1+\frac{\boldsymbol{\gamma}\, y_i}{\boldsymbol{\sigma}}\right)^{-\frac{1}{\gamma}-1} \text{ and}$$

$$p^*(\boldsymbol{\gamma};\boldsymbol{\sigma}) = \frac{1}{\boldsymbol{\sigma}} e^{-\gamma}$$

Therefore using the two equations above, the following posterior is obtained:

$$\boldsymbol{\pi}(\boldsymbol{\gamma};\boldsymbol{\sigma} \mid X) = \prod_{i=1}^{n} \left(\frac{1}{\boldsymbol{\sigma}}\right)^{n} \left(1+\frac{\boldsymbol{\gamma}\, y_i}{\boldsymbol{\sigma}}\right)^{-\frac{1}{\gamma}-1} \ .\ \frac{1}{\boldsymbol{\sigma}} e^{-\gamma} \text{ , hence,}$$

taking logs:

$$Ln\left[\boldsymbol{\pi}(\boldsymbol{\gamma};\boldsymbol{\sigma} \mid X)\right] = -\,n\ Ln(\boldsymbol{\sigma})(-1)(\frac{1}{\boldsymbol{\gamma}}+1)\sum_{i=1}^{n} Ln\left[1+\frac{\boldsymbol{\gamma}\, y_i}{\boldsymbol{\sigma}}\right] - Ln(\boldsymbol{\sigma}) - \boldsymbol{\gamma}$$

and re-arranging:

$$Ln\left[\boldsymbol{\pi}(\gamma;\boldsymbol{\sigma} \mid X)\right] = (n+1)\, Ln(\boldsymbol{\sigma})\ (\frac{1}{\gamma}+1) \sum_{i=1}^{n} Ln\left[1+\frac{\gamma\, y_i}{\boldsymbol{\sigma}}\right]\ -\gamma \qquad \text{Eq 4.32}$$

This equation of the log-posterior distribution is equivalent to the line "lpost" in the code in Fig 4.3 above.

The equation: $\hat{x}_q = t + \frac{\hat{\boldsymbol{\sigma}}}{\hat{\boldsymbol{\gamma}}}\left[\left(\frac{n\,(1-q)}{N_t}\right)^{-\hat{\gamma}} - 1\right]$ is re-arranged

as $$\hat{x}_q = t + \frac{\hat{\boldsymbol{\sigma}}}{\hat{\boldsymbol{\gamma}}}\left[\left(\frac{N_t}{n\,(1-q)}\right)^{\hat{\gamma}} - 1\right]$$

Whereby, in the code of Fig 4.3, $"m" = \frac{N_t}{n\,(1-q)}$

and "*n*" $= N_t$ and, still in the same code, <u>"U" represents the predictive distribution</u> derived from a whole set of $\hat{x}_q$'s which are in turn derived from a whole set of $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\gamma}}$ being simulated ("alphas" and "xis" respectively in the code)

Another way that this may also get a point-estimate through the formulae (see [2] Beirlant et al) is to estimate the percentiles, by letting $z_p$ = upper $p^{\text{th}}$ percentile of $\text{GPD}_{\gamma,\sigma}(x\text{-}\boldsymbol{t})$; then the estimate of $z_p$ is:

$$\hat{x}_p = t + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{P(X>t)}{p}\right)^{\hat{\gamma}} - 1\right] \equiv$$

$$\hat{x}_q = t + \frac{\hat{\boldsymbol{\sigma}}}{\hat{\boldsymbol{\gamma}}}\left[\left(\frac{n\,(1-q)}{N_t}\right)^{-\hat{\gamma}} - 1\right]$$

where the probability $P(X > \boldsymbol{t})$ is greater than a value $p$ and $P(X > \boldsymbol{t})$ is estimated by the sample proportion exceeding $\boldsymbol{t}$.

That is $P(X > \boldsymbol{t}) = N_t / n$ and $p = (1 - q)$ and inverting by making the exponent $\gamma$ negative.
Let us say for instance that we would like to position our threshold at the 99.05th percentile, this means that the sample proportion $[(n - N_t) / n] = 0.9905$, since $n = 928$,
$N_t = 928.(1\text{-}0.9905) = 8.8$, i.e. we ought to have approx. 9 observations above our threshold.
By choosing 2126 GWh, there is consistency with the results in paragraph 4.2.2 as that is the point of minimum variance and by using the parameters $\hat{\gamma} = 0.2869 \quad \hat{\sigma} = 768.76$
In the above equation, then we get:

$$\hat{x}_{0.9905} = 2126 + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{928\,(1-0.9905)}{9}\right)^{-\hat{\gamma}} - 1\right] = 2141.9 \text{ GWh}$$

Let us now operate more into the tail, say at the 99.75th percentile, therefore expecting 2 or 3 observations at this level.

$$\hat{x}_{0.9975} = 2126 + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{928\,(1-0.9975)}{3}\right)^{-\hat{\gamma}} - 1\right] = 2331.1 \text{ GWh}$$

From the simulations above on fitting the GPD (Fig 4.3), the predicted simulated distributions are shown in the graphs. There one may remark that the computations at $\hat{\sigma} = 768.76$ (the mode, in Fig 4.3) might have the tendency to underestimate the percentiles at the upper end of the tail (compare 2331.1 with the empirical 3724.6 at the 0.9970 percentile), therefore the value of 908.5 for sigma seems to be more appropriate. More experimenting on the sensitivity of the sigma (see coding in Fig 4.3 the value of 0.04 in “alphas”) would be necessary and then benchmark to the Table 3.6 (derived from Eq 3.12). More research in this area is necessary.

## 4.4 Relationship between MEF and the Tail of the GPD

In paragraph 4.2.2, the relationship between the Mean Excess Function and the tail probability was shown in detail following the works of [23] Kedem et al.
There it was shown that there is a relationship between the Mean Excess Function (MEF)
$e_t = E\,[(X - \boldsymbol{t}) \mid X > \boldsymbol{t}]$ and the tail probability $p_t = P\,(X > \tau \mid X > \boldsymbol{t})$ for some threshold $\tau$ conditional that we only consider excess losses above $\boldsymbol{t}$ (= 484 GWh in our case).
The relationship that was considered was

$$e_t = E[X - t \mid X > t] = \boldsymbol{\beta}_{\hat{\theta}}(t)\,.\,P[X > \boldsymbol{\tau}) \mid X > t]$$

$$e_t = \boldsymbol{\beta}_{\hat{\theta}}(t)\,.\,p_t$$

As shown in Table A1.1 in Appendix A, that the mean excess function for the GPD is given by $e_t = \dfrac{\sigma + \gamma\, t}{1-\gamma}$, following the arguments of [23] Kedem et al, by means of partial derivatives with respect to $\gamma$ and $\sigma$ respectively and using the Fisher Information Matrix and its inverse, as well as invoking the Delta Method, we obtained the variance of the maximum likelihood estimate of $\beta_{\hat{\theta}}(t)$ and resulting in the finding of $\tau$ that minimizes this variance. Experimenting, we found initially that the threshold where the tail of the GPD starts effecting the mean excess function, is $\tau = 449$ (while for the Log Normal it was 484). The corresponding estimate for $\beta_{\hat{\theta}}(t)$ being 1253.5.

Testing further, choosing a threshold of 550, we got estimates of $\hat{\gamma} = 0.5102$ $\hat{\sigma} = 272.75$; the Q-Q plot is shown in Fig 4.4 below. The fit of the extreme values appears to be quite good. As a subjective interpretation, the value of 550 GWh seems to coincide with the suppression of the outlying GWh mentioned in Chapter 3 Fig 3.9 (see paragraph above figure) as the background loss within an extreme value that "misbehaved".

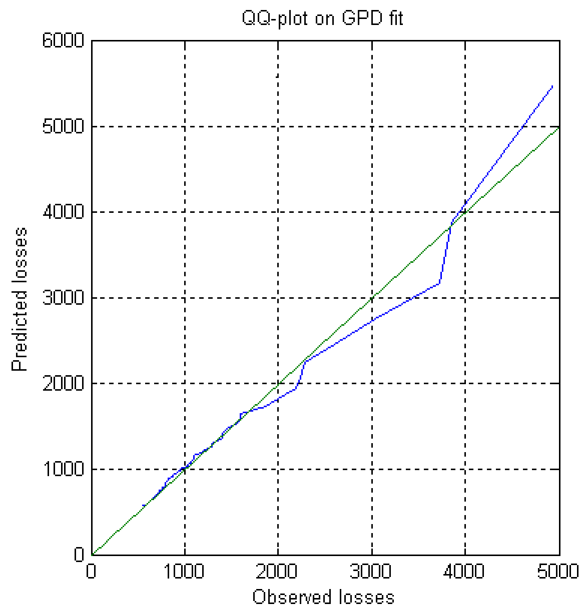



**Fig 4.4 Q-Q Plot GPD** $\hat{\gamma} = 0.5102$ $\hat{\sigma} = 272.75$ and $t = 550$

Extra assessments into the relationship, yielded a threshold at 850 GWh at which the GPD tail shows the best relationship with the mean excess over the threshold; this estimates the $\beta_{\hat{\theta}}(t)$ at 5310.5 GWh

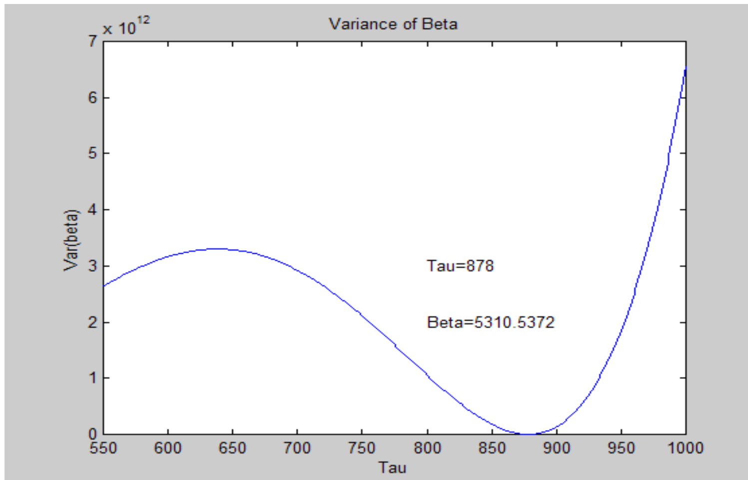



**Fig 4.5 Minimum Variance** found at ***t = 878***

## 4.5 The Dirichlet Process

As seen in Fig 3.10 of Chapter 3, in the assessment at when the Mean Excess Function becomes linear becomes difficult. [10] De Waal and [11] De Waal et al worked on that challenge and utilised the method of the Negative Differential Entropy (NDE). There the choice of the threshold (using the POT method) is applied to model extremes through the NDE of the Dirichlet process using a Bayesian approach. This methodology also can be used as a goodness of fit technique.

Let $x_1,..,x_k$ be $k$ ordered observations exceeding the threshold $\boldsymbol{t}$ of a random sample of $n$ observations from an unknown distribution with a distribution function $F_X(x \mid t)$. The aim is to estimate $F_X(x \mid t)$ if we assume a proposed distribution $\mathrm{GPD}_{\gamma,\sigma}(x\text{-}\boldsymbol{t})$. The $F_X(x \mid t)$ then is modelled as a Dirichlet process with the GPD as its parameters. Then from [20] Honkela, [11] De Waal et al, gets (see Eq C2.14 in Appendix C):

NDE = E[Log $p(\boldsymbol{p})$]

E[Log $p(\boldsymbol{p})$]= $J_k$

$$J_k = -Log\, Z(\boldsymbol{u}) + \sum_{i=1}^{n} (u_i - 1)\ [\Psi_0(u_i) - \Psi_0(u_0)] \qquad \text{Eq 4.33}$$

Subsequently, as proposed by [11] De Waal et al, one needs to select $k$ (i.e. the threshold) such that the NDE, in Eq 4.33 is at minimum, conversely this means that the information is at maximum (see the principles of entropy in Appendix C paragraphs 2.8 and 2.9).

A set of results is given in Fig 4.6 here below, together with the algorithm used.

As explained above, this kind of method provides the means to select, at the minimum NDE, the threshold with the GPD parameters that were estimated. By substituting the GPD parameters with new estimates, new levels of NDE are obtained; higher NDE levels would indicate a poorer GPD fit. Not sufficient credit is given to this method in this thesis, suffice it to say that it was experimented with and at prima-facie it looks promising.
A more detailed rendition is given in Appendix C paragraphs 2.8 and 2.9

The results obtained from the graph in Fig 4.6 indicated that the most contributing values to fit of the GPD tail were the 6 largest extremes.

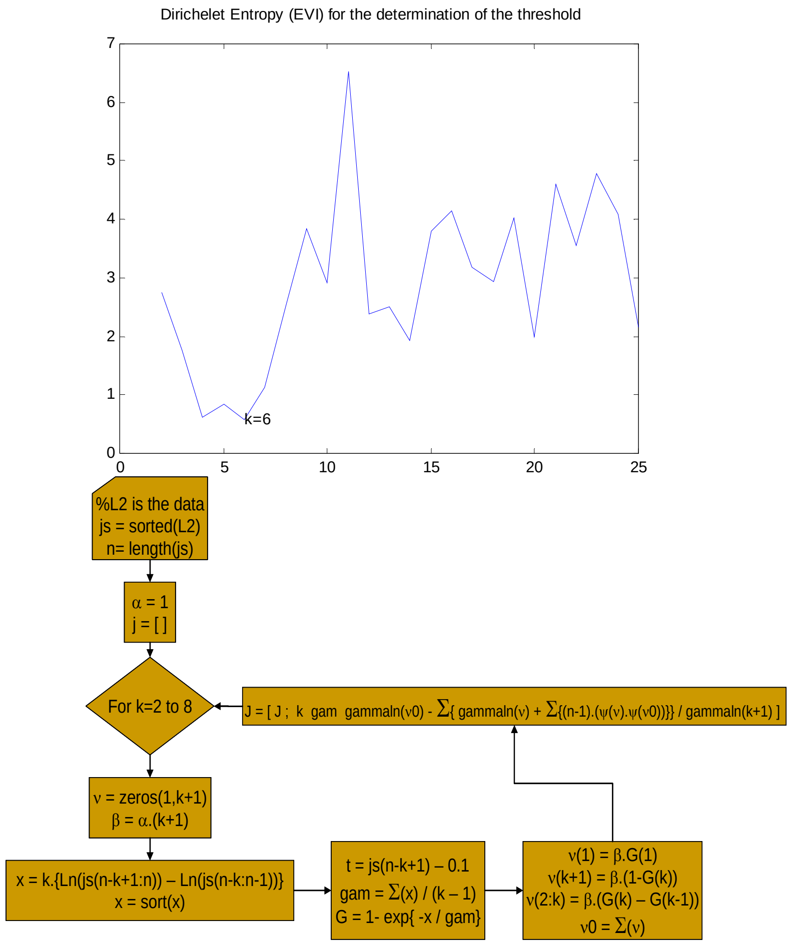



**Fig 4.6 Graphic Result & Matlab Algorithm of NDE**

# CHAPTER 5

# Conclusions

## 5.1 Categorisation

The means of using Cluster Analysis provided an initial method to classify large losses incurred by the GUs. In particular the Two-Way Joining method (results shown in Fig 2.3) provided initial "limits" with which these categories could be formed. These were practically applied in our business as what we termed "layers" for the GWh Losses.

The yearly total System GWh Loss indicated that a Loglogistic distribution fit is appropriate with a threshold of 1763 GWh.

This was shown to be useful but not sufficient; hence the usage of EVT became necessary for predicting extreme values in a more detailed manner.

## 5.2 Fitting of the Tail

The semi-parametric (Eq 3.8 and Figures 3.13 to 3.15) and full-parametric POT methodology proved very useful in the provision for the GPD fit of the tail of the GWh Losses. Figures 3.13 to 3.15, with particular emphasis on the latter provide the highlight of Chapter 3 with an interesting methodology to estimate the initial parameters (the 3-LM method). Chapter 3 concluded with the final adopted model for the GWh Loss quantiles being:

$$\hat{x}_q = 1841.8 + \frac{908.5}{0.2802}\left[\left(\frac{928(1-q)}{11}\right)^{-0.2802} - 1\right]$$

for GWh Losses above 765.8 GWh.

Illustrated by Fig 3.20 and tabulated in Table 3.6

## 5.3 Bayesian Methods

Classical methods allowed us to estimate the GPD parameters (by 3LM methodology) and infer on the distribution GPD ($y \mid \sigma,\gamma,\boldsymbol{t}$) , however as pointed out in Chapter 4, this did not make allowance for the errors incurred when estimating $\sigma,\gamma,$ inducing a false expectation of the predictive distribution, as the classical methodology did not cater for the fact that a future value is a stochastic event in itself (see Eq 4.31).
The Dirichlet methodology provides an interesting angle to the Bayesian philosophy in analysing the problem of the GWh Losses.

## 5.4 General

From an engineering perspective other interesting findings was that further research in the extremes with the introduction of covariates such as the Load Factor (LF) is a necessity. The assumptions in this thesis are that the LF is constant and the Installed Capacity (IC) does not change, which implies that the forced outages, UCLF (the GWh Losses) are in stationary mode of “bathtub hazard function”; what happens if, as is expected in the near future, these are not constant? For instance (see pp 89-90 for more details):

1] IC changes, i.e. New Capacity (large, significant generating units) is installed; this introduces high "infant mortality" rate (i.e. higher than expected UCLF) which "lifts" the function LF vs UCLF, therefore higher than expected losses $\Rightarrow$ a different $\gamma$, possibly different MDI? This increase in capacity is expected in 2011.

2] the LF changes according to the increase in demand and according to the LF vs UCLF function, the GWh losses (UCLF) would increase.

Care ought to be taken of the changes in $\gamma$ [i.e. in the behaviour of the losses] due to time – but in our case it's actually due to changes in LF that in turn, changes with time.

## Closing Comment

The bullets given in page vii) at the beginning of the thesis were fully investigated and answered. Extreme Value Theory (EVT) methods did quantify the meaning of "Large" (or "excessive") GWh Losses. Within the preamble of EVT and using Cluster Analysis, it was possible to categorise the losses into "layers". A scientific model was developed (see 5.2 above) which needs to be periodically validated and maintained. Latest Statistical techniques (such as the Dirichlet one) were researched and applied with promising results. And in conclusion, as having provided the answers above, it is felt that EVT resolved our challenge. In addition, it opened pockets (as pointed in various parts of the thesis) of new areas of research.

# APPENDIX A
## Notation and Classic EVT Workings

## 1 Notation

Most of the work in this part of the Appendix is referenced to [2] and [3] Beirlant, [13] Embrechts and [25] McNeil. The more practical and applied workings of EVT are particularly referenced to [2] and [3] Beirlant while the more rigorous theoretical ones are referred to [13] Embrechts.

Let X be a continuous random variable (rv), then $F_X(x)$ is its CDF, i.e. $F_X(x) = P(X \leq x)$
The tail, or survivor function is denoted by:
$1 - F_X(x) = P(X > x)$ and where it exists, the probability density function (pdf) is $f_X(x)$ and satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(x).dx \qquad \text{A1.0}$$

For $q \in (0\,;1)$, the $q$-th quantile of the distribution of $X$ is:

$$F^{-1}(q) = \inf\{x \in \mathfrak{R}: F_X(x) \geq q\} \qquad \text{A1.1}$$

The expectation of X is,

$$\mathrm{E}[X] = \int_{-\infty}^{+\infty} x.f_X(x).dx \qquad \text{A1.2}$$

The variance of X is,

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] \qquad \text{A1.3}$$

The covariance of $X$ and $Y$ is,
Cov $[X,Y] = E[(X-E[X]) \cdot (Y-E[Y])]$ A1.4

The conditional expectation of $Y$, given $X$ is,

$$E[Y|X=x] = \int_{-\infty}^{+\infty} y.f_{Y|X}(y|x).dy \qquad \text{A1.5}$$

## 2 Classical EVT

In studying the maxima of rv's (or minima), it is important to find the limiting distribution of these maxima (or minima). It is hence with this objective in mind that statisticians of note have researched the possible distributions of the maxima. The objective being twofold [2] Beirlant:

1] find all possible (non-degenerate) distributions G that can appear as a limit. This is known as the ***Extremal Limit*** problem.

