IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 2, 2020, accepted August 6, 2020, date of publication August 10, 2020, date of current version August 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015259

Analysis of Extreme Peak Loads Using Point Processes: An Application Using South African Data

JERRY BOANO-DANQUAH, CASTON SIGAUKE, AND KWABENA A. KYEI
Department of Statistics, University of Venda, Thohoyandou 0950, South Africa
Corresponding author: Jerry Boano-Danquah (jerrykd1979@gmail.com)

This work was supported by the National Research Foundation (NRF) of South Africa.

ABSTRACT Extreme value modelling of peak load process is critical to the reliable specification of power generation, distribution and maintenance purposes during both peak and off-peak periods. In this study, a frequency assessment of extreme peak electricity demand for the four seasons of the year using South African data for the period, January 1997 to December 2013 is carried out. A point process approach from extreme value theory is proposed as an ingenious extreme value theory approach. The data are made stationary by using a time-varying threshold which has a positive shift factor. The non-linear detrended datasets are then grouped into summer, spring, winter and autumn according to the calendar dates in the Southern Hemisphere. The datasets were declustered to keep the series relatively independent using Ferro and Segers automatic declustering method. A stationary point process model is then fitted to each of the cluster maxima. The modelling framework, which is easily extensible to other peak load parameters, assumes that peak power follows a Poisson point process. The parameters of the developed model are estimated using the maximum likelihood method. Empirical results show that daily peak electricity demand could be experienced approximately 27, 16, 7 and 15 days per year in winter, spring, summer and autumn, respectively. The modelling approach could assist system operators of utility companies in scheduling maintenance of generating units including long term planning.

INDEX TERMS Extreme value theory, daily peak electricity demand, peaks-over threshold, maximum likelihood estimation, point process.

NOMENCLATURE

ARMA	Autoregressive moving average
DJF	December January and February
EIEs	Energy-intensive enterprises
ESKOM	South African power utility company
EVT	Extreme value theory
DPED	Daily peak electricity demand
DJF	December, January, February
GEVD	Generalised extreme value distribution
GEV	Generalised extreme value
GPD	Generalised Pareto distribution
GSP	Generalised single Pareto
LCL	Lower control limit
JJA	June, July, August
MAM	March, April, May
MLE	Maximum likelihood estimation
POT	Peaks over threshold
PP	Point process
QQ	Quantile to quantile plot
RL	Return level
SON	September, October, November
UCL	Upper control limit

The associate editor coordinating the review of this manuscript and approving it for publication was Xiao-Sheng Si.

I. INTRODUCTION

Since 1994, the increased electricity demand driven by population growth and industrialisation has led to sustainability issues in South Africa. The electricity generating capacity does not show potential signs of meeting the country's demand and this has impacted negatively on the national grid. As a result, the national grid has been operating in a risky and vulnerable state, leading to disturbances such as load shedding in the past decade. In particular, it is of greater interest to have sufficient information about the extreme value of the stochastic load processes in time for proper planning, designing the generation and distribution system and storage devices. Effective planning ensures efficiency in the electrical energy sector and consequently maintains discipline in the national grid system.

A point process characterisation of extremes is presented in the paper. The study illustrates the importance of the modelling approach using daily peak electricity demand (DPED) data from South Africa. Extreme value models are normally used to describe the distribution of rare events. The univariate extreme value theory (EVT) is now relatively standard to model the tail of the distribution of a scalar random variable. Typically, an asymptotically motivated extreme value model is applied to approximate the tail of some population distribution. High electricity demand is normally driven by very low (winter) and very high temperatures (summer) ([1], [2]; among others).

A. BACKGROUND

The motivation behind this study stems from the need to assess the magnitude and frequency of occurrence of extreme peak electricity demand over a specified threshold using South African electricity data for each of the four seasons of the year, i.e. winter, spring, summer and autumn, respectively. Both grid companies and large consumers want to estimate their electricity consumption in the future, to ensure secured and economic system operation ([3]). Reference [4], further claim that, in many cases, energy-intensive enterprises (EIEs) have their generating plants, thus forming a grid. Notably, in many industrialised countries, electrical energy consumption in EIEs constitute a significant part of the country's total energy use ([4]).

To have a better understanding of the frequency of occurrence of extreme peak loads it is important to focus on the upper tail of the distribution of electricity demand. Modelling the residual tail distribution of peak electricity demand may help system operators and forecasters understand better how extreme events, such as heatwaves, prolonged heavy snowfall may affect load forecasts ([5]). Such events can create a surge in electricity demand which may affect the stability of the grid leading to possible brownouts including blackouts and load shedding. On the same note, [6] argue that the reliability of a network grid could be improved through an assessment of the frequency of occurrence of extreme peak electricity demand. These studies used either the block maxima approach or the peaks over threshold (POT) approach.

In their study, [7] outlined the use of the block maxima approach to estimate the maximum load forecast errors (residuals), up to several decades ahead to assess the risk inherent in long-term electricity demand projection taking into account climate effects. The study discussed the maximum residuals in terms of return levels and return periods that would occur with a finite time series. Empirical results showed that using a generalised extreme value (GEV) model, the number of return observations exceeding a given return level was in line with what was predicted.

Reference [2] presented work on modelling extreme peak electricity demand during a heatwave period. A frequency analysis of the occurrence of monthly cluster maxima of peak electricity demand was carried out. Areas for future study were then proposed. Our paper extends this work to frequency analysis of extreme peak electricity demand in the four seasons of the year. Reference [8] suggested the use of the generalised Pareto distribution (GPD) to explain how extreme electricity demand changes with time in the United Kingdom. The main idea underlying their study was to allow the GPD parameters to vary linearly and in a non-stationary sinusoidal manner. This was done to capture some patterns exhibited by electricity demand data.

A hybrid model called an autoregressive moving average-exponential generalised autoregressive conditional heteroscedasticity-generalised single Pareto (ARMA-EGARCH-GSP) was developed in [9]. The model was used for estimating extreme quantiles of inter-day increases in peak electricity demand. The authors carried out a comparative analysis with an ARMA-EGARCH model. Empirical results from this study showed that for estimating extreme tail quantiles the ARMA-EGARCH-GSP outperformed the ARMA-EGARCH model. The current paper focuses on the use of point processes in the assessment of the occurrence of extreme peak loads.

The point process characterisation of extremes introduced by [10] has been used by several authors and researchers including [11]–[13] who emphasised the importance of a point process technique in EVT modelling. The two main features of the point process (PP) are singled-out: firstly, the PP provides the analysis of extremes through merging the block maxima and POT approaches; secondly, the point process is more associated with the variations in the excesses above the threshold than the peaks-over-threshold approach ([12], [13]; among others).

A characterisation of extreme peak loads using point processes is presented in [14]. The study was then extended to power system reliability. Modelling and assessment of the risk of electricity peak loads using EVT are discussed in [15]. The authors present a detailed discussion of how EVT models can be used to forecast electricity peaks on a low voltage network. In a related study, [16] propose the use of point processes in modelling and forecasting of extreme electricity demand. The authors argue that their modelling approach is not only flexible and realistic but also that consistent outcome can be guaranteed.

Reference [11] applied the EVT in modelling the environmental data that were captured in Houston, Texas. The non-stationary generalised extreme value distribution (GEVD), non-stationary generalised Pareto distribution (GPD) as well as the point process were used. The study aimed to investigate the existence of a trend in the data which consisted of hourly measurement of ozone from April 1973 to December 1986. The study aimed at estimating the frequency with which specified high levels are exceeded. Reference [11], concluded that there was a downward trend in the crossing rates at high levels. This indicates that the work of the regulatory bodies has a significant effect by at least reducing the frequency of the occurrences of high emissions and without such regulations, one could argue that there would be an increasing trend.

B. RESEARCH HIGHLIGHTS

The main highlights of this study are as follows: The main contribution of this paper is the estimation of the intensity function of the point process which measures the rate of occurrence of daily peak electricity demand per year. The knowledge of the intensity function coupled with the 95% return level estimation could be useful to Eskom, the South African power utility company in ensuring stability in the grid system. In this study, the focus was on the frequency of occurrences of extreme peak electricity demand in each of the four seasons of the year, i.e. winter, summer, spring and autumn following the calendar dates in the Southern Hemisphere. Use of the reparameterisation approach of the Poisson-generalised Pareto distribution has shown that extreme daily peak electricity demand could be experienced approximately 27, 16, 7 and 15 days per year in winter, spring, summer and autumn, respectively. The modelling approach could assist system operators of utility companies in scheduling maintenance of generating units including long term planning for capacity expansion.

The rest of the paper is organized as follows: Section 2 presents the models. The empirical results are presented in Section 3. Section 4 presents the discussion of the results while Section 5 concludes with policy implications.

II. METHODOLOGY

A. POINT PROCESS CHARACTERISATION OF EXTREME DAILY PEAK ELECTRICITY DEMAND

In the analysis of unusual events, EVT provides three important approaches. Firstly, the block maxima approach approximates the parent distribution's tail with distribution for the maxima over time blocks of equal size ([17]). The second is the peaks-over-threshold (POT) approach which approximates the behaviour of extremes with distribution for the values over a sufficiently high threshold. The point process (PP) is the third and it unifies the first two approaches ([11], [12]). The PP describes the occurrences in time or space and can be used to model threshold excesses as occurrences in time. The study made use of the stationary PP characterisation of extreme peak loads.

Electricity data exhibit a large degree of non-stationarity, hence, the data is initially made stationary by non-linear detrending. The study covers stationary dependent series and the use of extremal mixture models where a non-parametric kernel density is fitted to a fixed threshold on the positive exceedances above a time-varying threshold. The extremal index which is the measure of dependence or independence structure of the daily peak electricity data is estimated.

The PP statistical approach was introduced by [11], although the fundamental probability theory upon which it derives had already existed in literature. The modelling framework of [18] illustrated that more light can be shared on PP using the viewpoints of EVT. Under this modelling framework, the times at which high-threshold exceedances occur and the excess values over the chosen threshold are unified into one process based on a two-dimensional plot of exceedance times and values instead of considering the two events as separate processes. Here, the asymptotic theory of threshold exceedances shows that under suitable normalisation, this process behaves like a non-homogeneous Poisson process. As explained in [19] and discussed in detail by [20], a PP is a stochastic model of points that are randomly scattered in some space, where the points may denote times of phenomena or location of objects that are characterised by a stochastic system. According to [11], [12], a PP serves as an approach for unifying and extending EVT modelling based on both block maxima and threshold methods.

In order to fit a PP model, the block maxima $M_n = \max\{X_1, X_2, \ldots, X_n\}$ should be distributed with the GEVD for the normalizing constants $\{a_n > 0\}$ and b_n. This leads to the sequential point process T_n on R^2 which is given in Eq. (1).

$$T_n = \left\{ \left(\frac{i}{(n+1)}, \frac{(Y_i - b_n)}{a_n} \right), i = 1, .., n \right\} \tag{1}$$

such that an axis of time passes through the closed interval $(0, 1)$; and the second point ensures stability in the occurrence of extremes as $n \to \infty$ such that on $[0, 1] \times [\tau, \infty]$, $K_n \to K$ as $n \to \infty$, where K is a heterogeneous Poisson process as emphasized in [13], [19], [20]. For a higher threshold, τ and for a given space of the form $A = [0, 1] \times (u, \infty)$, all the values of T_n possess a p chance of happening within A, where

$$p = P\left\{ \frac{X_i - b_n}{a_n} > \tau \right\} \approx \frac{1}{n} \left[1 + \xi (\frac{\tau - \mu}{\sigma}) \right]^{-1/\xi} . \tag{2}$$

As the binomial mass approaches the limiting Poisson distribution, then as $n \to \infty$, $T_n(A)$ obeys $\mathrm{Poi}(\Lambda(A))$ such that for all spaces that satisfy $A = [t_1, t_2] \times (u, \infty)$, with $[t_1, t_2] \subset [0, 1]$, the limiting distribution of $T_n(A)$ is also $\mathrm{Poi}(\Lambda(A))$, where

$$\Lambda(A) = \begin{cases} (t_2 - t_1) \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} & \xi \neq 0 \\ (t_2 - t_1) \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] & \xi = 0 \end{cases} \tag{3}$$

occurs as a homogeneous result of the process in the direction of time as emphasized in [13], [19]. The corresponding intensity function is then given as given in Eq. (4):

$$\lambda(x) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} & \xi \neq 0 \\ \frac{1}{\sigma} \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] & \xi = 0 \end{cases} \tag{4}$$

Normally a scaling factor n_x which denotes the number of years of observations to the intensity function is used.