2] characterize the distributions $F$ for which there exist sequences $\{a_n\ ;\ n \geq 1\}$ and $\{b_n\ ;\ n \geq 1\}$ such that

$P((M_n - b_n) / a_n \leq x) \rightarrow G_X(x)$ holds for any such specific limiting distribution [2] Beirlant, where

$M_n = max(X_1, X_2, \ldots , X_n)$ be the maximum of the sample of n. In this thesis, $X$ could be associated to the GWh Losses out of a sample of $n$ GU's in a particular year.

$\therefore P(M_n \leq x) = P(X_1 \leq x , X_2 \leq x , \ldots , X_n \leq x) = F^n(x)$

1] Has been resolved in the works of [15] Fisher et al known as the Fisher-Tippet Theorem

2] Has been dealt with by [13] Embrechts and [2] Beirlant as follows

One can now find sequences of $\mathfrak{R}$, $a_n > 0$ and $b_n > 0$ such that $(M_n - b_n) / a_n$ , i.e. the sequence of normalised maxima, converges in distribution.

If $G_X(x)$ is a non-degenerate distribution function of $X$ , which will specify a unique type of distribution, and [13] Embrechts

$P((M_n - b_n) / a_n \leq x)$
$= F^n(a_n x + b_n) \rightarrow G_X(x),$ as $n \rightarrow \infty$ A2.0
then we say that:

**F is in the maximum domain of attraction (MDA) of G, i.e.**
**F ∈ MDA (G)**

Two rv's U and V, are of the same type (i.e. equality in type) iff $U \overset{d}{=} a.V + b$

re-scaling re-location
This means same type of distribution under same scaling and location

In terms of their CDF's, $F_U$ and $F_V$ , this means:
$F_U(x) = F_V((x - b)/a)$ A2.1

This means that rv's of the same type have the same CDF's, up to possible changes of scale and location [25] McNeil.
It also means that if $G$ is the possible limiting distribution function for the sequence $(M_n - b_n) / a_n$ , then the set of such distribution functions exhibiting attraction to their limits is called the domain of attraction of $G$ [2] Beirlant.

## 2.1 The Generalized Extreme Value Distribution and Index (GEV & EVI)

The GEV of a rv X is given by [2] Beirlant:

$$G_\gamma(x) = \begin{cases} exp\ \{\ -(1+\gamma.x)^{-1/\gamma}\ \}\ \ , \text{if } \gamma \neq 0 \\ \\ exp\ \{\ -exp\{-x\}\ \}\ \ , \text{if } \gamma = 0 \end{cases} \qquad \gamma \in \Re \quad \text{A2.2}$$

where $1 + \gamma.x > 0$ and $\gamma$ is the shape parameter and is also termed the ***Extreme Value Index*** (EVI). The EVI forms an integral and fundamental part of EVT analyses.

In A2.2, $x$ can be replaced by $x' = (x - \mu) / \sigma$ , in which case the GEV is termed the **Standard Generalized Extreme Value** Distribution (SGEV) with the following conditions:

$$x > \mu - (\sigma / \gamma) \text{ for } \gamma > 0$$
$$x < \mu - (\sigma / \gamma) \text{ for } \gamma < 0$$
$$x \in \Re \qquad \text{for } \gamma = 0$$

Here $\gamma$ is called the ***shape*** parameter, $\mu$ the ***location*** parameter and $\sigma$ the ***scale*** parameter

Since this parameterization is continuous in $\gamma$, three extremal types of distribution can be derived (by using A2.1, [15] Fisher, [13] Embrechts and [2] Beirlant using the Helly-Bray Theorem ):

$\gamma > 0 \Rightarrow$ Fréchet: $G_\gamma((x-1)/\gamma) = \Phi_{1/\gamma}(x)$
$\gamma < 0 \Rightarrow$ Weibull: $G_\gamma(-(x+1)/\gamma) = \Psi_{-1/\gamma}(x)$
$\gamma = 0 \Rightarrow$ Gumbel: $G_0(x) = \Lambda(x)$

By substitution in A2.2, the following CDF's are obtained:

Fréchet: $\gamma > 0$
$$\Phi_\alpha(x) = \begin{cases} 0 & x \le 0 \\ exp\{-x^{-\alpha}\} & x > 0 \end{cases} \qquad \alpha > 0 \qquad \text{A2.3}$$

Weibull: $\gamma < 0$
$$\Psi_\alpha(x) = \begin{cases} exp\{-(-x)^{\alpha}\} & x \le 0 \\ 1 & x > 0 \end{cases} \qquad \alpha > 0 \qquad \text{A2.4}$$

Gumbel: $\gamma = 0$
$$\Lambda(x) = exp\{-e^{-x}\} \qquad x \in \Re \qquad \text{A2.5}$$

By differentiating the CDF (using A1.0) the following pdf's can be obtained:

Fréchet:
$$\phi_\alpha(x) = \begin{cases} 0 & x \le 0 \\ \alpha x^{-\alpha-1}.exp\{-x^{-\alpha}\} & x > 0 \end{cases} \qquad \alpha > 0 \qquad \text{A2.6}$$

Depicted below in Fig A1.1

Weibull: $\psi_\alpha(x) = \begin{cases} \alpha(-x)^{\alpha-1}.exp\{-(-x)^{\alpha}\} & x \leq 0 \\ 0 & x > 0 \end{cases}$ $\alpha > 0$ A2.7

Depicted below in Fig A1.2

Gumbel: $\lambda(x) = e^{-x}.exp\{-e^{-x}\}$ $x \in \Re$ A2.8

Depicted below in Fig A1.3

In the Fréchet domain, the:

**1]** Burr distribution is characterized by:
for $x > 0;\ \eta,\ \tau,\ \lambda > 0$:
$F_X(x) = 1 - [\ \eta / (\eta + x^{\tau})\ ]^{\lambda}$
$f_X(x) = -\lambda\tau\,[x^{\tau-1}]\ [\ \eta / (\eta + x^{\tau})\ ]^{\lambda-1}$
**EVI:** $\boldsymbol{1/(\lambda\tau)}$

**2]** Generalised Pareto (GP) distribution is characterized by:
for $x > 0;\ \sigma,\ \gamma \neq 0$:
$F_X(x) = 1 - [\ 1 + (\gamma x / \sigma)\ ]^{-1/\gamma}$
$f_X(x) = (1 / \sigma)\ [\ 1 + (\gamma x / \sigma)\ ]^{-(1/\gamma)-1}$
**EVI:** $\boldsymbol{\gamma}$

**3]** Pareto (Pa) distribution is characterized by:
for $x > 1;\ \gamma > 0$ for the mean to exist and $x > 2$ for the variance to exist:
$F_X(x) = 1 - [x]^{-1/\gamma}$
$f_X(x) = (1 / \gamma)\ [x]^{-(1/\gamma)-1}$
**EVI:** $\boldsymbol{\gamma}$

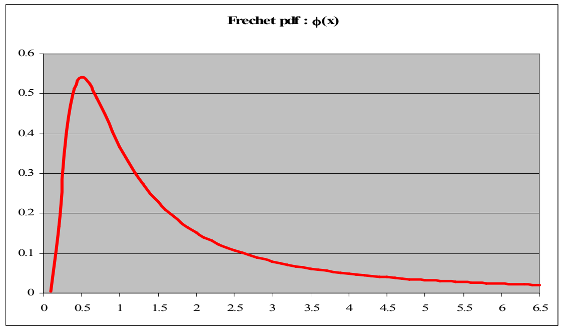



**Fig A1.1: Fréchet's pdf** (GEV with $\gamma = 1$)

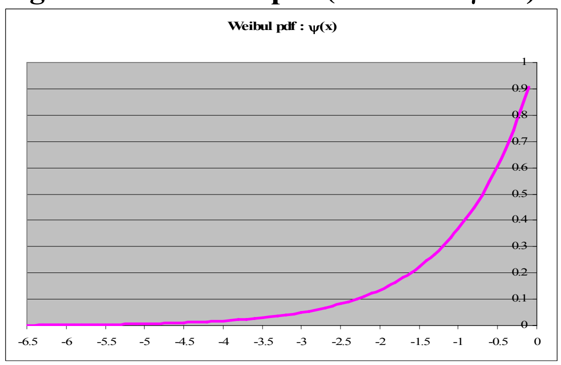



**Fig A1.2: pdf of Weibull** (GEV with $\gamma = - 1$)

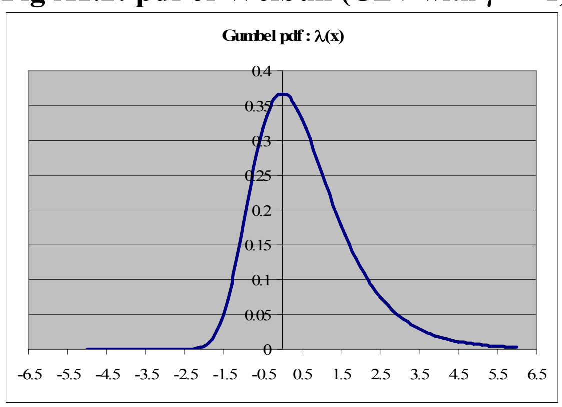



**Fig A1.3: pdf of Gumbel** (GEV with $\gamma = 0$)

**2.2 The Fréchet-Pareto Case: $\gamma > 0$** [2] Beirlant

In this part, attention is focused on this particular domain of the GEV distribution. In some publications this type is also called the ***EVD type II***. Subclasses of this type are called the ***Hall Class of distributions***.
Since in this kind of domain, the value of $\gamma$ is not known, [2] Beirlant recommends that from a statistical perspective, the analyst ought to work with the Quantile functions to estimate the $\gamma$.
When looking at this domain, it is important to define one of the most important underlying distributions, the Generalised Pareto distribution or GPD.

**The Generalised Pareto (GPD) distribution** [13] Embrechts
There are special circumstances whereby the maxima derived from a series that behaves as a GPD follow a GEV of the form:

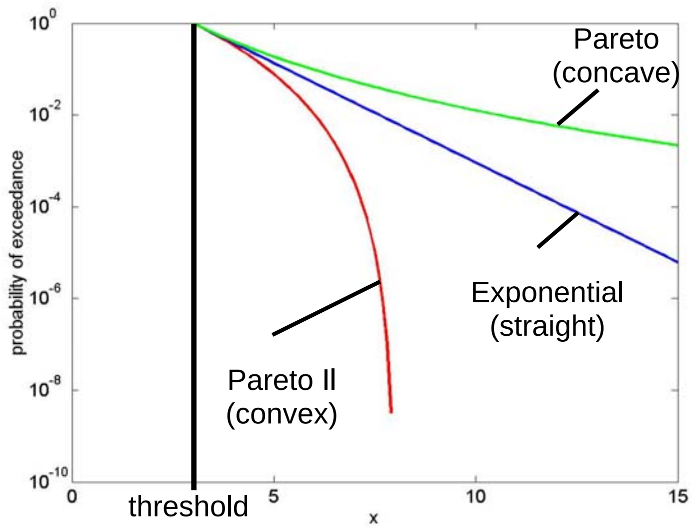
$$GPD_{\gamma,\sigma}(x) = \begin{cases} 1-(1+\gamma.x/\sigma)^{-1/\gamma}, \text{ if } \gamma \neq 0 \\ \qquad\qquad x \in (0,\infty), \gamma \in \Re \\ 1-\exp\{-x/\sigma\}, \text{ if } \gamma = 0 \end{cases} \quad \text{A2.9}$$

From above, if $\gamma = 0$ then it's Exponential, however, when $\gamma < 0$ then
$GPD_{\gamma,\sigma}(x) = 1-(1+\gamma.x/\sigma)^{-1/\gamma}$ and $x \in (0,-\sigma/\gamma)$ and the distribution becomes known as a Pareto II type

***Usually the GPD*** ([31] Pickands) ***is used with a given*** threshold ***t***; ***in that case***, $x$ in A2.9 ***is replaced by*** $(x - t)$,

***whereby*** $(x - \boldsymbol{t}) \geq 0$ for $\gamma \geq 0$ and $0 \leq (x - \boldsymbol{t}) \leq -\sigma/\gamma$ for $\gamma < 0$.
***Refer to the POT method below for more information on the*** threshold $\boldsymbol{t}$.

Under certain conditions the GP becomes a Strict Pareto (SP) with the following CDF:
$\mathrm{SP}_{1/\gamma}(x) = 1 - x^{-1/\gamma}$ , if $x > 1$, $\gamma > 0$




*from: van Gelder PHAJM*

## 2.3 Methods of Block Maxima (BM) and Peak-Over-Threshold (POT)

In the EDA phase of the analytical process there are two classical groups in dividing the data. Both groups divide the data into consecutive blocks, but:

i] the one that focuses on the series of Maxima (or Minima) in the blocks is termed **Block Maxima** (Minima) (BM), and

ii] the one that focuses on the series of events exceeding a certain High (or Low) Threshold is termed the **Peak-Over-Threshold** (POT) or also known as the ***Probability View***.

**It is also proposed in this thesis that the Block Minima is referred to as Bm and the Threshold for the Minima group be referred to as the Trough-Under-Threshold (TUT) method; in this case it is likely that, more often than not, the probability density function is bounded (e.g. at zero) and therefore more research would be needed in this area.**

In both these groups, the series are modeled separately from the rest of the observations, [2] Beirlant, [13] Embrechts, [25] McNeil and parametric estimation for the quantiles is used.

These methods deal with the aspect of tail estimations. It encompasses all values of the EVI ($\gamma$). While the BM method is inspired by the limiting behaviour of the normalised maximum of a rv, the POT method considers the conditional distribution of the excesses over a relatively high threshold [2] Beirlant.

***A third approach (to the BM and POT) called the Quantile method may also be used*** [2] Beirlant.

As pointed out in [2] Beirlant, the utilization of the set of maxima is a weakness in the sense that not the whole data set is utilized. This is compounded in determining the block-size especially in the case of Time-Series. To deal with the BM problem, POT was one of the methods developed to deal with it.
When dealing with Time-Series and one structures the series into a time dependent model (such as AR or GARCH dealing with Returns or squared-Returns (catering for the conditional volatility) respectively, say [see Appendix B, for more information on Returns]), then the Residuals can be analysed using EVT.

**Tail estimation of parameters**
Out of the numerous methods in estimating the GEV families parameters, the most important ones are reflected below, [13] Embrechts, [2] Beirlant, [21] Hosking:

**I] Maximum Likelihood Estimation (MLE)**
The log-likelihood of a sample $Y_i$ from a GEV, is given by:
$logL(\sigma,\gamma,\mu)$

$$= -m\ln(\sigma) - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{m}\ln\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\ln\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right)^{-\frac{1}{\gamma}} \quad \text{A2.10}$$

Provided $1 + \gamma((Y_i - \mu) / \sigma) > 0$, when $\gamma = 0$, the log-likelihood function reduces to *logL(σ,γ,μ)*

$$= -m\ln(\sigma) - \sum_{i=1}^{m} e^{\left(-\frac{Y_i-\mu}{\sigma}\right)} - \sum_{i=1}^{m}\left(\frac{Y_i-\mu}{\sigma}\right) \qquad \text{A2.11}$$

[2] Beirlant

The MLE for $(\sigma,\gamma,\mu)$ are derived by maximizing A2.10 – A2.11 in equating the partial derivatives (w.r.t .the parameters) of A2.10 to zero and solving for the parameters.

A condition for MLE method is that the data must be free of the parameters that need estimating, this is usually referred to as regularity.

The GEV is known to be “non-regular” for $\gamma \neq 0$.

**II] Probability Weighted Moments (PWM)**

The PWM estimator for $(\sigma,\gamma,\mu)$ is solution to the simultaneous moments equations below [2] Beirlant, p134:

$M_{1,0,0} = \mu - \sigma \,.\, \{1-\Gamma(1-\gamma)\} / \gamma$ A2.12

$2M_{1,1,0} - M_{1,0,0} = \sigma \,.\, \{\Gamma(1-\gamma)(2^{\gamma}-1)\} / \gamma$ A2.13

$(3M_{1,2,0} - M_{1,0,0}) / (2M_{1,1,0} - M_{1,0,0})$
$= (3^{\gamma}-1) / (2^{\gamma}-1)$ A2.14

The asymptotic consistent estimator for $M_{1,r,0}$ is:

$$\hat{M}_{1,r,0} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n+1}\right)^{r} Y_{i;n} \qquad \text{A2.15}$$

By using A2.15, $M_{1,0,0}$ corresponds to the arithmetic mean which would estimate $\mu$; one would then proceed to compute $M_{1,1,0}$ and $M_{1,2,0}$ by using A2.15, the LHS of

A2.14 is then solved numerically to estimate $\gamma$, subsequently A2.13 is solved for $\sigma$.

For both methods above, one can apply inferential techniques to obtain confidence intervals on the estimators [2] Beirlant p137.

**III] Regular Variation Approach (RVA)**

If expected $\gamma > 0$ then one may utilize this type of estimation for $\gamma$. This estimator is essentially based on the same principles as the Pickands estimator even though it bears some resemblance to the Hill estimator.

$$\hat{\gamma}_{m,n}^{(RVA)} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{ln}\left(X_{i,n}\right) - \mathbf{ln}\left(X_{m+1,n}\right)$$

**IV] L-Moments (L-M)**

In the works of [21] Hosking, one may see that the use of L-M is an efficient method in estimating parameters of a wide range of distributions from small samples. The Probability Weighted Moments (PWM) are the precursors to L-M. The L-M are linear combinations of PWM that have simple interpretations as measures of location, scale (dispersion) and shape. These linear combinations of PWM make them less subjected to outliers.
If $F_X(x)$ is given then, the PWM are, by definition:

$$\beta_r = \int x\{ F_X(x)\}^r \,.\, d F_X(x), \quad r = 0,1,2,...$$

The L-M are linear combinations of $\beta_r$ transformed by means of shifted Legendre polynomials, which are orthogonal on the interval (0;1) with a constant weight function.

$$\therefore \quad \left.\begin{aligned} \lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned}\right\} \qquad \text{A2.16}$$

The LM ratios are defined as, for $0 < \tau < 1$ : $\tau = \lambda_2 / \lambda_1$ and $\tau_r = \lambda_r / \lambda_2$

$\tau_3$ is a measure of skewness and $\tau_4$ is a measure of kurtosis, $-1 < \tau_3 ; \tau_4 < +1$

The L derived Coefficient of Variation is: L-CV = $\tau$

From a sample size $n$, the PWM can be derived from

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$b_r = \frac{1}{n}\sum_{i=r+1}^{n} \frac{(i-1)(i-2)...(i-r)}{(n-1)(n-2)...(n-r)} X_i$$

the $\beta_j$ PWM can be estimated by $b_j$ , $j$ =0,1,…,$r$
subsequently the $\lambda_i$ can be estimated from $\ell_i$ derived from the set of equations A2.16
by substituting $b_j$ for each $\beta_j$ and for each $\lambda_i$ linearised moment required. Subsequently the $\tau$ and $\tau_r$ ratios can also be estimated by $u$ and $u_r$

The workings of [40] van Gelder and [30] Pandey have shown the application of LM in the estimation of the GPD parameters.

The Location ($t$), the Scale ($\sigma$) and the shape ($\gamma$) can be then estimated by:

$$\left.\begin{aligned} \hat{\gamma} &= (3\ u_3 - 1)/\ (u_3 + 1) \\ \hat{\sigma} &= (1 - \hat{\gamma})\ (2 - \hat{\gamma})\ \boldsymbol{\ell}_2 \\ \hat{t} &= \boldsymbol{\ell}_1 - (\ 2 - \hat{\gamma})\ \boldsymbol{\ell}_2 \end{aligned}\right\} \qquad \text{A2.17}$$

This method when the 3 parameters need estimating is known as the 3-LM method, Pandey [30] et al; in that work the efficiency of this method was illustrated.

When ***t***, the location parameter, is known, e.g. a given threshold, the method reduces to what is termed the 2-LM method and the estimates may be computed as follows:

$\hat{\gamma} = 2 - \{(\lambda_1 - \boldsymbol{t}) / \lambda_2\}$

and

$\hat{\sigma} = (1 - \hat{\gamma})\ .\ (\lambda_1 - \boldsymbol{t})$

Pandey [30] states in the workings that LM are efficient estimators of the GPD when using POT methodology to estimate the tail of extremes.

### 2.3.1 Method of BM

This method is particular to the Fisher-Tippett Theorem which deals with the limit law for the BM with n being the block size.

Let $Y_i$ be an iid GEV rv derived as sample maxima from observations $X_j$

where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$ (in Fig A1.4 $n$=4 and $m$=20).

By using Eq. A2.2 and substituting $x$ for $(x - \mu)/\sigma$ one can then use the methodologies described above ( **I]** to **IV]** ) to estimate the GEV distribution parameters and corresponding confidence intervals (see **II]** above) [2] Beirlant p137.

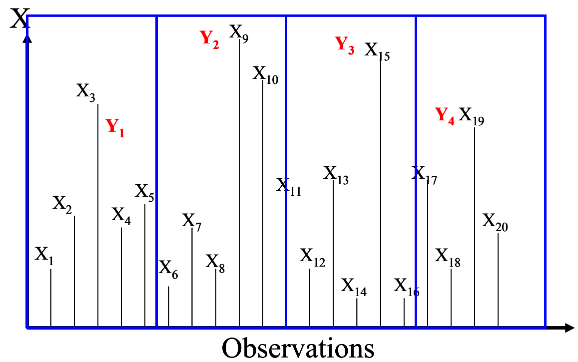



**Fig A1.4: Observations vs Exceedances (BM)**

The inversion of the GEV distribution function yields the estimates of the extreme quantiles:

$$Q_{Y;p} = \begin{cases} \mu + \sigma/\gamma\,[(-\log(1-p))^{-\gamma} - 1]\,, & \gamma \neq 0 \\ \mu - \sigma \log(-\log(1-p)), & \gamma = 0 \end{cases}$$

Whereby $\mu$, $\sigma$ and $\gamma$ are estimated according to either methods **I]** to **IV]**.
When the GEV is used as an approximation of the largest observation in a block sample, then:

$$q^*_{Y;p} = \begin{cases} \mu + \sigma/\gamma\,[(-log(1-p)^n)^{-\gamma} - 1]\ ,\ \gamma \neq 0 \\ \mu - \sigma \log\,(-log(1-p)^n),\ \gamma = 0 \end{cases}$$

where $n$ is the block length, [2] Beirlant

## 2.3.2 Method of POT

Here one is dealing with the conditional survival function of the exceedances (or peaks) $Y = X - \boldsymbol{t}$ , where $\boldsymbol{t}$ is the threshold. In some literature (and above) this is also referred as $\boldsymbol{u}$

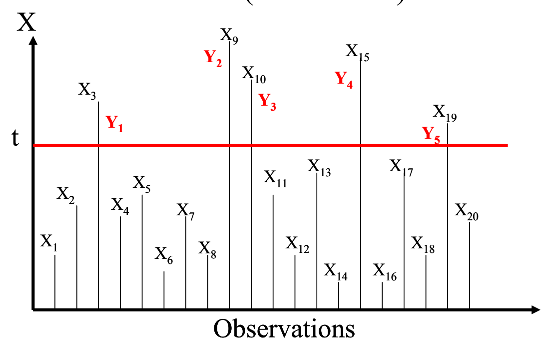



**Fig A1.5: Observations vs Exceedances (POT)**

Fig A1.5 above depicts the series $Y$ over a threshold $\boldsymbol{t}$. The differences between the BM and POT methods are clearly visible: the BM method yields $Y_1$ to $Y_4$, while the POT method yields $Y_1$ to $Y_5$ (inclusion of $X_{10}$ in the maxima dataset).

Although the ML and PWM present themselves in this paragraph, the methodology utilised here to estimate the EVI, is termed the Elemental Percentile Method (EPM) [2] Beirlant. This method overcomes some of the difficulties in the ML and PWM methods ($\gamma$ restrictions). In the EPM there are no $\gamma$ restrictions.

Following up from the visual of Fig A1.5, let $Y_i$ be an iid GEV rv derived as sample maxima from observations $X_j$ where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

A level $\boldsymbol{t}$ is chosen as a threshold as a point of departure whereby the exceedances of $X$ are observed above $\boldsymbol{t}$ (see Fig. A1.5).

This means that $P[M < \boldsymbol{t}] = \{P[Y_i < \boldsymbol{t}])^N$

Therefore there are two important issues to consider when using the POT method:

i) the selection of $\boldsymbol{t}$, and

ii) the independence of the peaks

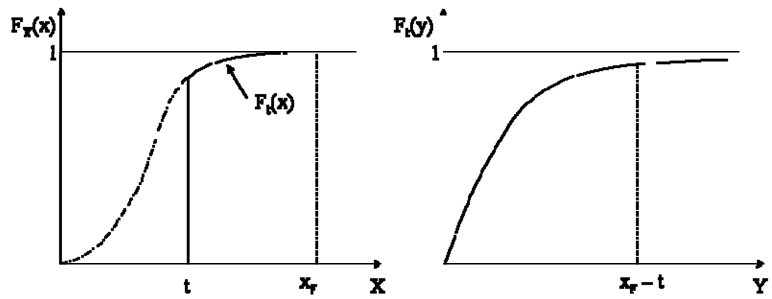



**Fig A1.6: CDF of X and Conditional Dist. Function on *t***

The discussions and debates on the selection of ***t*** still ensue today ([11] De Waal, et al). It may be chosen graphically ([2] Beirlant [13] Embrechts) and visually.
In this method one would "choose" a threshold or estimate it by the MEF or EDP techniques (see below). Then one would estimate the parameters of the GPD by the methods **I]** to **IV]** above. This does not go through without challenges, as the estimation of ***t*** and the GPD parameters are intertwined.
The selection of this threshold is main issue: if ***t*** is too high then there are too few exceedances and therefore too high variance of the parametric estimators; if ***t*** is too low the estimators become biased and a GPD approximation is not possible; hence the challenge to obtain an "optimal ***t***". The choice of ***t*** is critical for a good fit of the excesses to the distribution function. This threshold may be estimated as a trade off between the goodness-of-fit (it must be noted that this increases as ***t*** is set higher [for maxima]) and the available number of excesses above ***t***. It may be derived by using the mean squared error of the Hill estimator ([11] De Waal, et al) or by the usage of the graphical choice (Visual Method - VM) which implies a visual determination of a "kink" in the tail of the GPD Q-Q plot; the threshold level is determined by inversing the Q function. A spin-off of this method is the determination of the EVI by isolating the values to the right of the "kink". Subsequently a linear regressional fit is applied to these values and the estimated slope determines the EVI.
Three methods were researched in this thesis to arrive at a solution for this challenge: the Visual Method (VM), the Mean Excess Function (MEF) method and the Entropy of Dirichlet Process method (EDP) [10] De Waal.

In the POT methodology, at the threshold ***t*** one observes the exceedances of ***t*** and this means that the distribution of the excesses is given by:

$$F_t(y) = Pr\{ X \leq t+y \mid X > t\} = \frac{F_X(t+y) - F_X(t)}{1 - F_X(t)} \quad , y > 0$$

The asymptotic form of $F_t(y)$ was given first by [31] Pickands, which states that if the EVD of $F$ exists, there are constants $c_t > 0$ such that as ***t*** tends to either $-\infty$ or $+\infty$ , then:

$F_t(c_t \,.\, z) \rightarrow$ GPD, as given in Eq A2.9

One benefit of POT over BM is that since each exceedances is associated with a specific event, it is possible to make the parameters $\sigma$ and $\gamma$, in Eq. A2.9 depend on covariates [34] Smith, et al.

For instance, one could be assessing the probability of a certain exceedances in the UCLF, as a function of the Load Factor, which is the energy demand imposed on the installed capacity.

It is generally accepted that from a time-series analysis perspective the POT approach is preferred to the BM one (albeit its challenges: does not handle stochastic volatility well in financial time series).

If we allow time to play an important role in POT, then one can express Eq A2.9 in the form of an intensity set A of exceedances within a time window ($T_1$ ; $T_2$) as:

$\Lambda\,(A) = (T_2 - T_1)\,(1 + \gamma.(y - \mu)\,/\zeta)^{-1/\gamma}$

The interpretation of these parameters is equivalent to those in Eq. A2.9, suffice it to say that, if T is in years, then it reflects the probability that the set $A$ is “empty” of exceedances or that the probability of an annual maximum is $\leq y$. The GPD can be derived as a consequence of $\Lambda\,(A)$

which ties in with the POT method. In the author's opinion, research ought to be done in this camp as the distribution of T might be inverse-Gaussian when ($T_1$ ; $T_2$) is not considered as fixed but rather as a time variant interval, under which exceedances occur (i.e. the exceedances is the event that triggers the time-interval, e.g. time-to-fail or rather time-to-exceed ***t***).

**2.4 Mean Excess Function**

This type of diagnostic may be used to test the assumptions on the variable's behaviour under a threshold condition.
As mentioned, the ***t*** may be derived by the use of the MEF on the GPD [13] Embrechts. The MEF can be a useful tool for distinguishing distribution functions in their right tails (and its estimations).

Example using the MEF (refer to Fig A1.6):
Let us say that we model $Y$ as a $Max(X) - \boldsymbol{t} \sim Exp(\lambda)$, i.e.
$y_i = Max(x_j) - \boldsymbol{t}$ , with CDF: $F(y) = 1 - e^{(-y\lambda)}$
Let $e(\boldsymbol{t}) = 1/\lambda$, be the Mean Excess Function, then
$1/\lambda$ is the <u>mean exceedances above ***t***</u> and can be estimated by $\frac{1}{n}\sum_{i=1}^{n} y_i$
Other distributions' MEF's can also be derived, e.g. the Pareto as indicated in the table below, [13] Embrechts.

| Distribution | pdf | e(t) |
|---|---|---|
| Pareto | $\dfrac{\alpha.\kappa^{\alpha}}{x^{\alpha+1}}$ | $\dfrac{\kappa+t}{\alpha-1}$ , α >0 |
| Exponential | $\lambda.e^{(-\lambda x)}$ , $x,\lambda$ >0 | $\dfrac{1}{\lambda}$ |
| Burr | $\dfrac{\tau.\alpha.\lambda^{\alpha}(x^{\tau-1})}{(\lambda+x^{\tau})^{\alpha+1}}$ | $\dfrac{t}{\alpha\tau-1}$ , α >0 |
| GPD | $(1/\sigma)\,[1+(\gamma x/\sigma)]^{-(1/\gamma)-1}$ | $\dfrac{\sigma+\gamma t}{1-\gamma}$ , $1>\gamma>0$ |

**Table A1.1: Mean Excess Functions based on threshold *t***

From Table A1.1 one would plot various thresholds ***t*** on the *x*-axis versus its relative MEF (*y*-axis).
In the case of a GPD, from Table A1.1,
$e(\mathbf{t}) = [\{\sigma/(1-\gamma)\} + \{\gamma/(1-\gamma)\}.\ \mathbf{t}]$, one would expect that by plotting ***t*** vs $e(\mathbf{t})$, one should observe a straight line with slope $\{\gamma/(1-\gamma)\}$. Hence this has the benefit of selecting a threshold ***t*** , given that we estimated $\gamma$. However some caution needs to be exercised, since (as mentioned before) for large ***t*** there are a few observations and therefore a high variability in the mean.
Let a threshold $\mathbf{t_0}$ be of a "high" value and let $(y-\mathbf{t_0})$ follow a GPD with parameters $\gamma$ and $\sigma$ with $0<\gamma<1$ and the
$E\{(y-\mathbf{t_0})|\ y>\mathbf{t_0}\} = \sigma/(1-\gamma)$, then $\forall\ \mathbf{t}>\mathbf{t_0}$, the MEF is by definition:
$e(\mathbf{t}) = E\{(y-\mathbf{t})|\ y>\mathbf{t}\} = [\sigma-(-\gamma).(\mathbf{t}-\mathbf{t_0})]/(1-\gamma)$
$= [\sigma+\gamma.(\mathbf{t}-\mathbf{t_0})]/(1-\gamma)$

which is of the form given in Table A1.1 for the GPD.
Hence for a given $\gamma$, $e(\boldsymbol{t})$ is a linear function of $\boldsymbol{t}$ and provides a method to infer $\boldsymbol{t_0}$ for the GPD.
Let the empirical MEF be

$\hat{e}(\boldsymbol{t}) = \frac{1}{N_t}\sum_{i=1}^{N_t}(y_i - t)$, where $Y$ is the ordered-$X$, then by plotting $\hat{e}(\boldsymbol{t})$ on $\boldsymbol{t}$ we get the Mean Excess plot, expected to be linear at $\boldsymbol{t} > \boldsymbol{t_0}$.
It can be shown that a reduction of $\hat{e}(\boldsymbol{t})$ is $\hat{e}(\boldsymbol{t}) = -t + \frac{1}{N_t}\sum_{i=1}^{N_t}(y_i)$, for each of the $N_t$ 's, this tends to simplify the computations.

Although, practically, it may be difficult to interpret the plot because for large $\boldsymbol{t}$ there are few exceedances as mentioned above, the plot's real purpose is to detect significant shifts in the slope at different $\boldsymbol{t}$'s.

More sophisticated methods may be used when sampling $Y$'s off the $X$'s especially under a Time Series condition, whereby the peaks ought not to be associated with cyclic behaviour of the variable $X$ (see, for instance, [13] Embrechts p270 on the effects of autocorrelation when using the Hill estimator).

**2.5 GEV Mean and Variance**

Here the Mean and Variance of the GEV is related to $\mu$ and $\sigma$ (see SGEV in 2.1 above)

E[$X$] = $\mu + \sigma$ . $\{\Gamma(1-\gamma) - 1\} / \gamma$ $\quad\rightarrow\quad$ $\mu +$ 0.5772 $\sigma$ for $\gamma \rightarrow 0$

Note that $\psi(1)$ = diGamma = -0.5772, and 0.5772 is Euler's constant

and

Var[$X$] = $\sigma^2$. $\{\Gamma(1-2\gamma) - \Gamma^2(1-\gamma)\} / \gamma^2$

$\rightarrow$ [($\pi^2$)/6].$\sigma^2$ for $\gamma \rightarrow 0$

Note that $\psi'(1)$ = triGamma = ($\pi^2$)/6

where $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$

(***See Appendix C paragraph 2.7 for more details***)

# APPENDIX B

## Work on Price

To perform work in the area of electricity prices in a power pool, an understanding of the meaning of Returns is necessary.

**Definition A**: **Returns**
Returns are defined as the change in value of the asset plus the cumulative cash flow over the original value of the asset [38] Wilmott
Let $S_t$ be the asset value at unit time t.
Then the Returns from time unit t to time unit t + 1 is defined [38] Wilmott:

$$\mathbf{R_t = (S_{t+1} - S_t) / S_t} \qquad \text{B1.0}$$

## 1. Modelling Returns: First Approach

To guide us in the choice of what parameter would represent the randomness of the Returns, one could select a measure of spread within a certain time differential.
For instance, one could choose the standard deviation of the Returns, defined as

$$\mathbf{StDev\ (R_t)} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(R_t - \overline{R})^2} \qquad \text{B1.1}$$

Where, $\overline{R} = \frac{1}{n}\sum_{t=1}^{n} R_t$

**Assumption 1.1**: Let the Returns be Normally distributed, i.e. $\mathbf{R \sim N(\mu;\sigma^2)}$
This implies the following:

$$\mathbf{f_R(r)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}$$

Let $\mathbf{Z = (R - \mu) / \sigma}$ , to standardise.

Implying $\mathbf{R = \mu + Z.\sigma}$ B1.2

With,

$$\mathbf{f_Z(z)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}}$$ , and $\mathbf{Z \sim N(0;1)}$

***The Returns are now being expressed by means of (B1.1) a measure of expectation $\mu$ and variation $\sigma$.***

### 1.1 Effect of time upon Returns

It is however expected that the time unit of choice would also have some effect on the behaviour that is attempted to be modelled (If one samples hourly, daily, weekly,…, the results would be different).

Under the absence of randomness, (B1.2) would be:

$\mathbf{R = \mu}$ B1.3

**Assumption 1.2**: Let $\boldsymbol{\delta t}$ be a determined time differential (e.g. $\delta t$ could be a time step of 1 hr, 1 day,…)

$\mathbf{R = \mu \,.\, \delta t}$ B1.4

if $\delta t = 1$ , (B1.3) =(B1.4)

Equation A can now be expressed in the form:

$\mathbf{R_t = (S_{t+1} - S_t) / S_t = \mu \,.\, \delta t}$

Re-arranging:

$\mathbf{S_{t+1} = S_t (1 + \mu \,.\, \delta t)}$ B1.5

Starting at t = 0, equation (1.5) becomes after one time step $t = \delta t$,

$\mathbf{S_1 = S_0 (1 + \mu \,.\, \delta t)}$ B1.6

after the second time step $t = 2\delta t$,

$\mathbf{S_2 = S_1 (1 + \mu \,.\, \delta t)}$ B1.7

Substituting (B1.6) in (B1.7):

$\mathbf{S_2 = S_0 (1 + \mu \,.\, \delta t)^2}$

after n time steps $t = n\delta t$

$\mathbf{S_n = S_0 (1 + \mu \,.\, \delta t)^n}$

Let $\mathbf{X = e^{\ln(X)}}$ , $\mathbf{X^n = e^{n.\ln(X)}}$

therefore $\mathbf{S_0 \,.(X)^n = S_0} \,.\, \mathbf{e^{n.\ln(X)}}$ , by letting $\mathbf{X = (1 + \mu \,.\, \delta t)}$, we get:

$\mathbf{S_0 (1 + \mu \,.\, \delta t)^n = S_0} \,.\, \mathbf{e^{n.\ln(1 + \mu \,.\, \delta t)}}$ B1.8

Since $\mathbf{(1 + \mu \,.\, \delta t)^n \approx e^{n.\,\mu\,.\,\delta t}}$, by Taylor's expansion and Infinite series
and substituting in (B1.8),

$$\mathbf{S_n = S_0\,(1 + \mu \,.\, \delta t)^n \approx S_0 \,.\, e^{n.\,\mu\,.\,\delta t}}$$

Then, if after n time steps, let T = n.δt,

$$\mathbf{S_n = S_0\,(1 + \mu \,.\, \delta t)^n \approx S_0 \,.\, e^{\mu\,.T}} \qquad \text{B1.9}$$

**1.2 Random representation**

In the derivation of equation (B1.9), no variation was taken account of. If now the variation effect is taken into account and imbedding the time component, then equation (B1.1) is affected by n.δt terms by using the same argument in the paragraph above.
Since T = n.δt , n = T / δt , substituting in (B1.1):

$$\mathbf{StDev\,(R_t)} = \sqrt{\frac{\delta t}{T-\delta t}\sum_{i=1}^{T/\delta t}(R_i - \overline{R})^2} \qquad \text{B1.10}$$

Hence, the deviation of the Returns is of the order $\mathbf{(\delta t)^{1/2}}$, if it were to be represented in equation (B1.4) in taking cognisance of the variation.
From equation (B1.2), the Returns could be modelled as:

$$\mathbf{R_t = (S_{t+1} - S_t) / S_t = \mu.\delta t + Z\sigma\sqrt{\delta t}}$$

Which may be written as:

$$\boxed{\underbrace{\mathbf{(S_{t+1} - S_t)}}_{\text{Change in Asset price}} = \underbrace{\mathbf{\mu.S_t\delta t + \sigma S_t Z\sqrt{\delta t}}}_{\text{Model representing a Random process}}} \qquad \text{B1.11}$$

Another way of looking at equation (B1.10) is:

**Var ($R_t$) = E [$(R_t - \overline{R})^2$] = $\sigma^2$,**

Over small time unit intervals, the variance is: **Var ($R_t$) . $\delta t$ = $\sigma^2 . \delta t$**

Therefore the standard deviation of the Returns can be expressed by:

**Stdev ($R_t$) = $\sigma\sqrt{\delta t}$** B1.12

Combining equations (B1.2), (B1.4) and (B1.12), the same result as equation (B1.11) can be obtained.