This results in the estimation of the GEVD parameters, μ, σ, ξ. The intensity measure is then defined as given in Eq. (5).

$$\Lambda(A) = \begin{cases} n_x(t_2 - t_1)\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}, & \xi \neq 0 \\ n_x(t_2 - t_1)\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right], & \xi = 0 \end{cases} \quad (5)$$

Let the PP P_n be defined on the space $A = (0, 1] \times [\tau, \infty)$ with corresponding intensity measure $\Lambda(A; \theta)$ as defined in Eq. (5). Given that the probability density (*pdf*) of the points of a Poisson process in a set A, then the likelihood function is given as (Coles, 2001):

$$L(\theta; x_1, \ldots, x_n) = \exp\{-\Lambda(A;\theta)\} \frac{(A;\theta)^{n_\tau}}{n_\tau!} \prod_{i=1}^{n_\tau} \frac{\lambda(X_i;\theta)}{\Lambda(A;\theta)}, \quad (6)$$

where $\Lambda(A;\theta) = \int_A \lambda(X_i;\theta)dx$. This results in the generalised likelihood function given as

$$L(\tau, \mu, \sigma, \xi | x) \propto \begin{cases} \exp\left\{-n_x\left[1 + \xi\left(\frac{\tau-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \\ \times \prod_{i=1}^{n_\tau} \frac{1}{\sigma}\left[1 + \xi\left(\frac{\tau-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \exp\left[-n_x \exp\left(\frac{\tau-\mu}{\sigma}\right)\right] \\ \times \prod_{i=1}^{n_\tau} \frac{1}{\sigma}\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right], & \xi = 0 \end{cases} \quad (7)$$

B. THRESHOLD SELECTION, STATIONARITY AND DECLUSTERING

If the observations $x_1, \ldots, x_n$ represent the independently and identically distributed (i.i.d) average daily peak electricity demand and suppose τ denote an arbitrarily and sufficiently high threshold, then the observations $x_1, \ldots, x_k$ are the k positive residuals above the sufficiently high threshold τ if $x_i : x_i > \tau$ and the observations $y_j = x_i - \tau$, for $i = 1, \ldots, k$ are the threshold excesses ([12]).

Reference [12] also emphasizes that the selection of threshold processes is always a trade-off between the bias and variance and, if the chosen threshold is too small, it violates the asymptotic properties underlying the derivation of the GPD. However, if the chosen threshold is too high, the excesses $(x - \tau)$ above the threshold becomes too small to estimate the shape and the scale parameters, leading to high variance. Hence the threshold selection requires proper analysis to determine whether the limiting model provides a sufficiently good approximation versus the variance of the parameter estimates.

There are numerous diagnostic models used to determine the threshold, such as, the Pareto quantile plot, mean excess plot, threshold stability plot, extremal mixture plots among others. In this study extremal mixture models discussed in [21] are going to be used. A standard kernel density estimator representing the bulk model and the tail model represented by the PP or the GPD is discussed in [21]. The distribution function is given by [21]:

$$F(x|X, \alpha, \tau, \sigma_\tau, \xi, \phi_\tau) = \begin{cases} (1-\phi_\tau)\frac{H(x|X,\alpha)}{H(\tau|X,\alpha)}, \\ x \leq \tau, \\ (1-\phi_\tau) + \phi_\tau \times G(x|\tau, \sigma_\tau, \xi), \\ x > \tau, \end{cases} \quad (8)$$

where $H(.|X, \alpha)$ is the distribution function of the kernel density estimator, α is the band width, τ is the threshold, ξ denotes the shape parameter (extreme value index), σ_τ represents the scale parameter for the exceedances and ϕ_τ is the estimated sample proportion of the data above τ, calculated as $\phi_\tau = \frac{n_\tau}{n}$.

To ensure homogeneous processes, two main approaches were used to make the data stationary, i.e. non-linear detrending and differencing approaches. The data is initially detrended using a penalised regression cubic smoothing spline with a positive shift factor given in Eq. (9). See [21] for details.

$$\eta(t) = \sum_i^n (x_i - f(t_i))^2 + \lambda \int (f''(t))^2 dt + \tau, \quad (9)$$

where y_i denotes the daily peak electricity demand and λ is a smoothing parameter and $\tau \epsilon \Re$ is a shift factor which should be large enough to accommodate asymptotic conditions when the PP is fitted to the observations exceeding τ ([22]). The study uses extremal mixture models in determining the value of τ and observations above the time-varying threshold $\eta(t)$ are extracted without the shift factor. Following the work by [22], the positive shift factor τ is estimated using extremal mixture model given in Eq. (9). Reference [23] emphasised that the use of penalised cubic smoothing splines as a time-varying threshold has attractive features such as deseasonalising and detrending at the same time.

The initial stage is to identify the clusters in the data, an approach that is commonly known as declustering. The exceedances will be declustered using the intervals estimator method discussed in [24]. This will be followed by fitting a stationary PP model to the cluster maxima.

III. EMPIRICAL RESULTS

A. EXPLORATORY DATA ANALYSIS

The datasets used in this study consist of daily peak electricity demand (DPED) from Eskom, South African power utility company. The length of the data is 16 years spanning from the period, January 1997 to December 2013 constituting 6209 observations. The analysis presented here applies to the non-linear detrended datasets. The data are divided into winter, spring, summer and autumn according to the calendar dates in the Southern Hemisphere. Winter is defined as June, July and August (JJA), spring as September, October and November (SON), summer as December, January and February (DJF) and autumn as March, April and May (MAM). The Poisson point process model is used in the modelling of the non-linear detrended DPED data for the four seasons. Table 1 shows summary statistics of DPED data for the sampling period, January 1997 to December 2013. From Table 1, Q_1, Q_2 and Q_3 represent the first quartile, second quartile (median) and the third quartile, respectively. The DPED data from Table 1 have a mean value less than the median value, indicating negatively skewed data as confirmed by the skewed value of -0.248 from the table. The value of the kurtosis which is 2.295 inferred the distribution of DPED is platykurtic. A visual inspection of the time series plot in Fig. 1 shows that the DPED data is non-stationary and have a high seasonality and a deterministic upward trend. We divide the non-linear detrended DPED data into four seasons, winter, spring, summer and autumn according to the calendar dates in the Southern Hemisphere for yearly frequency analysis. Figs. 2-5 show plots of the DPED data from the four seasons superimposed with the time-varying thresholds (non-linear trend lines). The box plots showing the distribution of DPED in the different seasons of the year are shown in Fig. 6. As shown in Fig. 6 the highest demand for DPED is experienced in winter and the least demand is in summer. The distributions of demand in various seasons are important to system operators including decision-makers

TABLE 1. Summary statistics of DPED.