**<u>The model represented in equation B1.11 is fundamental in modelling the behaviour of returns</u>**.

***Remarks 1:*** ***1.1] The prices themselves take on discrete values and change at fixed point in time***
***1.2] Changes in the values of prices over time are uncertain or random***
***1.3] The following approximations apply:***
***a) the random process is continuous over time***
***b) the spot prices vary continuously***

# 2. Stochastic Processes

**<u>Definition 2.1</u>**: **Markov Process** (MP)

A Markov Process is a stochastic process where only the value of a variable at present time t, is relevant for predicting the future [38] Wilmott.

In statistical terms, it's the conditional distribution of X(t), given the information up until $\tau < t$ , depends only on $X(\tau)$.[38] Wilmott

**<u>Definition 2.2</u>**: **Wiener Process** (WP) or Brownian Motion
It is a particular kind of MP and is given as the limiting process of a random walk as the time differential between the steps tends to zero. [38] Wilmott
In statistical terms:
If W ~ WP, then:
1) $\mathbf{\Delta W = Z\sqrt{\Delta t}}$ and $\mathbf{Z \sim N(0;1)}$ B2.1
2) Values of $\mathbf{\Delta W}$, for any two different short time intervals $\Delta t$ are independent $\therefore$ $\mathbf{\Delta W}$ are not correlated over time.
**Properties of a WP**:
Let W ~ WP, then:

$$\begin{aligned}\mathbf{E[\Delta W]} &= \mathbf{E[Z\sqrt{\Delta t}\,]}\\ &= \mathbf{E[Z]\,.\,E[\sqrt{\Delta t}\,]}\\ &= \mathbf{0\,.\,E[\sqrt{\Delta t}\,]}\\ &= \mathbf{0}\end{aligned}$$

$$\mathbf{Var[\Delta W]} = \mathbf{E[\{\Delta W\ -\ E[\Delta W]\}^2}$$

The expected value of $\Delta W$ is zero (see above), hence:

$$\begin{aligned}\mathbf{Var[\Delta W]} &= \mathbf{E[\{\Delta W\}^2]}\\ &= \mathbf{E[\{Z\sqrt{\Delta t}\,\}^2]}\\ &= \mathbf{\Delta t\,.\,E[\{Z\}^2]} \quad \text{but } E[Z^2] = Var(Z) = 1\\ &= \mathbf{\Delta t}\end{aligned}$$

**<u>Definition 2.3</u>**: **Generalised Wiener Process** (GWP)
A GWP for a variable X, is defined as the summation of a drift and a stochastic component, in the form of:

$$\mathbf{dx = a.dt + b.dW}$$

where,

**a and b** $\in \Re$
**W ~ WP**
**dW = Z**$\sqrt{\Delta \mathbf{t}}$

**Properties of a GWP**

- If b = 0 , then dx/dt = a (constant). By integrating, $x = a.t + x_0$ ; $x_0 = x(0)$
  Hence, X has an expected drift rate of "a", per unit time.
- If a =0 , then dx = b. Z$\sqrt{\Delta \mathbf{t}}$ , this is a WP with zero drift and variance rate of b2, per unit time.

**<u>Definition 2.4</u>**: **Itô Process** (IP)
An IP is a GWP where "a" and "b" are functions of time "t" and a variable X expressed in the following form:

$$\mathbf{dx = a(x;t).dt + b(x;t).dW}$$

**Lemma**: **(Itô's Lemma)**
Suppose that dx follows an IP, then:
Some function dG of X and t follows:

$$\mathbf{dG} = \left(\frac{\partial G}{\partial x}\mathbf{a} + \frac{\partial G}{\partial t} + \frac{1}{2}\,.\frac{\partial^2 G}{\partial x^2}\mathbf{b^2}\right)\mathbf{dt} + \frac{\partial G}{\partial x}\mathbf{b.\,dW} \qquad \text{B2.2}$$

# 3. Modelling Returns: second method

Modelling Returns as a Generalised Wiener Process
Let

$\mathbf{S_t}$ = Spot price @ time t
$\mathbf{t}$ = time of observation
$\mu$ = drift rate
$\sigma$ = volatility
$\mathbf{dW = Z\sqrt{\Delta t}}$

and $\mathbf{dS_t = (S_{t+1} - S_t)}$

$\therefore$ Looking at the form of a GWP and it could be said that $\mathbf{S_{t+1}}$ is the stochastic price following a GWP given $\mathbf{S_t}$ . Then from Definition 2.3,

$$\mathbf{dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dW}$$
$$\mathbf{= \mu\, S_t\, dt + \sigma\, S_t\, Z\sqrt{\Delta t}} \quad \mathbf{, dt << 1} \qquad \text{B3.1}$$

or as dt $\rightarrow$ 1,

$$\mathbf{S_{t+1} - S_t = \mu\, S_t\, \delta t + \sigma\, S_t\, Z\sqrt{\delta t}}$$

<u>which is similar to Eq. B1.11, which is similar to Eq. B3.1 when $\mathbf{\delta t} \rightarrow 0$</u> (see [38] Wilmott p61 last two paragraphs of 3.7).

***But Eq. B3.1 lacks continuity and therefore cannot be solved using normal calculus (remember: Z is stochastic, Z ~ N(0;1) ).***

Hence, Eq. B3.1 can be solved by either using a numerical method (such as Monte Carlo simulation) or Stochastic integration (e.g. using the Gamma Function, its Lemma and moment Generating Functions).
The important point of using a GWP, is that now build up a continuous-time model instead of discrete time events.

Due to the exponential properties of the spot prices (e.g. Eq. B1.9, $\therefore$ $\ln(S_n/S_o) = \mu T$), returns are generally modelled in terms of the logs.
Hence, the widely accepted model for returns, equities, currencies, commodities and indices, is expressed in terms of the stochastic equation:

$$\boxed{\mathbf{dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dW}} \qquad \text{B3.2}$$

<u>**This is a continuous-time model of an asset price, and forms the foundation of Finance Theory**</u> (notice the similarity with Eq B1.11 and B3.1).

Using Itô's Lemma, Eq. B3.2 can be expressed in terms of $\ln(S_t)$

Let $G = \ln(S_t)$; $\quad \dfrac{\partial G}{\partial S_t} = \dfrac{1}{S_t}$ **;** $\quad \dfrac{\partial^2 G}{\partial S_t^{\,2}} = \mathbf{-}\ \dfrac{1}{S_t^{\,2}}$ **;** $\quad \dfrac{\partial G}{\partial t} = 0$;

$a = \mu\, S_t$ and $b = \sigma\, S_t$
Substituting in Eq. B3.2, we get:

$$\mathbf{d\{ln(S_t)\} = (}\frac{\partial G}{\partial S_t}\mathbf{\mu S_t} + \frac{\partial G}{\partial t} + \frac{1}{2}\ \mathbf{.}\ \frac{\partial^2 G}{\partial S_t^{\ 2}}\mathbf{\sigma^2 S_t^2)\ dt}$$

$$+\ \frac{\partial G}{\partial S_t}\mathbf{\sigma\ S_t.\ dW}$$

$$= (\mu - \frac{1}{2}\sigma^2)\ \mathbf{dt + \sigma\ .\ dW} \qquad \text{B3.3}$$

Therefore, $\ln(S_t)$ follows a GWP with rates:
drift = $\mu - \frac{1}{2}\sigma^2$ and variance = $\sigma^2$
$\therefore$ if $dt = T - t$, this means:

$$\mathbf{d\{ln(S_t)\} = ln\ S_T - ln\ S_t \sim N\{(\mu - \frac{1}{2}\sigma^2)(T - t); \sigma^2 (T - t)\}}$$

$$\therefore\quad \mathbf{ln}\ \{\frac{S_T}{S_t}\} \sim \mathbf{N\{(\mu - \frac{1}{2}\sigma^2)(T - t); \sigma^2 (T - t)\}}$$

From Eq. B3.3,

$$\mathbf{ln}\ \{\frac{S_T}{S_t}\} = \mathbf{(\mu - \frac{1}{2}\sigma^2)(T - t) + \sigma (W_T - W_t)}$$

$$\therefore \quad \boxed{\mathbf{S_T = S_t\ .\ exp\{(\mu - \frac{1}{2}\sigma^2)(T - t) + \sigma (W_T - W_t)\}}}$$

***Hence, the spot price at time T, contingent on the spot price at t, was derived.***

In addition:

$$\mathbf{E[S_T] = S_t \,.\, exp\{(\mu - \frac{1}{2}\sigma^2)(\,T - t)\}\,.\, E[exp\{\sigma\,(W_T - W_t)\}]}$$

$$\mathbf{= S_t \,.\, exp\{(\mu - \frac{1}{2}\sigma^2)(\,T - t)\}.\, exp\{\frac{1}{2}\sigma^2\,(T - t)}$$

$\therefore$ $\mathbf{E\,[S_T\;] = S_t \,.\, e^{\,\mu\,.(T - t)}}$

$\therefore$ $\mathbf{E\,[S_T\;] = S_t \,.\, e^{\,\mu\,.dt}}$ B3.4

(Compare this result to Eq. B1.9 for interest)

Equation B3.4 implies the following important points:

- The expected spot prices grow exponentially over time with an expected return m (remember the Nominal Annual continually Compounded
  $A = P.e^{rt}$, where P = Investment, r = Interest and
  t = maturity time)
- The spot prices are always positive
- No mean reversion was represented
- No seasonality was modelled in Eq. B3.4
- The volatility is constant

**<u>REMARKS</u>**:

- A point of interest here is that the violation of the price positivity, would destroy Eq. B3.2 which forms the basis of Finance Theory (e.g. fair pricing of all types of options would be destroyed and infer chaotic behaviour in the derivatives markets).
  The Australian market experimented at paying the customers to take the energy for production purposes. This induces negative pricing, destroying Eq. B3.2.
  The argument still stands that if the customer uses the energy still (as it's not storable), then why not sell it at the very minimal unit price (e.g. R1, US$ 1, €1, A$1, £1, ¥1, etc…).

- Another interesting aspect is the distribution of the Returns. Is it important for $\frac{\delta S}{S}$ to be $\mathbf{N(\mu;\sigma^2)}$?
  Let $\mathbf{Y} = \frac{\delta S}{S}$ be a random variable with a determined distribution. What matters as far as transaction costs, is not the mean of Y, nor its standard deviation ([38] Wilmott p346, 24.11) but $|\overline{Y}|$ (the absolute average value of the relative returns).

- Let $^{A}\mathbf{R_t}$ be the arithmetic representation of the Returns (also termed ***Net Returns***) and $^{G}\mathbf{R_t}$ the geometric representation (also termed ***Log Returns***) of the Returns. Then mathematically they can be expressed [42] van Zyl as:

$$^{A}R_t = \frac{S_{t+1} - S_t}{S_t} \quad \text{and} \quad ^{G}R_t = \ln\left(\frac{S_{t+1}}{S_t}\right)$$

$$\therefore \qquad ^{G}R_t = \ln\left(\frac{S_{t+1} - S_t + S_t}{S_t}\right) = \ln(1 + {}^{A}R_t)$$

$$= {}^{A}R_t - \frac{{}^{A}R^2{}_t}{2} + \frac{{}^{A}R^3{}_t}{3} - \frac{{}^{A}R^4{}_t}{4} + \ldots$$

[17] Giordano

If $^{A}R_t \ll 1$, then

$$^{G}R_t \approx {}^{A}R_t$$

(However the problem comes into using the standard deviation of the Arithmetic returns, as we'll see below)

Also,

$$^{G}R_t = \ln(1 + {}^{A}R_t)$$

$$^{G}R_t \ \ln(e) = \ln(1 + {}^{A}R_t) \text{ , since } \ln(e) = 1$$

i.e.

$$\ln(\exp\{{}^{G}R_t\}) = \ln(1 + {}^{A}R_t)$$

$$\therefore \qquad \exp\{{}^{G}R_t\} = 1 + {}^{A}R_t$$

$$\boxed{\exp\{{}^{G}R_t\} - {}^{A}R_t = 1}$$

***This is an exact relationship between the arithmetic and geometric interpretations of the Returns***

Scenario 1 (S1): Take a change between 100 @ time t, down to 20 @ time t+1, then

$$^{A}\mathbf{R_t} = \frac{\mathbf{S_{t+1} - S_t}}{\mathbf{S_t}} = (20 - 100)/100 = -80\% \;;$$

$$^{G}\mathbf{R_t} = \mathbf{ln}\ (\frac{\mathbf{S_{t+1}}}{\mathbf{S_t}}) = \ln(20/100) = -1.61$$

Scenario 2 (S2): Take a change between 20 @ time t, up to 100 @ time t+1, then

$$^{A}\mathbf{R_t} = \frac{\mathbf{S_{t+1} - S_t}}{\mathbf{S_t}} = (100 - 20)/20 = 400\% \;;$$

$$^{G}\mathbf{R_t} = \mathbf{ln}\ (\frac{\mathbf{S_{t+1}}}{\mathbf{S_t}}) = \ln(100/20) = +1.61$$

This clearly indicates that taking the standard deviation of all the $^{A}\mathbf{R_t}$ induces artificial bias in the distribution of the Net Returns (different values [-80% ; 400%] for the same change of 80).

Hence expressing the volatility as a percentage of the Log Returns seems the more plausible thing to do. ***However, this could result in a debate in terms of Eq. B1.11 and Eq. B3.1***

# APPENDIX C

## 1. Exploratory Data Analysis (EDA)

In EDA, some methods in Data Mining were used:

- Data exploration, i.e. investigating:
    - the Problem Space: Extreme GWh losses events occurring in the production of electricity
    - the Solution Space: application of EVT to understand the behaviour of these extreme events
    - the Data Space itself: data collection, preparation, use of EDA techniques
- Data preparation
    - Data cleansing: Care and caution was taken so as no data tampering happened.
    - Meticulous scrutiny of the data was carried out to assure that:
        - No duplication occurred (software transfer between media)
        - There were no inconsistencies (e.g. same acronym for different Power Stations, like Mt for both Matimba and Matla)
        - Default values were correct (e.g. a "blank" for missing values and not a zero)
- Data transformation
    - Transforming UCLF percentages into GWh losses
    - Encoding by scale augmentation: Clustering of GWh losses into event categories (such as Main Event, Semi-catastrophic an Catastrophic)
    - Basic data enhancement such as deriving Total GWh losses for each year over the period, descriptive statistics and basic plots.

The working files are named: “Units 1990-2005.xls” and “GWh Losses EDA v3.xls”, “Workbook.xls” and “Workbook (Entropy-GEV).xls”

**Software Used**:

StatsticA for Windows v 5.1
Matlab v5.3 Used at the Risk Laboratory at the University of the Free State
ForecastPro for Windows
Excel 2002 Including macros and Visual Basic coding
Powerpoint 2002
Word 2002

# 2. General Statistical Work

## 2.1 Bayes Theorem

From Bayes Theorem:

$f_{X|Y}(x|y) = f_{X,Y}(x,y) \,/\, f_Y(y)$ C2.0

When X and Y are independent:

$f_{X,Y}(x,y) = f_X(x) \,.\, f_Y(y)$ C2.1

By substituting C1.1 in C1.0

$f_{X|Y}(x|y) = f_X(x)$

## 2.2 Binomial Distribution

Usually expressed as the probability of successes, in this thesis it is expressed as the probability of failures.

Let:

$X$ (an integer value) be the No. of failures in the sample $X = \{0, 1, \ldots, n\}$

$p$ = success rate

$q$ = 1- $p$ = failure rate

$n$ = sample size

Then the Binomial Distribution, given $n$ and $q$ is expressed by:

$$B(n;q)\ ; \quad P(X = x) = \frac{n!}{x!(n-x)!} q^x (1-q)^{n-x} \qquad \text{C2.2}$$

and P(X = x) is interpreted as the probability of x failure(s) out of n objects

**2.3 Poisson Distribution**

This distribution is derived from a Poisson process which exists if one can observe discrete events in a certain continuous window (e.g. defects per inch <u>or</u> per second <u>or</u> per cubic metre).
The Poisson model [ $Psn(\lambda)$] is derived as the limit of the Binomial as $n \rightarrow \infty$ and $q \rightarrow 0$ and the product $n \,.\, q = \lambda$ being constant, the expectation $E[x] = n \,.\, q$
The Poisson Distribution, given $\lambda$, is expressed by:

$$Psn(\lambda): \quad P\,(X = x) = \frac{e^{-\lambda}.\lambda^{x}}{x!} \qquad \text{C2.3}$$

Example: Say we have observed, in the Catastrophic category, 1 event over the period of 930 Unit Years, then $q$ = 1/ 930 = 0.00107, if in a particular year we have a fleet of 64 GU's then E[x] = 64 . $q$ = 0.0688, therefore by using Eq C2.3, the chance of having zero Catastrophic events in a particular year with 64 GU's is:
$Psn(0.068)$: $P\,(X = 0) = exp(0\text{-}0.0688)\ .\ 1\ /\ 1 = 0.9335$

**2.4 Relationship between the Poisson and Binomial Distributions**

Let $X_1$ and $X_2$ be two iid $Psn(\lambda_i)$ rv's, and let $Y = X_1 + X_2$
From C1.0, we get:
$f_{X|Y}\,(x|y) = P(X_1 = x_1 \,,\, Y = y) \,/\, P(Y = y)$

$= P(X_1 = x_1 \,,\, X_2 = y - x_1) \,/\, P(Y = y)$

From C1.1, we get:

$f_{X|Y}(x|y) = [P(X_1 = x_1). P(X_2 = y - x_1)] / P(Y = y)$

$$= \frac{\dfrac{\lambda_1^{x_1} e^{-\lambda_1}}{x_1!} \cdot \dfrac{\lambda_2^{y-x_1} e^{-\lambda_2}}{(y-x_1)!}}{\dfrac{(\lambda_1+\lambda_2)^y e^{-\lambda_1-\lambda_2}}{y!}}$$

$$= \frac{y!}{x_1!(y-x_1)!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{y-x_1}$$

$$= \frac{y!}{x_1!(y-x_1)!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1} \left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{y-x_1} \qquad \text{C2.4}$$

Therefore, given the sum $Y$ of $X_1$ and $X_2$ being two iid *Psn* $(\lambda_i)$ rv's, the conditional distribution of $X_i$ is binomial (see Eq1.2 in main body of thesis) with parameters $n = y$ and $q = (\lambda_i) / (\lambda_1 + \lambda_2)$

Expanding for $Y = \sum_{i=1}^{n} X_i$

$f_{X|Y}(x|y)$

$$= \frac{\left(\sum_{i=1}^{n} X_i\right)!}{x_1!(\sum_{i=1}^{n} X_i - x_1)!} \left(\frac{\lambda_1}{\sum_{i=1}^{n}\lambda_i}\right)^{x_1} \left(1-\frac{\lambda_1}{\sum_{i=1}^{n}\lambda_i}\right)^{\sum_{i=1}^{n} X_i - x_1}$$

C2.5

Therefore, given $Y = \sum_{i=1}^{n} X_i$ of all $X_i$ being *n* iid *Psn* $(\lambda_i)$ rv's, the conditional distribution of $X_i$ is binomial with parameters $n = y$ and $q = (\lambda_i) / \sum_{i=1}^{n} \lambda_i$

Note here that the joint pdf may be constructed by the product of the conditional and the marginal.

$$\therefore \quad f_{X,Y}(x,y) = f_{X|Y}(x|y) \, . \, f_Y(y)$$

## 2.5 Quantiles

From A1.1,
For $q \in (0\,;1)$, the $q$-th quantile of the distribution of a rv $X$ is:
$F^{-1}(q) = inf\{x \in \mathfrak{R} : F_X(x) \geq q\}$

<u>Exponential Distribution (Gumbel domain)</u>
Let $X$ be exponentially distributed with pdf
$f_X(x) = \lambda.exp(-\lambda x)$, for $\lambda > 0$ and $x \geq 0$.
Hence, its CDF (of $X$) is:
$F_X(x) = 1 - exp(-\lambda x)$, as given (see Appendix A Eq. A1.0).
Consequently, $Q_x(p) = inf\{x: P(X \leq x) \geq p\}$, from A1.1 above.
Thus, $p$-th Quantile, for any given $0 < p < 1$, is the root of $F_X(x) = p$ , i.e.
$1 - exp(-\lambda x) = p$ or $exp(-\lambda x) = 1 - p$,
hence $-\lambda x = ln\,(1 - p)$
Therefore, $Q_x(p) = -\,[\{ln\,(1 - p)]\,/\,\lambda$
Since $\lambda$ is a constant,
for a **<u>Q-Q plot</u>**,
use { - [{$ln\,(1 - p)$] ; $x$ } as { Horiz. ; Vert. } co-ordinates

<u>Weibull Distribution (Weibull domain)</u>
Let $X$, essentially positive, be Weibull distributed with the following CDF
$F_X(x) = 1 - exp(\,-\lambda x^{\tau}\,)$ , $x > 0$
Therefore: $1 - exp(\,-\lambda x^{\tau}\,) = p$ or $exp(\,-\lambda x^{\tau}\,) = 1 - p$
Hence: $(-\lambda x^{\tau}) = ln\,(1 - p)$,
Therefore $\tau.ln\,(x) = ln\,[(-\,1/\lambda)\,ln\,(1 - p)]$

$ln\ [Q_x(p)] = 1/\ \tau\ \ ln\ [(-\ 1/\lambda)\ ln\ (1 - p)]$
Since $\lambda\ \ \tau$ are constants,
for a **Q-Q plot**, use
$\{\ ln\ [-\ ln\ (1 - p)];\ ln\ (\ x\ )\}$ as { Horiz.;Vert. } co-ordinates

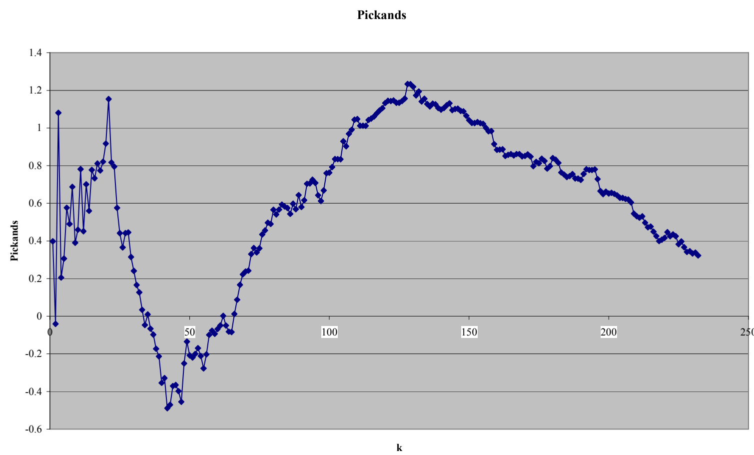
Pareto Distribution (Fréchet domain)
Let $X$, essentially positive, be Pareto distributed, then
$F_X(x) = 1 - x^{(-1/\gamma)},\ x > 0$
Therefore: $1 - x^{(-1/\gamma)} = p$ or $x^{(-1/\gamma)} = 1 - p$
Hence: $(-1/\gamma)\ ln\ (x) = ln\ (1 - p)$,
Therefore $ln\ (x) = -\ \gamma\ ln\ (1 - p)$
$ln\ [Q_x(p)] = -\ \gamma\ ln\ (1 - p)$ C2.6
Since $\gamma$ is a constant,
for a **Q-Q plot**, use
$\{\ -\ [\{ln\ (1 - p)]\ ;\ ln\ (\ x\ )\}$ as { Horiz. ; Vert. } co-ordinates

**2.6 Pickands Estimator**

This estimator for the EVI is given by:

$$\mathrm{EVI}_k = \hat{\gamma}_{P,k} = \frac{1}{Ln(2)} Ln\left\{ \frac{X_{n-k+1} - X_{n-2k+1}}{X_{n-2k+1} - X_{n-4k+1}} \right\},\ k_{\max} \text{ is } n/4$$

Utilising the model above on the GUs GWh losses, the results of $\mathrm{EVI}_k$ , plotted against $k$, are shown in the graph here below:




The large volatility of the EVI is evident, and by the time that the limit of observations was reached, there was no stability in the indicator. A possibility is that it may have reached it at a value of between 0.2 and 0.4.

In [2] Beirlant, it is mentioned that more efficient changes in this estimator are being proposed (such as Segers[2004]).

### 2.7 Gamma Series Functions

It is possible to estimate the Gamma function near a point by using a series expansion at that point.
To be able to do that one needs to invoke some important findings.

#### 2.7.1 Euler's Function

In 1730, Euler introduced the following definition:

$$\Gamma(x) = \int_0^1 [-Ln(t)]^{x-1} dt \quad \textbf{or} \quad \Gamma(x) = \int_0^\infty t^{x-1}.e^{-t} dt$$

by substituting $u = -Ln(t)$ in the 1st function.

#### 2.7.2 The Weierstrass Formula

The relation $p^x = e^{x ln(p)} = e^{x[ln(p)-1-1/2-\ldots-1/p]} \, . \, e^{x+x/2+\ldots+x/p}$, involves:

$$\Gamma_p(x) = \frac{1}{x}.\frac{1}{x+1}.\frac{2}{x+2}\cdots\frac{p}{x+p}\, p^x$$

$$\Gamma_p(x) = \frac{1}{x}.\frac{e^x}{x+1}.\frac{e^{x/2}}{x+2}\cdots\frac{e^{x/p}}{1+\frac{x}{p}}\ [e^{x\{Ln(p)-1-\frac{1}{2}-\ldots-\frac{1}{p}\}}]$$

Euler's constant (also called the Euler-Mascheroni constant) is defined as:

$$\varepsilon = \underset{p\to\infty}{Lim}\left(1+\frac{1}{2}+...+\frac{1}{p}-Ln(p)\right) = 0.5772156649\ldots$$

and the Weierstrass formula for the Gamma Function is (from Weierstrass theorem):
$\forall\ x \in \Re^{+}\ ,\ \notin\ \{0,\Im^{-}\}$ (i.e. all real numbers except 0 and negative integers)

$$\frac{1}{\Gamma(x)} = x.e^{\varepsilon.x}\prod_{p=1}^{\infty}\left(1+\frac{x}{p}\right)e^{-\frac{x}{p}} \qquad \text{C2.7}$$

by taking the logs of C2.7, the following is obtained:

$$-Ln[\Gamma(x)] = Ln(x)+\varepsilon.x+\sum_{p=1}^{\infty}\left(Ln\,(1+\frac{x}{p})-\frac{x}{p}\right), \text{ or}$$

$$Ln[\Gamma(x)] = -Ln(x)-\varepsilon.x-\sum_{p=1}^{\infty}\left(Ln\,(1+\frac{x}{p})-\frac{x}{p}\right) \qquad \text{C2.8}$$

by differentiating C2.8

$$\frac{d}{dx}\{Ln[\Gamma(x)]\} = -\frac{1}{x}-\varepsilon+\sum_{p=1}^{\infty}\left(\frac{1}{p}-\frac{1}{x+p}\right)$$

$$= -\varepsilon+\sum_{p=1}^{\infty}\left(\frac{1}{p}-\frac{1}{x+p-1}\right)$$

for x $\neq$ 0, -1,-2,….

$$= -\varepsilon + \sum_{p=1}^{\infty}\left(\frac{x-1}{p(x+p-1)}\right)$$

for x $\neq$ 0, -1,-2,….

$$\Psi(x) = \frac{d}{dx}\{Ln[\Gamma(x)]\} = -\varepsilon + \sum_{p=1}^{\infty}\left(\frac{x-1}{p(x+p-1)}\right) \qquad \text{C2.9}$$

for x $\neq$ 0, -1,-2,….

The equation C2.9 is defined as the ***psi*** or ***digamma function*** for any non-zero or negative integer
One may continue differentiating to obtain ***Polygamma*** functions:

$$\Psi^{'}(x) = \frac{d}{dx}\{\Psi(x)\} = \sum_{p=1}^{\infty}\left(\frac{1}{(x+p-1)^2}\right)$$

$$\Psi^{''}(x) = \frac{d}{dx}\{\Psi^{'}(x)\} = -\sum_{p=1}^{\infty}\left(\frac{2}{(x+p-1)^3}\right)$$

or, in general form, the ***Polygamma*** can be expressed as:

$$\Psi_n(x) = \sum_{p=1}^{\infty}\left(\frac{(-1)^{n+1}n!}{(x+p-1)^{n+1}}\right) \qquad \text{C2.10}$$

this means that equation C2.9 is usually expressed in the form:

$$\Psi_0(x) = \Psi(x) = -\varepsilon + \sum_{p=1}^{\infty}\left(\frac{x-1}{p(x+p-1)}\right) \qquad \text{C2.11}$$

$\Psi_0(x) = \Psi(x) = -\varepsilon + \sum_{k=1}^{x-1}\left(\frac{1}{k}\right) \quad = -\varepsilon + H_{x-1}$ , where $H_x$ is a harmonic number

For a continuous *x*, equation C2.11 can be expressed as

$\Psi_0(x) = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1-e^{-t}}\right) dt$, or also as

$$\Psi_0(x+1) = -\varepsilon + \int_0^1 \left(\frac{1-t^x}{1-t}\right) dt$$

$\Psi_1(x)$ is defined as the ***trigamma function***.

Special cases for the ***digamma*** and ***trigamma*** :

$$\Psi_0(1) = \Psi(1) = -\varepsilon$$

$$\Psi_0(\frac{1}{2}) = \Psi(\frac{1}{2}) = -\varepsilon - 2.Ln(2)$$

$$\Psi_0(\frac{1}{2} + x) = \Psi_0(\frac{1}{2}) + 2.\sum_{k=1}^{x}\left(\frac{1}{2.k-1}\right) \quad = -\varepsilon + H_{x-\frac{1}{2}}$$

$$\Psi_1(1) = \frac{\pi^2}{6} \ ,$$

***NB* Other Theorems/Corollary of interest**:

Legendre's Theorem:

$$\Gamma(x).\Gamma(x+\frac{1}{2}) = \sqrt{\pi}\,\Gamma(2x)$$

<u>Gauss</u>' Theorem:

$$\Gamma(x).\Gamma(x+\frac{1}{n}).\Gamma(x+\frac{2}{n})....\Gamma(x+\frac{n-1}{n}) = (2\pi)^{\frac{(n-1)}{2}}.n^{\frac{1}{2}-nx}.\Gamma(nx)$$

<u>Euler</u>'s Corollary (by setting $x = 1/n$ in Gauss'):

$$\Gamma(\frac{1}{n}).\Gamma(\frac{2}{n})....\Gamma(\frac{n-1}{n}) = \frac{(2\pi)^{\frac{(n-1)}{2}}}{\sqrt{n}}$$

## 2.8 Entropy

Entropy referred here is Statistical Entropy. It originates from the works in thermodynamics by L Boltzmann, an Austrian phycicist who was researching entropy using probability theory. In later works, the definition of Statistical Entropy was defined in terms of a macrostate variable to be considered as an expression of a function of microstate variables, or mathematically:

$$\mathcal{H}(\boldsymbol{p}) = -\lambda\sum_{i=1}^{k}\left[p_i \, .Log(p_i)\right] \qquad \text{C2.12}$$

Where, $\boldsymbol{p} = \{p_1, p_2, \ldots, p_k\}$ is the probability density vector of element I; i.e. $p_i$ is the probability that i will be in a given microstate and all $p_i$'s are evaluated for the same macrostate; $\lambda$ is a constant (Boltzmann constant in thermodynamics), but may be set arbitrarily to 1 without affecting the concept of entropy in Statistical Mechanics. Since all $p_i$'s are probabilities between 0 and 1, this implies that by taking the Log, the results would be negative, to counteract this effect, a negative sign is placed in front of the equation. **The Log is taken to the base 2 to reflect the presence/absence of the element i in the microstate**.

If all $p_i$'s are equal (e.g. distributed Uniformly), then:

$\mathcal{H}(\boldsymbol{p}) = -\lambda \sum_{i=1}^{k} \left[ p_i \, . Log(p_i) \right]$ = -$\lambda$ *{k.p Log(p)}*, since all

***p***'s are equal, *k.p* =1 and *p* =1/*k* therefore the entropy equation reduces to:

$\mathcal{H}(p) = \lambda\{Log(k)\}$

Hence, under this condition, the relevant factor becomes the number of i's (states in microstate).

The larger the Entropy, the more unpredictable the outcome.

Although not utilised in the thesis, the Kullback-Leibler (KL) information-theoretic measure is worth mentioning. The KL is a value of similarity between a statistical model and a true distribution. In some cases, the statistical model would be the observed distribution while the true one may be generated (say by a simulating process). The KL value is essentially positive and zero when the 2 distributions coincide, but caution needs to be exercised as it is not a metric (it does not satisfy the triangle of inequality and is not symmetric; it is a convex function of the "true" distribution).

The KL may be thought of as the relative entropy of one distribution ***p*** with another distribution ***q***. One may think of the KL (using $Log_2$) as the observed average number of failures by maintaining events from a distribution ***p*** with a maintenance plan based on the distribution ***q***.

Since the KL is defined as the difference between an approximating entropy and the source entropy (with the latter always smaller), as an experiment approaches, through experience, the distribution given, the KL decreases towards zero.
However since the KL in itself might be misleading (if source entropy is 1 and KL is 0.1, then the latter is small; if source entropy is 0.1 and KL is 0.1, then the latter is relatively large). Hence the Relative Kullback is generally utilised as: 100 (KL / $\mathcal{H}$), where $\mathcal{H}$ is the source entropy (referred to as the arc source entropy, this can be interpreted in our case as the minimum average number of fail/succeed tests which must be performed to determine a maintenance plan).
An interesting topic but outside the realm of this thesis.

## 2.9 Dirichlet Process

### 2.9.1 Distributional Properties

A few definitions need mentioning, namely the Multinomial distribution one and the Dirichlet one.
The Multinomial distribution is a discrete distribution which gives the probability of choosing a collection of *m* items from a set of *n* items with repetitions and the probabilities of each choice given by $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\}$, these probabilities form the parameters of the Multinomial distribution: $p_{\boldsymbol{X}}(\boldsymbol{x}) = n! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!}$
The Dirichlet distribution is the conjugate prior of the parameters of the Multinomial distribution. The pdf of the

Dirichlet for variables $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\}$, with parameters $\boldsymbol{u} = \{u_1, u_2, \ldots, u_n\}$, is defined as:

$$\mathrm{Dir}\,(\boldsymbol{p};\boldsymbol{u}) = \frac{1}{Z(\boldsymbol{u})}\prod_{i=1}^{n} p_i^{u_i-1} \qquad \text{C2.13}$$

with $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\} \geq \mathbf{0}$, $\sum_{i=1}^{n} p_i = 1$

and $\boldsymbol{u} = \{u_1, u_2, \ldots, u_n\} > \mathbf{0}$,

where $Z(\boldsymbol{u}) = \dfrac{\prod_{i=1}^{n}\Gamma(u_i)}{\Gamma(\sum_{i=1}^{n} u_i)}$

The parameters $\boldsymbol{u}$ can be interpreted as "prior observation counts" for events governed by $\boldsymbol{p}$.

Let $u_0 = \sum_{i=1}^{n} u_i$, then the mean and variance of equation C2.13 are:

$\mathrm{E}[p_i] = \dfrac{u_i}{u_0}$ and $\mathrm{Var}[p_i] = \dfrac{u_i(u_0 - u_i)}{u_0^2(u_0+1)}$

If $u_\mathrm{i} \to \mathbf{0}$ , then Dir $(\boldsymbol{p};\boldsymbol{u})$ becomes non-informative and $p_i$'s stay the same if all $u_\mathrm{i}$'s are scaled with the same multiplicative constant, but $\mathrm{Var}[p_i]$ will get smaller as $u_\mathrm{i}$'s increase.