Minimum	Q_1	Q_2	Mean
17600	25660	29090	28580
Q_3	Maximum	Kurtosis	Skewness
31420	36660	2.295	-0.248

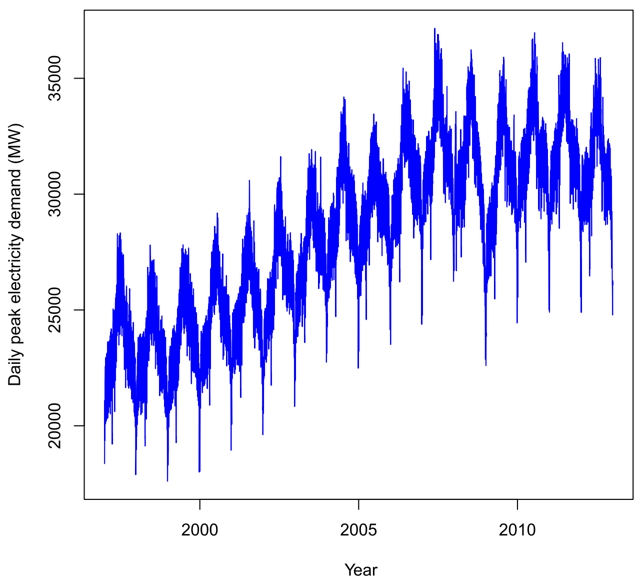


FIGURE 1. Plot of daily peak electricity demand.

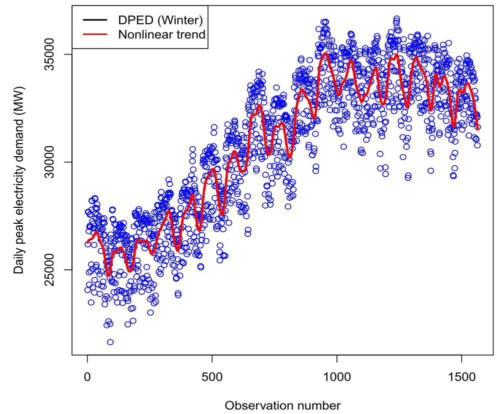


FIGURE 2. Spring nonlinear detrending.

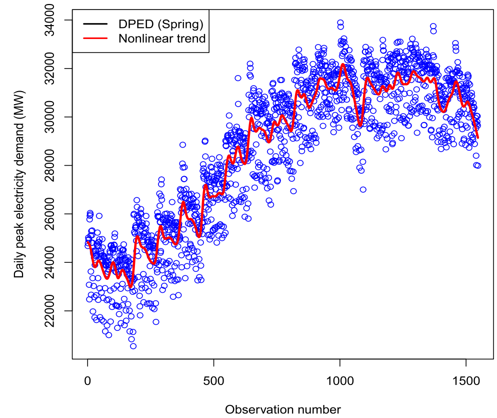


FIGURE 3. Spring nonlinear detrending.

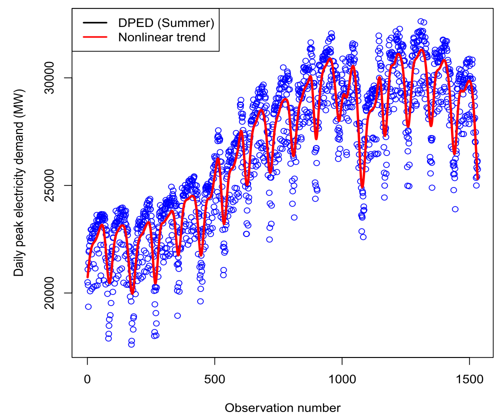


FIGURE 4. Summer nonlinear detrending.

in planning for scheduling of the maintenance of generating units.

Table 2 presents the descriptive statistics of the nonlinear detrended winter, spring, summer and autumn data set. In Table 2, Min denotes minimum, Max represents Maximum, Std Dev is the Standard Deviation, Kurt is Kurtosis and Skew represents Skewness. All the descriptive statistics were rounded to the nearest whole number except for kurtosis and skewness which were rounded to three decimal places.

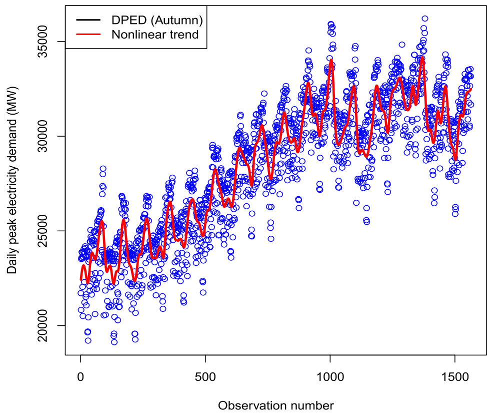


FIGURE 5. Autumn nonlinear detrending.

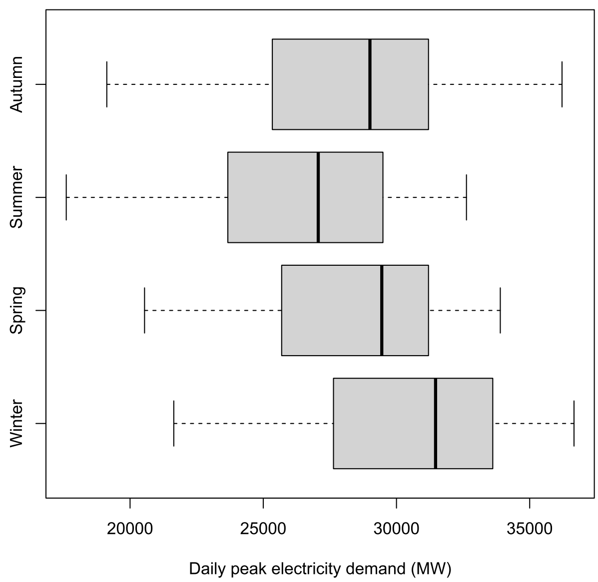


FIGURE 6. Box plots of daily peak electricity demand.

Table 2 representing the descriptive statistics of the winter, spring and autumn DPED show negative kurtosis value indicating lighter tails and flatter peak than the normal distribution. The kurtosis of the summer DPED data is positive meaning that the data are characterised by heavier tails and sharper peak than the normal distribution. All four datasets are negatively skewed. A comparative analysis of the four data sets reveals the spring data have the lowest standard deviation of 1086.4, inferring that the spring DPED were closely dispersed around the mean value of 47.86.

In Table 2, Min is minimum, Q_1 denotes quartile 1, Q_2 is quartile 2, Max is maximum, StDev is the standard deviation, Kurt is kurtosis and Skew is skewness.

B. POINT PROCESS ANALYSIS OF THE WINTER, SPRING, SUMMER AND AUTUMN DPED DATA

1) WINTER DPED

The winter period is defined as a period spanning from 1 June to August 31 (JJA). The original winter series has a length of 1235 observations. The data are initially detrended using penalised cubic smoothing splines. An initial threshold is set at zero and only the positive excesses above zero are considered. The non-parametric extremal mixture model is fitted to the observations to determine a sufficiently high threshold (which was found to be $\tau = 1516$) and the exceedances are then declustered using the intervals estimator technique discussed in [24]. The PP model is then fitted to the cluster maxima. The parameter estimates are used to compute the yearly frequency of occurrence of extreme peak DPED during the winter season. Fig. 7 is the time series plot of the non-linear detrended DPED winter data together with the density, normal quantile to quantile (QQ) and box plots.

TABLE 2. Summary statistics of detrended winter, spring, summer and autumn.

Winter								
Min	Q_1	Q_2	Mean	Q_3	Max	StDev	Kurt	Skew
-3494	-1066	404	25	1008	2728	1244	-0.849	-0.473
Spring								
Min	Q_1	Q_2	Mean	Q_3	Max	StDev	Kurt	Skew
-3039	-867	389	48	854	3187	1086	-0.659	-0.575
Summer								
Min	Q_1	Q_2	Mean	Q_3	Max	StDev	Kurt	Skew
-4588	-10540	489	-31	1010	2341	1363	0.171	-0.901
Autumn								
Min	Q_1	Q_2	Mean	Q_3	Max	StDev	Kurt	Skew
-4265	-11470	446	-40	975	2475	1332	-0.200	-0.764

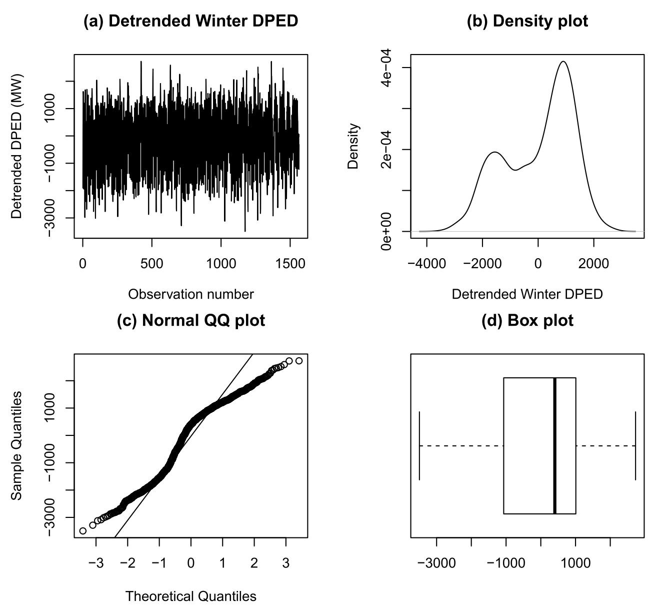


FIGURE 7. Non-linear detrended winter DPED data.

Fig. 8 shows the threshold estimation with the vertical line indicating the value of the threshold which was determined as $\tau = 1516$.

Following the interval estimation method of [24], the estimate of the extremal index is found to be $\hat{\theta} = 0.565$ (fairly independent stationary process) with the exceedances occurring in groups of $1.77 \approx 2$ and 47 identified cluster maxima were observed. The maximum likelihood

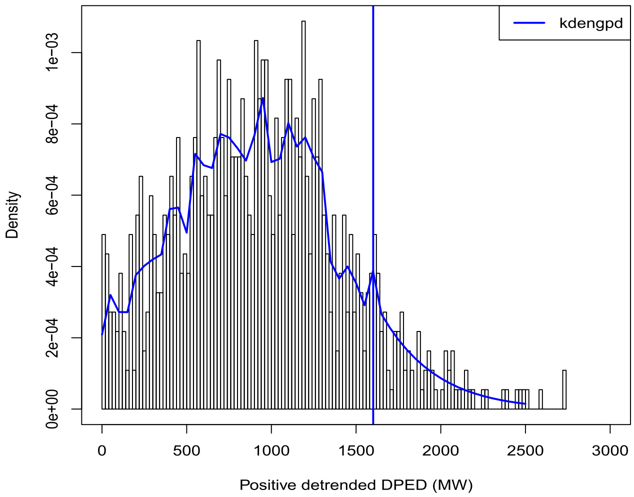


FIGURE 8. The threshold estimation using the non-parametric extremal mixture model, $\tau = 1516$.

estimates of the stationary point process fitted to the winter cluster maxima with their standard errors in parentheses are $\hat{\mu} = 3149.8(503.7)$, $\hat{\sigma} = 174.2(168.1)$, $\hat{\xi} = -0.131(0.198)$. In this study the reparameterisation approach discussed in [25] is used.

$$\hat{\sigma}^* = \sigma + \xi(\tau - \mu) \text{ and } \hat{\lambda}^* = \left[1 + \xi\left(\frac{\tau - \mu}{\hat{\sigma}^*}\right)\right]^{-\frac{1}{\xi}}$$

The new scale parameter, σ after reparameterisation is estimated as:

$$\begin{aligned}\hat{\sigma}^* &= 174.232 - 0.131(1601 - 3149.804)\\ &= 377.125\end{aligned}$$

The intensity of the point process which measures the frequency of the occurrence of the DPED demand is estimated per year as follows:

$$\begin{aligned}\hat{\lambda}^* &= \left(1 - 0.131\frac{(1601 - 3149.804)}{377.125}\right)^{7.633}\\ &= 26.732\\ &\simeq 27.\end{aligned}$$

The estimated intensity ($\hat{\lambda} \simeq 27$), indicates that, DPED would be experienced approximately 27 days in a year.