### 2.9.2 Negative Differential Entropy

Define the **Negative Differential Entropy** (NDE) as $\mathrm{E}[Log\, p(\boldsymbol{p})]$ and the expectation $\mathrm{E}[Log\, p_i]$, by reducing this expectation over a two dimensional space of the Dirichlet distribution represented by:

$(p;1-p) \sim \mathrm{Dir}(u_i\,; u_0 - u_i)$
the expectation is then given by:

$$\mathrm{E}[Log\, p_i] = \int_0^1 \frac{\Gamma(u_0)}{\Gamma(u_i).\Gamma(u_0-u_i)}\, p^{u_i-1}(1-p)^{u_0-u_i-1}\, Log\ p\ \ dp$$

Using the digamma functions, [20] Honkela evaluates the expectation as:

$$\mathrm{E}[\mathrm{Log}\, p_i] = \Psi_0(u_i) - \Psi_0(u_0)$$

Honkela utilises the above equation to evaluate the NDE as:

$$\mathrm{E}[\mathrm{Log}\, p(\boldsymbol{p})] = \mathrm{E}\left[ Log\, \frac{1}{Z(\boldsymbol{u})} \prod_{i=1}^{n} p_i^{u_i-1} \right]$$

$$= \left[ -Log\, Z(\boldsymbol{u}) + \sum_{i=1}^{n} (u_i - 1)\ \mathrm{E}[Log\ p_i] \right]$$

$$= \quad -Log\, Z(\boldsymbol{u}) + \sum_{i=1}^{n} (u_i - 1)\ [\Psi_0(u_i) - \Psi_0(u_0)] \qquad \text{C2.14}$$

**2.9.3 Comparative Formulae**
**Comparison between Honkela equation C2.14 and equation (2) in [11] De Waal, et al**

From equation C2.12

$\mathcal{H}(\boldsymbol{p}) = -\lambda \sum_{i=1}^{n} \left[ p_i \, .Log(p_i) \right]$, setting $\lambda = 1$, we get

$\mathcal{H}(\boldsymbol{p}) = -\sum_{i=1}^{n} \left[ p_i \, .Log(p_i) \right] = -\mathrm{E}[\mathrm{Log}\, p_i]$ , $\sum_{i=1}^{n} p_i = 1$

$\therefore \; \mathrm{E}[\mathrm{Log}\, p_i] \;=\; -\sum_{i=1}^{n} \left[ p_i \, .Log(p_i) \right]$,

Since $p_i \sim$ Dir $(\boldsymbol{p;u})$ ,

$$\mathcal{H}(\boldsymbol{p}) = \mathrm{E}[\mathrm{Log}\, p_i] = \mathrm{E}\left[ Log\left\{ \frac{1}{Z(\boldsymbol{u})} \prod_{i=1}^{n} p_i^{u_i - 1} \right\} \right]$$

$$= \mathrm{E}\left[ Log\left\{ \frac{\Gamma\left(\sum_{i=1}^{n} u_i\right)}{\prod_{i=1}^{n} \Gamma(u_i)} \prod_{i=1}^{n} p_i^{u_i - 1} \right\} \right]$$

$$\mathcal{H}(\boldsymbol{p}) = -Log\left\{ \frac{\prod_{i=1}^{n} \Gamma(u_i)}{\Gamma\left(\sum_{i=1}^{n} u_i\right)} \right\} + \sum_{i=1}^{n} (u_i - 1)\ \mathrm{E}[Log\ p_i] \quad ,$$

but $u_0 = \sum_{i=1}^{n} u_i$

$$\therefore\ \mathcal{H}(\boldsymbol{p}) = \quad Log\,\Gamma(u_0) - \sum_{i=1}^{n} Log\,\Gamma(u_i) + \sum_{i=1}^{n}(u_i - 1)\ \ \mathrm{E}[Log\ p_i]$$

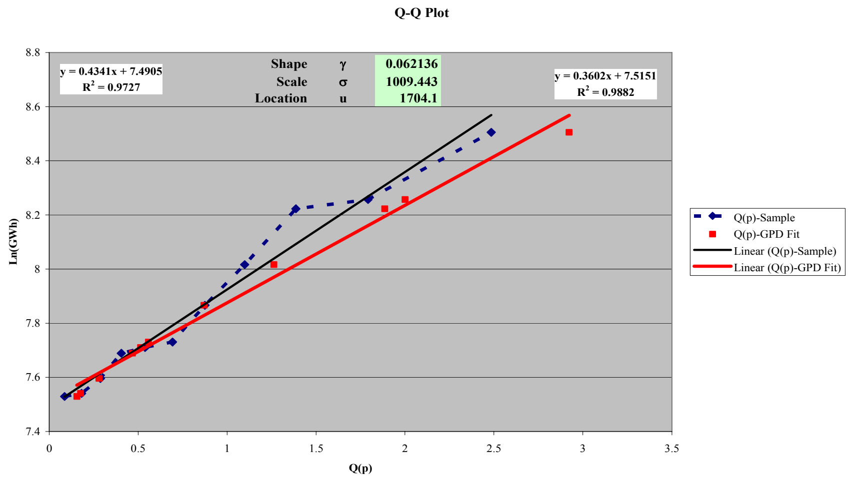
$$\therefore\ \mathcal{H}(\boldsymbol{p}) =$$

$$Log\,\Gamma(u_0) - \sum_{i=1}^{n} Log\,\Gamma(u_i) + \sum_{i=1}^{n}(u_i - 1)\ \ [\Psi_0(u_i) - \Psi_0(u_0)] \qquad \text{C2.15}$$

Which is similar to (2) in [11] De Waal et al, with $u_i = \nu_i$ and $p_i = y_i$ and $n = k+1$

Equations C2.14 and C2.15 are equivalent

## 2.10 Wobbles




By original 3-LM with a threshold of 1704 GWh (although not a good threshold choice, at k=8) : we can see the wobble (blue dotted line).The issue here would be to establish if the wobbles are important underlying distributional phenomena, and not manufactures of the sample, [19] Hernandez-Campos.

# 3 Bayesian Methods

## 3.1 Fisher Information Matrix (FIM)

To arrive to the FIM, it is necessary to define the ***Score Function*** ($V$). $V$ is the partial derivative with respect to a parameter $\theta$ of the Likelihood function.
Let $X$ be a r.v. and its corresponding Likelihood be $L(X\,|\,\theta)$, then the score function is:

$$V = \frac{\partial}{\partial\theta} Ln[L(X|\theta)] = \frac{1}{L(X|\theta)} \cdot \frac{\partial}{\partial\theta} L(X|\theta) \qquad \text{C3.1}$$

$V$ is a function of $\theta$ and $X$ and is a sufficient statistic for $\theta$.

***Sufficiency criterion***: If $T(X)$ is sufficient for $\theta$ then $f_X(X|\theta) = g\{T(X),\theta\}.\ h(X)$ for some functions $g$ and $h$.

Also, E[$V$] = 0
The variance of V is defined as the Fisher Information and is denoted by I($\theta$). Therefore:

$$\text{Var }[V] = \text{I}(\theta) = \text{E}\left[\left\{\frac{\partial}{\partial\theta} Ln[L(X|\theta)]\right\}^2 \,\Big|\, \theta\right] \qquad \text{C3.2}$$

It is worth noting that $0 < \text{I}(\theta) < \infty$ and I($\theta$) is not a function of a particular observation.
**Hence the I($\theta$) is the amount of information that an observed r.v. *X* carries about a parameter $\theta$, upon which, *L*($\theta$;*X*) i.e. its Likelihood, depends**.
If $L(X\,|\,\theta) = f_X(X|\theta)$ and the condition

$\int \frac{\partial^2}{\partial \theta^2} f_X(X|\theta)\, dx = 0$ is met, then C3.2 may be expressed as:

$$\mathrm{I}(\theta) = -\,\mathrm{E}\left[\frac{\partial^2}{\partial \theta^2} Ln[f_X(X|\theta)]\right] \qquad \text{C3.3}$$

Since this is the 2nd partial derivative of the *Ln* of *f* with respect to $\theta$, $\mathrm{I}(\theta)$ may be interpreted as measure of the spread near the maximum likelihood estimate of $\theta$ (even if the sign is negative). This kind of information became known as the Fisher Information Metric is a type of Riemannian metric on the space of probabilities.

***Properties:*** If two r.v.'s, *X* and *Y* are independent, then:

$\mathrm{I}_{X;Y}(\theta) = \mathrm{I}_X(\theta) + \mathrm{I}_Y(\theta)$

Interpreted, this means that provided two r.v.'s are independent, then the variance of their sum, is the sum of their variances.

From C3.3 and the sufficiency criterion as well as the fact that $h(X)$ is independent of $\theta$, we get:

$\frac{\partial^2}{\partial \theta^2} Ln[f_X(X|\theta)] = \frac{\partial^2}{\partial \theta^2} Ln[\mathrm{g}\{T(X), \theta\}]$ and if $T = t_X(x)$ is a statistic, then

$\mathrm{I}_T(\theta) \leq \mathrm{I}_X(\theta)$ with $\mathrm{I}_T(\theta) = \mathrm{I}_X(\theta) \Leftrightarrow T$ is a sufficient statistic. The inequality (Cramer-Rao) indicates that the reciprocal of the Fisher Information, is an asymptotic lower bound on the variance of any unbiased estimator for $\theta$.

When $\boldsymbol{\theta}$ is an *m* vector, then the Fisher Information becomes a *m* x *m* Fisher Information Matrix (FIM) with elements:

$$[\mathrm{I}(\theta)]_{i;j} = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} Ln[f_X(X|\,\theta)]\right] \qquad \text{C3.4}$$

$[\mathrm{I}(\theta)]_{i;j}$ is positive definite symmetric defining $\boldsymbol{\theta}$ in the m-dimensional space. Same definition as above (in bold underlined) but catering for multidimensionality.

Let $\boldsymbol{\theta}$ = ($\mu;\sigma$), i.e. a 2 parameter vector, then, from C3.4, $[\mathrm{I}(\theta)]_{i;j}$ is a 2 x 2 FIM with

$$[\mathrm{I}(\theta)]_{i;j} = -\mathrm{E}\begin{bmatrix} \frac{\partial^2}{\partial\theta_1\partial\theta_1} Ln[f_X(X|\,\theta)] & \frac{\partial^2}{\partial\theta_1\partial\theta_2} Ln[f_X(X|\,\theta)] \\ \frac{\partial^2}{\partial\theta_2\partial\theta_1} Ln[f_X(X|\,\theta)] & \frac{\partial^2}{\partial\theta_2\partial\theta_2} Ln[f_X(X|\,\theta)] \end{bmatrix}$$

$$= -\mathrm{E}\begin{bmatrix} \frac{\partial^2}{\partial\mu^2} Ln[f_X(X|\mu;\sigma)] & \frac{\partial^2}{\partial\mu\,\partial\sigma} Ln[f_X(X|\mu;\sigma)] \\ \frac{\partial^2}{\partial\sigma\,\partial\mu} Ln[f_X(X|\mu;\sigma)] & \frac{\partial^2}{\partial\sigma^2} Ln[f_X(X|\mu;\sigma)] \end{bmatrix} \qquad \text{C3.5}$$

**Example 1**: Log Normal case

for a Log Normal r.v. *X*, by transformation *Y=Ln(X), Y ~ N*($\mu;\sigma^2$)

i.e. the pdf is given by: $f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ ,

where $\mu$ and $\sigma^2$ are the log-mean and log-variance. Then the Likelihood function is: $f_Y(y|\,\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$

$$f_Y(y|\theta) = \frac{1}{\left(\sqrt{2\pi\sigma}\right)^n} e^{-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}} ,$$

and hence the Log-Likelihood is:

$$Ln\{f_Y(y|\theta)\} = -\frac{n}{2} Ln\{2\pi\sigma^2\} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$
$$= -\frac{1}{2} Ln\{\sigma^2\} - \frac{1}{2\sigma^2}(y-\mu)^2 + \text{Constant}$$

using partial derivatives on the Log-Likelihood with respect to $\mu$ and $\sigma^2$, we get:

$$\frac{\partial}{\partial\mu} Ln\{f_Y(y|\theta)\} = \frac{1}{\sigma^2}(y-\mu) \qquad \text{and}$$
$$\frac{\partial}{\partial\sigma} Ln\{f_Y(y|\theta)\} = -\frac{1}{\sigma} + \frac{1}{\sigma^3}(y-\mu)^2$$

$\therefore$ using matrix C3.5 we get as the elements of $[I(\theta)]_{i;j}$:

$$\frac{\partial^2}{\partial\mu^2} Ln\{f_Y(y|\theta)\} = -\frac{1}{\sigma^2} \qquad \text{and}$$
$$\frac{\partial^2}{\partial\mu\partial\sigma} Ln\{f_Y(y|\theta)\} = -\frac{2}{\sigma^3}(y-\mu)$$
$$\frac{\partial^2}{\partial\sigma\partial\mu} Ln\{f_Y(y|\theta)\} = -\frac{2}{\sigma^3}(y-\mu) \qquad \text{and}$$
$$\frac{\partial^2}{\partial\sigma^2} Ln\{f_Y(y|\theta)\} = \frac{1}{2\sigma^2} - \frac{3}{\sigma^4}(y-\mu)^2$$

Since $[I(\theta)]_{i;j}$ is an expectation (see C3.5) and since $E[Y]=\mu$, $E[y-\mu]=0$; also, $E[(y-\mu)^2]=\sigma^2$ , we can reduce the four equations above to:

$$[I(\theta)]_{i;j} = -\begin{bmatrix} -\frac{1}{\sigma^2} & 0 \\ 0 & -\frac{2}{\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix} \qquad \text{C3.6}$$

Det $[I(\theta)]_{i;j} = \dfrac{2}{\sigma^4} - 0$

This implies that the reference prior should be:

$$\pi\,(\mu;\sigma^2 \mid y) \propto \left(\frac{1}{\sigma^4}\right)^{\frac{1}{2}} = \frac{1}{\sigma^2}$$

In concordance with the usual prior for a $N(\mu;\sigma^2)$ which is $\pi\,(\mu;\sigma^2 \mid x) \propto \left(\dfrac{1}{\sigma^2}\right)$

**<u>Example 2</u>**: The GPD case

Let Y ~ GPD $(\gamma;\sigma)$ where $Y = X - \boldsymbol{t}$ of a r.v. X over a threshold $\boldsymbol{t}$ and assume $\gamma \neq 0$, then:

$$F_Y(y) = 1 - \left[1 - \frac{\gamma}{\sigma}y\right]^{\frac{1}{\gamma}} \quad ,\forall y > 0$$

$$\therefore f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{\gamma}\frac{\gamma}{\sigma}\left[1 - \frac{\gamma}{\sigma}y\right]^{\frac{1}{\gamma}-1} = \frac{1}{\sigma}\left[1 - \frac{\gamma}{\sigma}y\right]^{\frac{1}{\gamma}-1} \quad ,\forall y > 0$$

Likelihood:

$$f_Y(y\,|\,\theta) = \prod_{i=1}^{n} f_Y(y) = \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^{n}\left(1-\frac{\gamma\, y_i}{\sigma}\right)^{\frac{1}{\gamma}-1}$$

Log-Likelihood:

$$Ln\{f_Y(y\,|\,\theta)\} = -\,n\ln\sigma + \left(\frac{1}{\gamma}-1\right)\sum_{i=1}^{n}\ln\left(1-\frac{\gamma\, y_i}{\sigma}\right) \quad \text{C3.7}$$

The "trick" here is in the reduction of the equations into a form of $\left[1-\frac{\gamma}{\sigma}y\right]^r \quad ,\forall\,\gamma<\frac{1}{2}$; the reason being that one may then use the properties of expectations, as developed by Davison AC, see [41] van Noortwijk, when developing the Fisher Information Matrix .

$$\sum_{i=1}^{n} y_i = \bar{y}.\text{constant} = y \quad ,\text{let } y = y.\frac{\gamma}{\sigma}.\frac{\sigma}{\gamma}+\frac{\sigma}{\gamma}-\frac{\sigma}{\gamma}$$

$$\therefore\, y = \frac{\sigma}{\gamma}\left[1-\left\{1-\frac{\gamma\, y}{\sigma}\right\}\right] \quad \text{C3.8}$$

Taking 1st partial derivative of the Log-Likelihood w.r.t. $\gamma$:

$$\frac{\partial}{\partial\gamma} Ln\{f_Y(y\,|\,\theta)\} =$$

$$-\left(\frac{1}{\gamma^2}\right)\ln\left(1-\frac{\gamma\, y}{\sigma}\right)+\left(\frac{1}{\gamma}-1\right)\left(\frac{-\,y}{\sigma}\right)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}+\text{constant}$$

Using C3.8 in the equation above, we get:

$$\frac{\partial}{\partial\gamma} Ln\{f_Y(y\mid\theta)\} =$$

$$-\left(\frac{1}{\gamma^2}\right)\ln\left(1-\frac{\gamma\, y}{\sigma}\right)+\left(\frac{1}{\gamma}-1\right)\left(\frac{-1}{\sigma}\right)\frac{\sigma}{\gamma}\left[1-\left\{1-\frac{\gamma\, y}{\sigma}\right\}\right]\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}$$

$=$

$$-\left(\frac{1}{\gamma^2}\right)\ln\left(1-\frac{\gamma\, y}{\sigma}\right)+\left(\frac{1}{\gamma}-1\right)\left[\frac{-1}{\gamma}+\frac{1}{\gamma}\left\{1-\frac{\gamma\, y}{\sigma}\right\}\right]\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}$$

$\therefore$

$$\frac{\partial}{\partial\gamma} Ln\{f_Y(y\mid\theta)\} =$$

$$-\left(\frac{1}{\gamma^2}\right)\ln\left(1-\frac{\gamma\, y}{\sigma}\right)+\left(\frac{1}{\gamma}-1\right)\left(\frac{1}{\gamma}\right)\left[1-\left\{1-\frac{\gamma\, y}{\sigma}\right\}^{-1}\right] \quad \text{C3.9}$$

Taking 1st partial derivative of the Log-Likelihood (C3.7) w.r.t. $\sigma$:

$$\frac{\partial}{\partial\sigma} Ln\{f_Y(y\mid\theta)\} = -\left(\frac{1}{\sigma}\right)+\left(\frac{1}{\gamma}-1\right)\left(\frac{\gamma\, y}{\sigma^2}\right)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1} + \text{constant} \quad \text{C3.10}$$

Using C3.8 in the equation above, and substituting in C3.10 we get:

$$\frac{\partial}{\partial\sigma} Ln\{f_Y(y\mid\theta)\} =$$

$$-\left(\frac{1}{\sigma}\right)+\left(\frac{1}{\gamma}-1\right)\left(\frac{\gamma}{\sigma^2}\right)\left(\frac{\sigma}{\gamma}\right)\left[1-\left\{1-\frac{\gamma\, y}{\sigma}\right\}\right]\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1} \quad \text{C3.11}$$

Let $p = (\frac{1}{\gamma}-1)$ and $k = (1-\frac{\gamma\, y}{\sigma})$ and substitute in C3.11, we get:

$$\frac{\partial}{\partial\sigma}Ln\{f_Y(y\mid\theta)\} = -\left(\frac{1}{\sigma}\right)+\left(p\right)\left(\frac{\gamma}{\sigma^2}\right)\left(\frac{\sigma}{\gamma}\right)\left[1-\{k\}\right]\left(k\right)^{-1}$$

$$=-\left(\frac{1}{\sigma}\right)+\left(p\right)\left(\frac{\gamma}{\sigma^2}\right)\left(\frac{\sigma}{\gamma}\right)\left[-\frac{\{k\}p}{p}+1\right]\left(k\right)^{-1}$$

$$=-\left(\frac{1}{\sigma}\right)+\left(p\right)\left(\frac{\gamma}{\sigma^2}\right)\left(\frac{\sigma}{\gamma}\right)\left[\frac{\{k\}(1-\frac{1}{\gamma})}{p}+1\right]\left(k\right)^{-1}$$

$$=-\left(\frac{1}{\sigma}\right)+\left(p\right)\left(\frac{\gamma}{\sigma^2}\right)\left(\frac{\sigma}{\gamma}\right)\left[\frac{k}{p}-\frac{k}{\gamma\, p}+1\right]\left(k\right)^{-1}$$

$$=-\left(\frac{1}{\sigma}\right)+\left(p\right)\left(\frac{\gamma}{\sigma^2}\right)\left(\frac{\sigma}{\gamma}\right)(-\frac{1}{p})\left[-k+\frac{k}{\gamma}-p\right]\left(k\right)^{-1}$$

$$=-\left(\frac{1}{\sigma}\right)-\left(\frac{1}{\sigma}\right)\left[-1+\frac{1}{\gamma}-p(k)^{-1}\right]$$

$$=-\left(\frac{1}{\sigma}\right)+\left(\frac{1}{\sigma}\right)-\frac{1}{\sigma}\frac{1}{\gamma}+\frac{1}{\sigma}p(k)^{-1}$$

$$=-\frac{1}{\sigma}\frac{1}{\gamma}+\frac{1}{\sigma}p(k)^{-1}$$

$$\therefore\ \frac{\partial}{\partial\sigma}Ln\{f_Y(y\mid\theta)\} = -\frac{1}{\sigma}\frac{1}{\gamma}+\frac{1}{\sigma}\left(\frac{1}{\gamma}-1\right)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1} \qquad \text{C3.12}$$

Using equation C3.9 and taking its 2nd partial derivative w.r.t. $\gamma$, letting $k=(1-\frac{\gamma\, y}{\sigma})$ and using equation 3.8, we get:

$$\frac{\partial^2}{\partial \gamma^2} Ln\{f_Y(y \mid \theta)\} =$$

$$\frac{2}{\gamma^3} Ln(k) - \frac{3-\gamma}{\gamma^3} + \frac{1}{\gamma^3} k^{-1} + \frac{1}{\gamma^2}(\frac{1}{\gamma} - 1)k^{-1} + \frac{2}{\gamma^3} k^{-1} - \frac{k^{-1}}{\gamma^2} + \frac{1}{\gamma^2}(\frac{1}{\gamma} - 1)\{-k^{-2}\}$$

$$= \frac{2}{\gamma^3} Ln(k) - \frac{3-\gamma}{\gamma^3} + \frac{4}{\gamma^3} k^{-1} - \frac{2}{\gamma^2} k^{-1} + \frac{1}{\gamma^2}(\frac{1}{\gamma} - 1)\{-k^{-2}\}$$

$$= \frac{2}{\gamma^3} Ln(k) - \frac{3-\gamma}{\gamma^3} + \frac{2(2-\gamma)}{\gamma^3} k^{-1} - (\frac{1-\gamma}{\gamma^3})(k^{-2})$$

$$\therefore \frac{\partial^2}{\partial \gamma^2} Ln\{f_Y(y \mid \theta)\} =$$

$$\frac{2}{\gamma^3} Ln(1 - \frac{\gamma y}{\sigma}) - \frac{3-\gamma}{\gamma^3} + \frac{2(2-\gamma)}{\gamma^3}\left(1 - \frac{\gamma y}{\sigma}\right)^{-1} - (\frac{1-\gamma}{\gamma^3})\left(1 - \frac{\gamma y}{\sigma}\right)^{-2} \qquad \text{C3.13}$$

Using equation C3.9 and taking its 2nd partial derivative w.r.t. $\sigma$, letting $k = (1 - \frac{\gamma y}{\sigma})$ and using equation 3.8, we get:

$$\frac{\partial^2}{\partial \boldsymbol{\sigma}^2} Ln\{f_Y(y \mid \boldsymbol{\theta})\} = \frac{1}{\sigma^2}\frac{1}{\gamma} - \frac{1}{\sigma^2}(\frac{1}{\gamma} - 1)\left(1 - \frac{\gamma y}{\sigma}\right)^{-2} \qquad \text{C3.14}$$

Since the matrix is symmetric we only need to evaluate one element of the opposite of the trace; using equation C3.12 and taking it's the partial derivative w.r.t. $\gamma$, letting $k = (1 - \frac{\gamma y}{\sigma})$ and using equation 3.8, we get:

$$\frac{\partial^2}{\partial\sigma\,\partial\gamma}Ln\{f_Y(y\,|\,\theta)\} =$$

$$\frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{\frac{1}{\sigma}\left(\frac{1}{\gamma}-1\right)\left(\frac{-1}{\sigma}\right)y\left(1-\frac{\gamma\,y}{\sigma}\right)^{-2} + \frac{1}{\sigma}\left(\frac{-1}{\gamma^2}\right)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-1}\right\}$$

$$=$$

$$\frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{\frac{1}{\sigma}\left(\frac{1}{\gamma}-1\right)\left(\frac{-1}{\sigma}\right)\frac{\sigma}{\gamma}\left[1-\left\{1-\frac{\gamma\,y}{\sigma}\right\}\right]\left(1-\frac{\gamma\,y}{\sigma}\right)^{-2} + \frac{1}{\sigma}\left(\frac{-1}{\gamma^2}\right)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-1}\right\}$$

$$=$$

$$\frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{\frac{1}{\sigma}\left(\frac{1}{\gamma}-1\right)\left(\frac{-1}{\gamma}\right)(k)^{-2} - \frac{1}{\sigma}\left(\frac{1}{\gamma}-1\right)\left(\frac{-1}{\gamma}\right)(k)^{-1} + \frac{1}{\sigma}\left(\frac{-1}{\gamma^2}\right)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-1}\right\}$$

$$=\frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)(k)^{-2} + \frac{k^{-1}}{\sigma}\left(\frac{1-\gamma}{\gamma^2}+\frac{1}{\gamma^2}\right)\right\}$$

$$=\frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)(k)^{-2} + \frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)(k)^{-1}\right\}$$

$$\therefore\ \frac{\partial^2}{\partial\sigma\,\partial\gamma}Ln\{f_Y(y\,|\,\theta)\} =$$

$$\frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-2} + \frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-1}\right\} \quad \text{C3.15}$$

The Fisher Information Matrix (FIM) for a GPD is then:

$$[\mathrm{I}(\gamma;\sigma)] = -\mathrm{E}\begin{bmatrix} C3.13 & C3.15 \\ C3.15 & C3.14 \end{bmatrix}$$

$$[\mathrm{I}(\gamma;\sigma)] =$$

$$-\mathrm{E}\begin{bmatrix} \frac{2}{\gamma^3}Ln(1-\frac{\gamma\,y}{\sigma}) - \frac{3-\gamma}{\gamma^3} + \frac{2(2-\gamma)}{\gamma^3}\left(1-\frac{\gamma\,y}{\sigma}\right)^{-1} - (\frac{1-\gamma}{\gamma^3})\left(1-\frac{\gamma\,y}{\sigma}\right)^{-2} & \frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-2} + \frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-1}\right\} \\ \frac{1}{\sigma}\frac{1}{\gamma^2} - \left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-2} + \frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-1}\right\} & \frac{1}{\sigma^2}\frac{1}{\gamma} - \frac{1}{\sigma^2}(\frac{1}{\gamma}-1)\left(1-\frac{\gamma\,y}{\sigma}\right)^{-2} \end{bmatrix}$$

$[\mathrm{I}(\gamma;\sigma)]=$

$$
\mathrm{E}\begin{bmatrix} -\frac{2}{\gamma^3}Ln(1-\frac{\gamma\, y}{\sigma})+\frac{3-\gamma}{\gamma^3}-\frac{2(2-\gamma)}{\gamma^3}\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}+(\frac{1-\gamma}{\gamma^3})\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2} & -\frac{1}{\sigma}\frac{1}{\gamma^2}+\left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}+\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}\right\} \\ -\frac{1}{\sigma}\frac{1}{\gamma^2}+\left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}+\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}\right\} & -\frac{1}{\sigma^2}\frac{1}{\gamma}+\frac{1}{\sigma^2}(\frac{1}{\gamma}-1)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2} \end{bmatrix}
$$

Using the expectation properties, see [41] van Noortwijk, with respect to $y$, i.e.

$E\left[\left(1-\frac{\gamma\, y}{\sigma}\right)^r\right]=\frac{1}{1+r\gamma}$ and $E\left[Ln\left(1-\frac{\gamma\, y}{\sigma}\right)\right]=-\gamma$ and

$E\left[\left(\frac{1}{(\sigma-\lambda y)^2}\right)\right]=\frac{1}{\sigma^2(1-2\gamma)}$

referred to as property I, II and III respectively: P-I, P-II and P-III

and all valid for $\gamma < \frac{1}{2}$.

Hence, taking the expectation of the 1st element of the FIM:

$$E\left[-\frac{2}{\gamma^3}Ln(1-\frac{\gamma\, y}{\sigma})+\frac{3-\gamma}{\gamma^3}-\frac{2(2-\gamma)}{\gamma^3}\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}+(\frac{1-\gamma}{\gamma^3})\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}\right]$$

$=$

$$-\frac{2}{\gamma^3}E\left[Ln(1-\frac{\gamma\, y}{\sigma})\right]+\frac{3-\gamma}{\gamma^3}-\frac{2(2-\gamma)}{\gamma^3}E\left[\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}\right]+(\frac{1-\gamma}{\gamma^3})E\left[\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}\right]$$

Using P-I and P-II above, we get:

$$=-\frac{2}{\gamma^3}(-\gamma)+\frac{3-\gamma}{\gamma^3}-\frac{2(2-\gamma)}{\gamma^3}\frac{1}{(1-\gamma)}+(\frac{1-\gamma}{\gamma^3})\frac{1}{(1-2\gamma)}$$

$$= \frac{2\gamma}{\gamma^3} + \frac{1}{\gamma^3}\left\{\frac{(3-\gamma)(1-\gamma)(1-2\gamma) - 2(2-\gamma)(1-2\gamma) + (1-\gamma)^2}{(1-\gamma)(1-2\gamma)}\right\}$$

$$= \frac{2\gamma}{\gamma^3} + \frac{1}{\gamma^3}\left\{\frac{(-2\gamma + 6\gamma^2 - 2\gamma^3)}{(1-\gamma)(1-2\gamma)}\right\}$$

$$= \frac{1}{\gamma^3}\left\{\frac{2\gamma(1-\gamma)(1-2\gamma) + (-2\gamma + 6\gamma^2 - 2\gamma^3)}{(1-\gamma)(1-2\gamma)}\right\}$$

$$= \frac{1}{\gamma^3}\left\{\frac{2\gamma^3}{(1-\gamma)(1-2\gamma)}\right\}$$

$$\therefore E\left[-\frac{2}{\gamma^3} Ln(1 - \frac{\gamma\, y}{\sigma}) + \frac{3-\gamma}{\gamma^3} - \frac{2(2-\gamma)}{\gamma^3}\left(1 - \frac{\gamma\, y}{\sigma}\right)^{-1} + (\frac{1-\gamma}{\gamma^3})\left(1 - \frac{\gamma\, y}{\sigma}\right)^{-2}\right] = \frac{2}{(1-\gamma)(1-2\gamma)} \qquad \text{C3.16}$$

Similarly, taking the expectation of the 2nd element (diagonally) of the FIM:

$$E\left[-\frac{1}{\sigma^2}\frac{1}{\gamma} + \frac{1}{\sigma^2}(\frac{1}{\gamma} - 1)\left(1 - \frac{\gamma\, y}{\sigma}\right)^{-2}\right]$$

$$= -\frac{1}{\sigma^2}\frac{1}{\gamma} + \frac{1}{\sigma^2}(\frac{1}{\gamma} - 1)\; E\left[\left(1 - \frac{\gamma\, y}{\sigma}\right)^{-2}\right]$$

Using P-I above,

$$= -\frac{1}{\sigma^2}\frac{1}{\gamma} + \frac{1}{\sigma^2}(\frac{1}{\gamma} - 1)\;\frac{1}{(1-2\gamma)}$$

$$= \frac{1}{\sigma^2}\left\{-\frac{1}{\gamma} + (\frac{1}{\gamma} - 1)\;(\frac{1}{(1-2\gamma)})\right\}$$

$$=\frac{1}{\sigma^2}\left\{-\frac{1}{\gamma}+\frac{1}{\gamma(1-2\gamma)}-\frac{1}{(1-2\gamma)}\right\}$$

$$=\frac{1}{\sigma^2}\left\{\frac{-(1-2\gamma)+1-\gamma}{\gamma(1-2\gamma)}\right\}$$

$$\therefore E\left[-\frac{1}{\sigma^2}\frac{1}{\gamma}+\frac{1}{\sigma^2}(\frac{1}{\gamma}-1)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}\right]=\frac{1}{\sigma^2(1-2\gamma)} \qquad \text{C3.17}$$

Again, in a similar way, taking the expectation of the remaining element of the FIM:

$$E\left[-\frac{1}{\sigma}\frac{1}{\gamma^2}+\left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}+\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}\right\}\right]$$

$$=$$

$$-\frac{1}{\sigma}\frac{1}{\gamma^2}+\left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\,E\left[\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}\right]+\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\,E\left[\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}\right]\right\}$$

$$=$$

$$-\frac{1}{\sigma}\frac{1}{\gamma^2}+\left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\frac{1}{(1-2\gamma)}+\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\frac{1}{(1-\gamma)}\right\}$$

$$=-\frac{1}{\sigma\gamma^2}\left\{1+(1-\gamma)\frac{1}{(1-2\gamma)}-(2-\gamma)\frac{1}{(1-\gamma)}\right\}$$

$$=-\frac{1}{\sigma\gamma^2}\left\{\frac{(1-2\gamma)(1-\gamma)+(1-\gamma)^2-(2-\gamma)(1-2\gamma)}{(1-\gamma)(1-2\gamma)}\right\}$$

$$= -\frac{1}{\sigma\gamma^2}\left\{\frac{\gamma^2}{(1-\gamma)(1-2\gamma)}\right\}$$

$$\therefore E\left[-\frac{1}{\sigma}\frac{1}{\gamma^2}+\left\{-\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(1-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-2}+\frac{1}{\sigma}\left(\frac{1}{\gamma^2}\right)(2-\gamma)\left(1-\frac{\gamma\, y}{\sigma}\right)^{-1}\right\}\right]=\frac{-1}{\sigma(1-\gamma)(1-2\gamma)}$$

C3.18

Hence the FIM becomes:

$$[\mathrm{I}(\gamma;\sigma)]=\begin{bmatrix}\dfrac{2}{(1-\gamma)(1-2\gamma)} & -\dfrac{1}{\sigma(1-\gamma)(1-2\gamma)}\\ -\dfrac{1}{\sigma(1-\gamma)(1-2\gamma)} & \dfrac{1}{\sigma^2(1-2\gamma)}\end{bmatrix} \quad \text{C3.19}$$

The determinant of which is:

$$\Delta = \frac{2}{(1-\gamma)(1-2\gamma)}\cdot\frac{1}{\sigma^2(1-2\gamma)}-\left(-\frac{1}{\sigma(1-\gamma)(1-2\gamma)}\right)^2$$

$$= \frac{1}{\sigma^2(1-\gamma)^2(1-2\gamma)^2}\ \left\{2(1-\gamma)-1\right\}$$

$$\therefore \Delta = \frac{1}{\sigma^2(1-\gamma)^2(1-2\gamma)}$$

Therefore, Jeffreys' Prior for a $Y \sim$ GPD $(\gamma;\sigma)$,
where $Y = X - \boldsymbol{t}$ is:

$$J\{GPD(\gamma;\sigma)\} = \sqrt{\Delta} = \frac{1}{\sigma(1-\gamma)\sqrt{(1-2\gamma)}}\ , \qquad \text{C3.20}$$

for $-\frac{1}{2} < \gamma < \frac{1}{2}$

Although the equations of the FIM elements, prior to taking expectations, are different in appearance but not in content, the results of C3.19 and C3.20 are equivalent to the one obtained by [41] van Noortwijk p J4 ([I($\gamma;\sigma$)] in this thesis and [I($\sigma;\gamma$)] in van Noortwijk), quid est demonstrandum.

Inverting the FIM, let $[\mathrm{I}(\theta)]_{i;j} = [\mathrm{I}(\theta)]$:

$$[\mathrm{I}(\gamma;\sigma)]^{-1} = \frac{1}{\Delta}[\mathrm{I}(\gamma;\sigma)]^{\mathrm{T}}$$

$$[\mathrm{I}(\gamma;\sigma)]^{-1} = \begin{bmatrix} \dfrac{2\sigma^2(1-\gamma)^2(1-2\gamma)}{(1-\gamma)(1-2\gamma)} & -\dfrac{\sigma^2(1-\gamma)^2(1-2\gamma)}{\sigma(1-\gamma)(1-2\gamma)} \\ -\dfrac{\sigma^2(1-\gamma)^2(1-2\gamma)}{\sigma(1-\gamma)(1-2\gamma)} & \dfrac{\sigma^2(1-\gamma)^2(1-2\gamma)}{\sigma^2(1-2\gamma)} \end{bmatrix}$$

$$[\mathrm{I}(\gamma;\sigma)]^{-1} = \begin{bmatrix} 2\sigma^2(1-\gamma) & -\sigma(1-\gamma) \\ -\sigma(1-\gamma) & (1-\gamma)^2 \end{bmatrix} \qquad \text{C3.21}$$

or in another form:

$$[\mathrm{I}(\gamma;\sigma)]^{-1} = \sigma^2(1-\gamma)^2 \begin{bmatrix} \dfrac{2}{(1-\gamma)} & -\dfrac{1}{\sigma(1-\gamma)} \\ -\dfrac{1}{\sigma(1-\gamma)} & \dfrac{1}{\sigma^2} \end{bmatrix} \qquad \text{C3.22}$$

### 3.2 The Delta Method (DM)

This method is a general procedure to determine estimates and provides the approximation of standard errors of arbitrary function of normally distributed random variables; it may also be used for the computation of confidence intervals for functions of MLE. This method uses a function that would be too complex to compute and by employing a linear approximation of that complex function, it formulates the variance of the simpler linear function that may be utilised for inference.

Let $\sqrt{n}\,(\hat{\theta}-\theta)\xrightarrow{L} N[\mathbf{0};Var(\hat{\theta})]$ $\quad , n\to\infty$ from MLE theory

Let $g(\theta)$ be a non-linear continuous function of $\theta$, e.g. predicted probabilities from, say, a Log-Normal , a GPD, etc…

Then, using Taylor series as an expansion, we get:

$$g(\hat{\theta}) = g(\tilde{\theta}) + \sum_{i=1}^{k}\frac{\partial g}{\partial \theta_i}(\hat{\theta}_i-\tilde{\theta}_i) + o\left(\left\|\hat{\theta}-\tilde{\theta}\right\|\right) \qquad \text{C3.23}$$

where, $\tilde{\theta}$ **is a value between** $\hat{\theta}$ **and** $\theta$.

Let $g'(\tilde{\theta}) = \dfrac{\partial g}{\partial \theta} = \left[\dfrac{\partial g}{\partial \theta_1};\dfrac{\partial g}{\partial \theta_2};...;\dfrac{\partial g}{\partial \theta_k}\right]^{\mathbf{T}}$, evaluated at $\tilde{\theta}$, then Eq C3.23 becomes:

$$g(\hat{\theta}) = g(\tilde{\theta}) + g'(\tilde{\theta})^{\mathbf{T}}(\hat{\theta}-\tilde{\theta}) + o\left(\left\|\hat{\theta}-\tilde{\theta}\right\|\right)$$

or

$g(\hat{\theta}) - g(\tilde{\theta}) = g'(\tilde{\theta})^{\mathbf{T}}(\hat{\theta} - \tilde{\theta}) + o\left(\left\| \hat{\theta} - \tilde{\theta} \right\|\right)$, representing the MLE of $g(\theta)$
by taking the variance:
$Var[g(\hat{\theta}) - g(\tilde{\theta})] = g'(\tilde{\theta})^{\mathbf{T}} Var[\hat{\theta}]\, g'(\tilde{\theta})$
where $Var[\hat{\theta}] = \sigma^2 \mathbf{D}$ C3.24
Let $g_i'$ be the $i^{\text{th}}$ element of $g'(\hat{\theta})$ and let $v_{ij}$ be the $ij^{\text{th}}$ element of the matrix $\sigma^2 \mathbf{D}$, then:

$$Var[g(\hat{\theta})] = \sum_{i=1}^{k} \sum_{j=1}^{k} g_i' g_j' v_{ij}$$

Then as $n \to \infty$,

$$\sqrt{n}\,[g(\hat{\theta}) - g(\tilde{\theta})] \xrightarrow{L} N\{\mathbf{0}; n\sigma^2[g'(\tilde{\theta})^{\mathbf{T}} \mathbf{D}\, g'(\tilde{\theta})]\} \qquad , n \to \infty \qquad \text{C3.25}$$

Practically the partial derivatives are evaluated at $\hat{\theta}$ relative to $\theta$.
Since Eq C3.24 can be expressed in terms of the Fisher Information Matrix (Eq 3.2 above), then Eq C3.25 may be expressed as:

$$\sqrt{n}\,[g(\hat{\theta}) - g(\theta)] \xrightarrow{L} N\{\mathbf{0}; [g'(\theta)^{\mathbf{T}} \mathbf{I}^{-1}(\theta)\, g'(\theta)]\} \qquad , n \to \infty \qquad \text{C3.26}$$

# 4 General Discussions and Remarks

## 4.1 Discussion on EDA

***Linear Correlations***

Correlations were found amongst GU's GWH Losses, eg for Arnot GU3 with the following GU's:

| | |
|---|---|
| TU1 | 0.827 |
| KE1 | 0.799 |
| ML3 | 0.795 |
| MB5 | 0.735 |
| LE1 | 0.733 |
| HE9 | 0.732 |
| TU4 | 0.696 |
| HE2 | 0.645 |
| HE6 | 0.637 |
| MJ3 | 0.602 |
| HE1 | 0.601 |
| ML1 | 0.595 |
| HE3 | 0.591 |

However, these correlations are not necessarily yielding usable information: Coal is not a factor (different and unique sources). Plant equipment and design also not and the same goes for the quality of water. Nodal connections are same for some and totally different for others. At this point in time it is difficult to harness these correlations (future work will be performed using Correspondence Analysis), in other words, for instance, how do we answer the question: "so what that Arnot GU 3 is highly correlated to Tutuka GU 1, in terms of yearly GWh Losses?"