Fig. 9 shows the diagnostic plots of the PP model fitted to the cluster maxima for the winter data. Visual inspection of the plots shows a fairly good fit. The formal goodness of fit tests based on Anderson-Darling test statistics A^2 and Cramer-Von Mises test statistics W^2 for each data set were consistent with Poisson GPD with their scale and shape parameters at 5% level of significance. The results of each formal test have been summarised in tables found in appendix B. The test statistics are consistent with the usual asymptotic arguments underpinning the Poisson GPD the model.

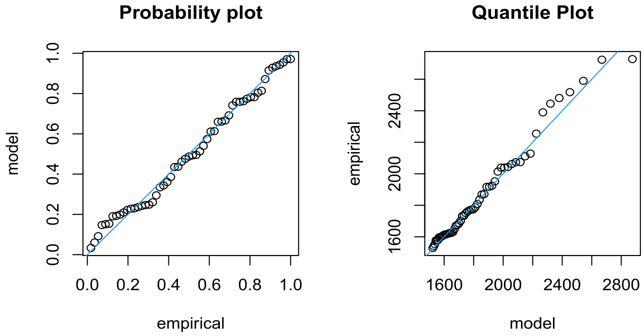


FIGURE 9. Diagnostic plots of the PP fitted to the winter cluster maxima.

2) SPRING DPED

The spring data spans from 1 September to November 30 (SON). The length of the original series is 963, with threshold $\tau = 1183$, interval estimator of extremal index $\theta = 0.5822$ inferring a fairly independent process as θ is closer to 1 than 0, run length of 1 and 91 identified cluster maxima observations. The maximum likelihood estimates with their standard errors in parentheses are $\hat{\mu} = 3708.0(200.0)$, $\hat{\sigma} = 109.0(58.4)$, $\hat{\xi} = -0.183(0.112)$.

The new scale parameter, σ after reparametrisation is estimated as:

$$\begin{aligned}\hat{\sigma}^* &= 109.005 - 0.183(2551.408 - 3707.971)\\ &= 320.656.\end{aligned}$$

and the intensity is estimated as:

$$\begin{aligned}\hat{\lambda}^* &= \left(1 - 0.183\frac{(2551.408 - 3707.971)}{320.656}\right)^{5.4645}\\ &= 15.953\\ &\simeq 16.\end{aligned}$$

The estimated intensity ($\hat{\lambda} \simeq 16$), shows 16 days of extreme DPED will be experienced per year.

3) SUMMER DPED

The summer DPED is from 1 December to February 28/29 (DJF) of each year depending on whether February is in a leap year or not. The length of the original series is 905, with threshold $\tau = 1125$, number of threshold exceedances 74, interval estimator of extremal index $\hat{\theta} = 0.3645$ (weak dependence in the cluster maxima series). The exceedances occur in groups of $2.74 \approx 3$. The maximum likelihood estimates with their standard errors in parentheses are $\hat{\mu} = 2357.52(98.97)$, $\hat{\sigma} = 53.62(23.07)$, $\hat{\xi} = -0.3894(0.0934)$.

The new scale parameter, σ after reparameterisation is estimated as:

$$\begin{aligned}\hat{\sigma}^* &= 53.62 - 0.3894(1125 - 2357.52)\\ &= 533.609.\end{aligned}$$

and the intensity is estimated as:

$$\hat{\lambda}^* = \left(1 - 0.3894\frac{(1125 - 2357.52)}{533.609}\right)^{2.5678}$$
$$= 5.194$$
$$\simeq 6.$$

The estimated intensity ($\hat{\lambda} \simeq 6$), indicates that approximately 6 days of extreme DPED will be experienced in a year.

4) AUTUMN DPED

The autumn is for the period 1 March to 31 May of each year. The estimated threshold is $\tau = 1550$, an estimated extremal index $\hat{\theta} = 0.5841$ (fairly independent stationary process) indicating that exceedances occur in groups of $1.71 \approx 2$ and 44 identified cluster maxima observations. The maximum likelihood estimates with their standard errors in parentheses are $\hat{\mu} = 2685.1(265.2)$, $\hat{\sigma} = 107.6(79.9)$, $\hat{\xi} = -0.190(0.2)$.

The new scale parameter, σ after reparametrisation is estimated as:

$$\hat{\sigma}^* = 107.633 - 0.190(1513.340 - 2685.083)$$
$$= 330.264.$$

The estimated intensity is:

$$\hat{\lambda}^* = \left(1 - 0.20.19\frac{(1513.340 - 2685.083)}{330.264}\right)^{5.263}$$
$$= 15.057$$
$$\simeq 15.$$

The estimated intensity ($\hat{\lambda} \simeq 15$), shows that approximately 15 days of extreme DPED will be experienced in a year.

Table 3 provides the predictive interval estimates for the return levels of the winter, spring, summer and autumn data, respectively. The statistical inference drawn from Table 3 for instance, shows that we are 95% confident that a daily peak electricity demand of 1656.2 megawatts is likely to be reached with a lower confidence limit of 1616.9 megawatts and upper limit of 1695.5 megawatts in 20 years for the winter data. The interval estimates provide sufficient information to ESKOM, South Africa's power utility about the extreme value of the stochastic load process in time for effective planning and maintenance of the national grid system.

TABLE 3. 95% predictive interval estimates for the return levels.

Winter			
Year	LCL (2.5)%	RL	UCL (97.5%)
20	1616.9	1656.2	1695.5
30	1731.9	1801.7	1871.5
40	1813.9	1898.8	1983.8
50	1876.9	1970.9	2064.9
60	1927.6	2017.8	2128.0
70	1969.5	2074.5	2179.5
80	2005.0	2114.0	2222.9
90	2035.5	2148.0	2260.5
100	2062.1	2177.9	2293.7
Spring			
Year	LCL (2.5)%	RL	UCL (97.5%)
20	1355.5	1406.8	1458.15
30	1468.0	1545.6	1623.3
40	1547.8	1642.3	1736.8
50	1609.3	1716.4	1823.4
60	1658.9	1776.2	1893.5
70	1700.4	1826.4	1952.4
80	1735.7	1869.5	2003.3
90	1766.5	1907.3	2048.1
100	1793.6	1940.9	2088.2
Summer			
Year	LCL (2.5)%	RL	UCL (97.5%)
20	1306.4	1363.8	1421.2
30	1443.5	1529.1	1614.7
40	1533.4	1631.5	1729.5
50	1589.9	1703.4	1807.8
60	1649.5	1757.6	1865.7
70	1690.3	1800.6	1910.9
80	1724.1	1835.8	1947.5
90	1752.7	1865.4	1978.0
100	1777.4	1890.7	2003.9
Autumn			
Year	LCL (2.5)%	RL	UCL (97.5%)
20	2617.7	2639.9	2662.1
30	2716.1	2779.3	2842.5
40	2728.1	2863.3	2944.6
50	2830.6	2921.1	3011.6
60	2858.5	2964.0	3059.6
70	2899.1	2997.6	3096.0
80	2924.6	3024.7	3124.8
90	2946.3	3047.3	3148.3
100	2965.1	3066.5	3167.9

IV. DISCUSSION

The PP characterisation of extreme peak electricity demand is discussed to model the rate of occurrence of these peaks. The modelling approach incorporates extremal mixture model for threshold estimation. The penalised regression cubic smoothing splines were used to stationarise the data. High dependence was observed within the threshold exceedances. The data was declustered to keep it relatively independent using the interval method discussed in [24].

The cluster maxima of each of the non-linear detrended data sets were extracted. This was followed by fitting the stationary point process models to the cluster maxima and the parameters were estimated using the maximum likelihood estimation method. The maximum likelihood estimates of the shape parameter (ξ) are all negative for the four data sets implying that the Weibull class of distributions is a good fit for these data. That means that DPED data are bounded above. The key interest is the frequency analysis of the occurrence of extreme daily peak load in each of the four seasons, i.e. winter, summer, spring and autumn based on the calendar dates in the Southern Hemisphere. The results show that daily peak electricity demand could be experienced approximately 27, 16, 7 and 15 days per year in winter, spring, summer and autumn, respectively. This is consistent with the modelling framework of [2], [14] with the winter season recording the maximum daily peak electricity demand. This high peak value may be due to excessive use of geysers, cookers and other heating appliances during the winter season. Another

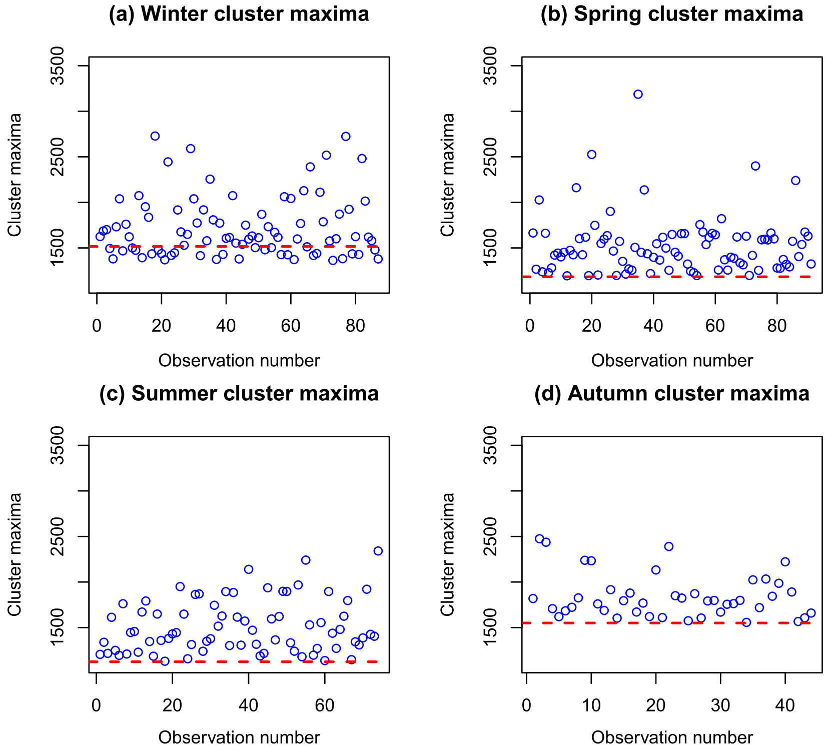


FIGURE 10. Cluster maxima for winter, spring, summer and autumn DPED data. The dashed horizontal lines represent the thresholds which are 1516, 1183, 1125 and 1183 for winter, spring, summer and autumn DPED data, respectively.

appealing contribution is the use of extremal mixture model for threshold estimation which makes redundant any major concerns about the trade-off bias-variance aspects of the POT and PP approaches.

Our modelling framework provides insightful feedback to electric power generating companies about the quantity of electricity to be reserved for the off-peak seasons on the national grid system if a reliable assessment of the extreme load levels and predictive distribution beyond the range of available data is of interest. Electricity is a key commodity used mainly as a source of energy in industrial, residential and commercial sectors. Effective monitoring of electricity demand is of great importance because demand that exceeds maximum power generated may lead to a power outage and load shedding. Consequently, decision-makers and demand-side managers should play a key role in managing the behavioural change in electricity usage, especially during peak seasons. Viable demand response strategies can be designed and implemented, wherein consumers are exposed to day time-based electricity pricing incentives.

V. CONCLUSION

This paper presented an application of the point process models in assessing the rate of occurrence of extreme peak electricity demand using South African data for the four seasons of the year. The study provided predictive interval estimates for the extremal quantiles for each data set. These

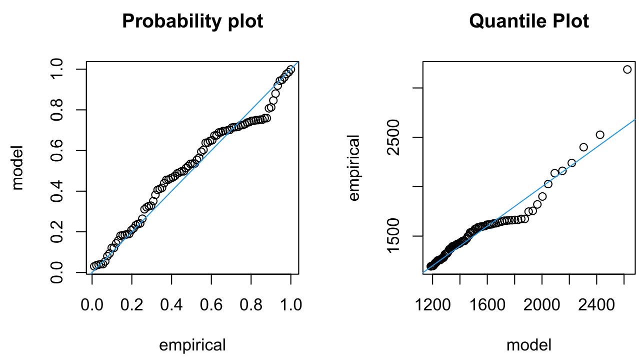


FIGURE 11. Spring DPED.

interval estimates can be helpful to ESKOM as they cater to the uncertainties surrounding the average peak loads.