The hypothesis at the moment that the author can theorize is Management intervention and Load Factor (LF) effects. The latter has been mentioned in paragraph 2.3.3, pp.30-31 and illustrated in Fig 2.14 p.30. As stated in that paragraph this would constitute further analyses with a different dataset of a much larger nature. This has already been commenced by the author and just to give an idea, the size of one Power Station file consisting of 6 GU's is 17 Gigabytes!

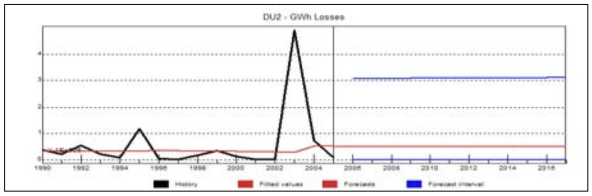
***Time Series Analyses***

This included analyses of the Autocorrelations and Partial Autocorrelations Functions. A sample of the results for 4 GU's (the others tend to follow similar results and patterns) are illustrated in the next two pages.

The only GU that showed significance (in the 1st order autocorrelation) was Duvha GU 4.

The Ljung-Box statistic indicated significance at the 95% level.

The reason is known to our business: there was a particular managerial intervention which was put into effect from 1996. The trend might be misleading for this interpretation and change-point analysis more appropriate in this case.




**Report for DU2**

**Analysis**

Using rule-based logic the choice was narrowed down to exponential smoothing or Box-Jenkins.

Series is too short to consider Box-Jenkins.

Exponential Smoothing is used.

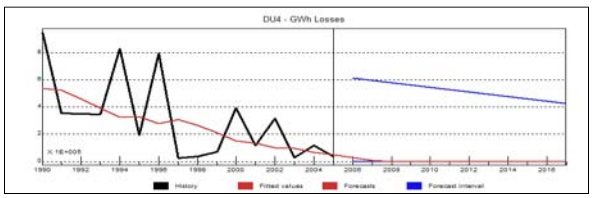
**Model Details**

**Selection**
**Simple exponential smoothing: No trend, No seasonality**

| Component | Smoothing Wgt | Final Value |
|---|---|---|
| Level | 0.05 | 499760 |

**Within-Sample Statistics**

| | | | |
|---|---|---|---|
| Sample size | 16 | No. parameters | 1 |
| Mean | 567318 | Std. deviation | 1208578.89 |
| R-square | 0 | Adj. R-square | 0 |
| Durbin-Watson | 2.04 | Ljung-Box(10) | 2.3 P=0.01 |
| Forecast error | 1247059.53 | BIC | 1316744.68 |
| MAPE | 467.94 | RMSE | 1207460.19 |
| MAD | 534168.34 | | |




**Report for DU4**

**Analysis**

Using rule-based logic the choice was narrowed down to exponential smoothing or Box-Jenkins.

Series is too short to consider Box-Jenkins.

Exponential Smoothing is used.

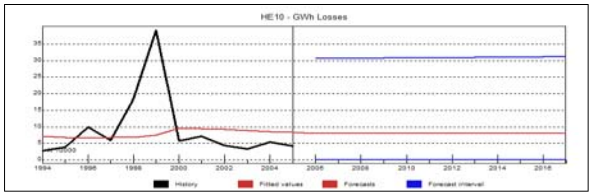
**Model Details**

**Selection**
**Holt exponential smoothing: Linear trend, No seasonality**

| Component | Smoothing Wgt | Final Value |
|---|---|---|
| Level | 0.06474 | 48106 |
| Trend | 0.9998 | -21647 |

**Within-Sample Statistics**

| | | | |
|---|---|---|---|
| Sample size | 16 | No. parameters | 2 |
| Mean | 310152.76 | Std. deviation | 304437.36 |
| R-square | 0.27 | Adj. R-square | 0.22 |
| Durbin-Watson | 2.46 | Ljung-Box(9) | 19.1 P=0.98 |
| Forecast error | 269720.36 | BIC | 300037.31 |
| MAPE | 178.2 | RMSE | 252300.29 |
| MAD | 197138.88 | | |




**Report for HE10**

**Analysis**

Using rule-based logic the choice was narrowed down to exponential smoothing or Box-Jenkins.

Series is too short to consider Box-Jenkins.

Exponential Smoothing is used.

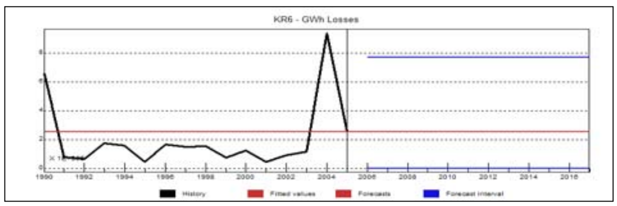
**Model Details**

**Selection**
**Simple exponential smoothing: No trend, No seasonality**

| Component | Smoothing Wgt | Final Value |
|---|---|---|
| Level | 0.06563 | 79929 |

**Within-Sample Statistics**

| | | | |
|---|---|---|---|
| Sample size | 12 | No. parameters | 1 |
| Mean | 91397.48 | Std. deviation | 104215.34 |
| R-square | 0 | Adj. R-square | 0 |
| Durbin-Watson | 1.48 | Ljung-Box(6) | 4.3 P=0.36 |
| Forecast error | 108509.78 | BIC | 115223.34 |
| MAPE | 82.47 | RMSE | 103890.2 |
| MAD | 66003.92 | | |




**Report for KR6**

**Analysis**

Using rule-based logic the choice was narrowed down to exponential smoothing or Box-Jenkins.

Series is too short to consider Box-Jenkins.

Exponential Smoothing is used.

**Model Details**

**Selection**
**Simple exponential smoothing: No trend, No seasonality**

| Component | Smoothing Wgt | Final Value |
|---|---|---|
| Level | 0 | 256318 |

**Within-Sample Statistics**

| | | | |
|---|---|---|---|
| Sample size | 16 | No. parameters | 1 |
| Mean | 205957.17 | Std. deviation | 244182.91 |
| R-square | 0 | Adj. R-square | 0 |
| Durbin-Watson | 1.65 | Ljung-Box(10) | 1.8 P=0.00 |
| Forecast error | 249660.85 | BIC | 263611.79 |
| MAPE | 162.09 | RMSE | 241733.08 |
| MAD | 186434.63 | | |

## 4.2 Steven's Theorem

The theorem originates from problem that the author was facing when trying to price Options using the Black-and-Scholes model in the realm of large volatilities in Energy Markets. The theorem was presented at the South African Statistical Conference (Cape Town) in November 2001.
"Steven" is an acronym for: **St**atistical **E**stimator for the **V**olatility – an **E**valuatio**n** which is the Title of the presentation.

<u>Steven's Theorem</u>:
**If we assume that the $\alpha$ angles formed by the spot prices (S) and the horizontal axis (time) follow a Uniform distribution, then the spot price difference (r) in the returns (R), behaves as a Cauchy distribution.**

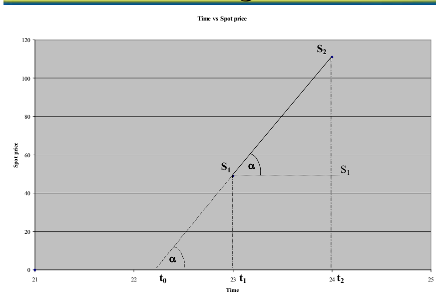
<u>Implications of Steven's Theorem</u>:
**This implies that the distribution of the returns (R) depends on the instantaneous distribution of the Spot prices.**
**If this is assumed to be N($\mu;\sigma$), then (R), by simulation, shows to be Cauchy**

The author chose to prove this geometrically:
One of the definitions for the volatility is that it's the Standard Deviation of the Returns

($R_i$) which in turn are defined as: $R_i = \dfrac{S_{i+1} - S_i}{S_i}$

19

## "the other angle"




$$\tan\alpha = \frac{S_2 - S_1}{t_2 - t_1}$$

$$\tan\alpha = \frac{S_1 - 0}{t_1 - t_0}$$

$$\therefore \frac{S_2 - S_1}{S_1} = \frac{t_2 - t_1}{t_1 - t_0}$$

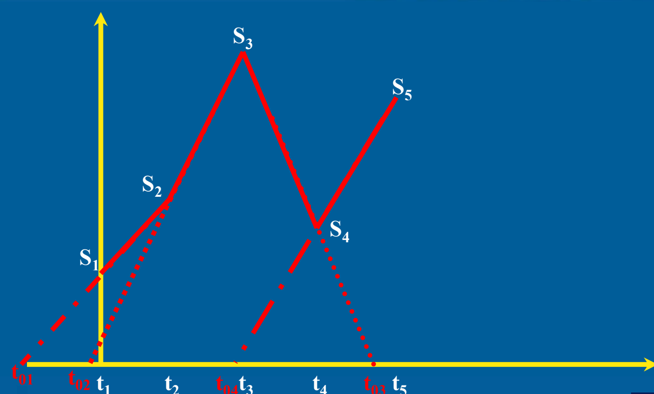
This equates to the instantaneous $R_1$, whereby $t_2$-$t_1$ would be unitary (e.g. hourly, daily, etc…)

$$\therefore R_1 = \frac{S_2 - S_1}{S_1} = \frac{1}{t_1 - t_0}$$

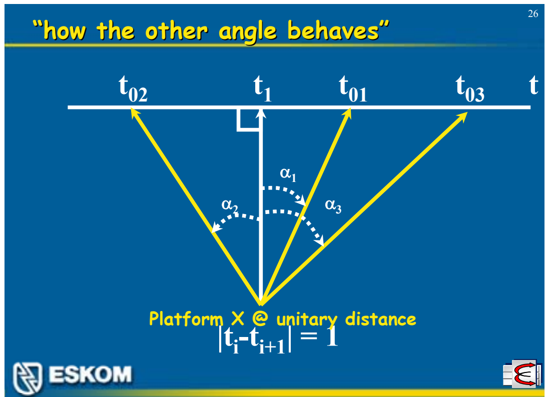
ESKOM

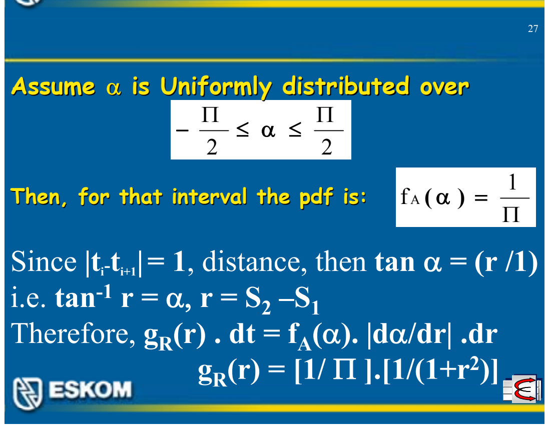
25

## Illustrated on the price graph




ESKOM




The Cauchy pdf is given by:

$$f_X(x) = \frac{1}{\pi\tau} \cdot \frac{1}{1 + (x+\mu)^2/\tau^2} \quad , \quad -\infty < x < +\infty$$

If $\mu = 0$ and $\tau = 1$, then the Cauchy above becomes:

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x)^2} \quad , \quad -\infty < x < +\infty$$

This is of the form as shown in the last slide [gR(r)].

**BIBLIOGRAPHY**

[1] Alexander SS, “Price Movements in Speculative Markets; Trends or Random Walks”, Industrial Management Review II, May, pp 7-26, 1961

[2] Beirlant J, Goegebeur, Segers J, Teugels J, “Statistics of Extremes: Theory and Applications”, John Wiley & Sons, 2004

[3] Beirlant J, Teugels JL, Vynckier, “Practical Analysis of Extreme Values”, Leuven University Press, 1996

[4] Clark PK, “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices”, Econometrica, January, pp 135-155, 1973

[5] Cootner PH, “Stock Prices: Random vs Systematic Changes”, Industrial Management Review III, Spring, pp 25-45, 1962

[6] Copeland TE, Weston JF, “Financial Theory and Corporate Policy”, Addison-Wesley 3rd Ed., 1992

[7] Davison AC, Smith RL, “Models of Exceedances over High Thresholds”, J. Royal Statistical Society, Series B,52,393-442, 1990

[8] De Waal D, “Goodness of Fit of a Strict Pareto: A Bayesian approach”, Technical Report No. 337, September 2004

[9] De Waal D, “Bayesian Goodness of Fit of a Strict Pareto”, South African Statistical Association (SASA) Conference, November 2004

[10] De Waal D, “Goodness of Fit Using the Entropy of the Dirichelet Process for Applying POT in Extreme Value Analysis”, UFS Workshop, June 2005

[11] De Waal D, Beirlant J, “Choice of the Threshold through the Entropy of the Dirichlet Process when applying POT”, South African Statistical Association (SASA) Conference, November 2005

[12] Dierckx  G, “Estimation of the Extreme Value Index”, PhD Thesis, Katholieke Universiteit Leuven, 2000

[13] Embrechts P, Kluppelberg C, Mikosch T, “Modelling Extremal Events”, Springer, 1997

[14] Fama EF, “The Behaviour of Stock Market Prices”, Journal of Business, January, pp 34-105, 1965

[15] Fisher RA, Tippett LHC, “Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample”, Proc. of the Cambridge Philosophical Soc., 24, pp 180-190, 1928

[16] Fuller WE, “ Flood flows”, American Society of Civil Engineers Transactions, v. 77, no. 1293, p. 564-617, 1914

[17] Giordano FR, Weir MD, Fox WP “Mathematical Modelling”, Thomson-Brookes/Cole 3rd Ed., 2003

[18] Griffith AA , “The phenomena of rupture and flow in solids”, Philosophical Transactions of the Royal Society A (London) 221, pp. 163–198, 1920

[19] Hernandez-Campos F, Marron JS, Samorodnitsky G, Smith FD “Variable Heavy Tailed Durations in Internet Traffic”, unpublished manuscript

[20] Honkela A, “Nonlinear Switching State-Space Models”, Master’s dissertation, Helsinki University of Technology, Espoo, 2001

[21] Hosking JRM, “Some Theoretical Results concerning L-Moments”, Report RC 14492, IBM Research Division, 1996

[22] Hull JC, White A, “Value at Risk when Daily Changes in Market Variables are not Normally Distributed”, Journal of Derivatives, Vol.5, No.3 (Spring), pp 9-19, 1998

[23] Kedem B, Pavlopoulos H, “On the Threshold Method for Rainfall Estimation: Chhosing the Optimal Threshold Level”, Journal of the American Statistical Association, Vol.86, No.415, pp 626-633, Sep 1991

[24] Mandelbrot B, “The Variation of Certain Speculative Prices”, Journal of Business, 36, October, pp 394-419, 1963

[25] McNeil A, “EVT and Copulas in Financial and Insurance Risk Management”, Federal Institute of Technology, Dept. of Mathematics, Course Notes, 2002

[26] Micali V, “On the Distribution of the Volatility of the System Marginal Prices in the Energy Sector”, SASA Conference, Cape Town, 2001

[27] Moore A, “A Statistical Analysis of Common Stock Prices”, PhD dissertation, Graduate School of Business, University of Chicago, 1962

[28] Nel A, de Waal D, van Gelder PHAJM, “Estimating Joint Tail Probabilities through the Logistic Copula”, A Collection of papers on the 50$^{th}$ anniversary of the Department, Department of Mathematical Statistics, University of the Free State, RSA, Nov 2006

[29] Nelson DB, “Conditional Heteroskedasticity in Asset Returns. A New Approach”, Econometrica, Vol.59, No.2 (March), pp 347-370, 1991

[30] Pandey MD, van Gelder PHAJM, Vrijling JK, “The Estimation of Extreme Quantiles of Wind Velocity using L-Moments in the Peak-Over-Threshold Approach”, Structural Safety, No.23, pp 179-192, 2001

[31] Pickands J, “Statistical Inference Using Extreme Order Statistics”, The Annals of Statistics, No.3, pp 119-131, 1975

[32] Reiss R, Thomas M “Statistical Analysis of Extreme Values”, Birkhäuser, Basel, 1997

[33] Shannon CE, “A Mathematical Theory of Communication”, The Bell System Technical Journal, Vol. 27, pp 379-423, 623-656, July, October 1948

[34] Smith RL and Shively TS, “A point Process Approach to Modelling Trends in Tropospheric Ozone”, Athmospheric Environment, No.29, pp 3489-3499, 1995

[35] Tippett LHC “On the Extreme Individuals and the Range of Samples Taken from a Normal Population”, Biometrika 17, pp 364-387, 1925

[36] Tsay RS, “Analysis of Financial Time Series”, John Wiley & Sons, 2002

[37] Tukey JW, Mosteller F “Data Analysis and Regression”, Addison-Wesley Publishing Co., 1977

[38] Wilmott P, “Paul Wilmott on Quantitative Finance”, John Wiley & Sons, 2000

[39] Wilson TC, “Infinite Wisdom”, Risk, Vol.6 (June), pp 37-46, 1993

[40] van Gelder PHAJM, “Performance of Parameter Estimation Techniques with Inhomogeneous datasets of Extreme Water Levels along the Dutch”, unpublished paper, TU Delft, Faculty of Civil Engineering, circa 1999

[41] van Noortwijk JM, Kalk HJ, Duits MT, Chbab EH. “Bayesian Statistics for Flood Prevention”, HKV Consultants and Institute for Inland Water Management and Waste Water Treatment (RIZA), Ministry of Transport, Public Works, and Water Management, Lelystad, The Netherlands, 2007

[42] van Zyl M, Informal presentation at the Risk Laboratory, University of the Free State, March 2004

[43] Zellner A, “Bayesian Analysis in Econometrics and Statistics”, Edward Elgar Publ. Ltd, 1997

***Additional References***

[44] Davison AC, “Modelling Excesses over High Thresholds”, Statistical Extremes and Applications, J. Tiago de Oliveira (ed.), D Reidel Publishing Co., pp 461-482, 1984

[45] Smith RL, “Threshold Methods for Sample Extremes”, Statistical Extremes and Applications, J. Tiago de Oliveira (ed.), D Reidel Publishing Co., pp 621-638, 1984

[46] Van Gelder, PHAJM “Statistical Methods for the Risk-Based Design of Civil Structures”, PhD Thesis, 1999-2000