This study is significant because careful modelling of daily peak electricity demand is critical to the reliable specification of power generation, distribution and maintenance in both peak and off-peak periods. Estimates of the intensity function with their associated predictive intervals of the extreme quantiles could be helpful to ESKOM in planning for future electricity generation. This is very crucial because, over a long period, management, policy makers and planners must embrace a probabilistic view of potential peak demand levels instead of point estimates only. Statistical inference based on the Autumn data set from Table 3 for instance, shows that a peak electricity demand of 3066.5 megawatts will be reached

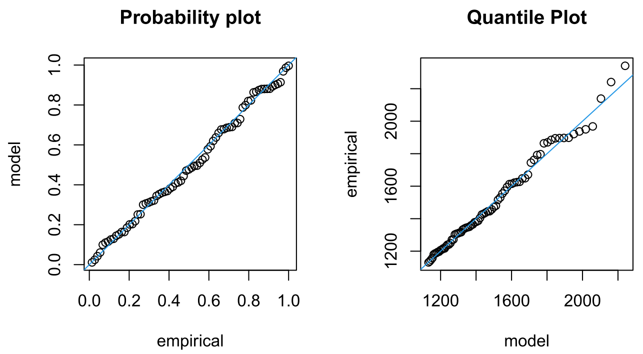


FIGURE 12. Summer DPED.

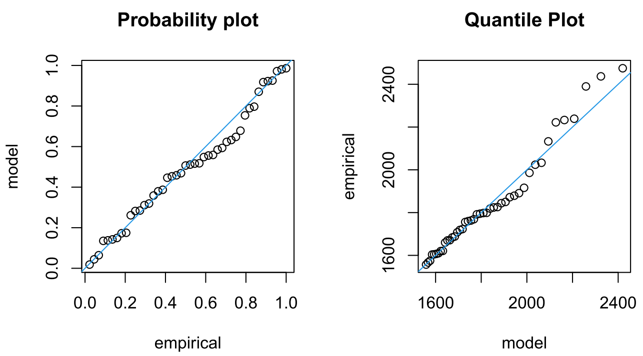


FIGURE 13. Autumn DPED.

with a lower confidence level of 2965.1 megawatts and upper limit of 3167.9 megawatts on average once in 100 years at 5% level of significance. This predictive distribution provides information about the extreme value of the stochastic load process in time which can assist ESKOM for proper planning, designing the generating and distribution system and storage devices to ensure sanity in the national grid system. The research is informative as it provides a potentially better warning and much insight into the frequency analysis of extreme peak electricity demand which can be helpful to the scientific community, power utility companies and energy-intensive enterprises for planning and forecasting during both peak and off-peak seasons.

Warning of extreme peak demand will be more reliable if future researchers could consider using the discrete-time Markov Chain (DTMC) and Bayesian estimation with informative priors to model the daily peak electricity demand. Covariates such as temperature could also be incorporated in the case of non-stationary dependent sequences as this study focused only on the stationary dependent sequences.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

APPENDIX A
FIGURES

Plot 10 shows the cluster maxima for winter, spring, summer and autumn DPED data. The plots 11 - 13 are the diagnostic plots of the stationary point process model fitted to cluster maxima of the daily peak electricity data based on the four seasons.

TABLE 4. Goodness-of-fit test statistics for the winter, spring, summer and autumn DPED data.

Winter			
Test	$\hat{\sigma}$	$\hat{\xi}$	p-value
Anderson-Darling	407.5	-0.1214	0.5927
Cramer-Von Mises	407.5	-0.1214	0.6337
Spring			
Test	$\hat{\sigma}$	$\hat{\xi}$	p-value
Anderson-Darling	461.0	-0.1430	0.5990
Cramer-Von Mises	461.0	-0.1430	0.6412
Summer			
Test	$\hat{\sigma}$	$\hat{\xi}$	p-value
Anderson-Darling	547.0	-0.3986	0.5340
Cramer-Von Mises	547.0	-0.3986	0.5412
Autumn			
Test	$\hat{\sigma}$	$\hat{\xi}$	p-value
Anderson-Darling	383.8	-0.2984	0.5706
Cramer-Von Mises	383.8	-0.2984	0.5979

APPENDIX B
TABLES

Table 4 presents formal goodness-of-fit test statistics (Anderson-Darling test and Cramer von Mises test) for testing goodness of fit for the models. The null hypothesis is that the model is a good fit to the data. In all cases the $p-$values which are given in column 3 in Table 4 are greater than the level of significance, $\alpha = 0.05$, meaning that the null hypothesis cannot be rejected. Therefore, the models are a good fit to the data.

ACKNOWLEDGMENT

The authors are indebted to Eskom, South African's power utility company for providing the data.

REFERENCES

[1] C. D and S. C, "A frequentist and Bayesian regression analysis to daily peak electricity load forecasting in south africa," *Afr. J. Bus. Manage.*, vol. 6, no. 40, pp. 10524–10533, Oct. 2012.

[2] C. Sigauke and M. M. Nemukula, "Modelling extreme peak electricity demand during a heatwave period: A case study," *Energy Syst.*, vol. 11, pp. 139–161, Feb. 2018, doi: 10.1007/s12667-018-0311-y.

[3] H. Zhang, Z. Wang, F. Gao, and D. Zhou, "A new method for the daily electricity consumption forecast of energy intensive enterprise," in *Proc. Asia–Pacific Power Energy Eng. Conf.*, Mar. 2012, pp. 1–4.

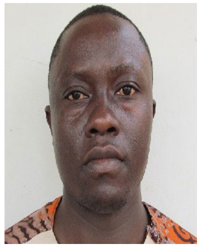
[4] P. Kou and F. Gao, "A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises," *Int. J. Electr. Power Energy Syst.*, vol. 55, pp. 144–154, Feb. 2014.

[5] C. Hor, S. Watson, D. Infield, and S. Majithia, "Assessing load forecast uncertainty using extreme value theory," in *Proc. 16th PSCC*, Jul. 2008, pp. 14–18. Accessed: Oct. 5, 2019. [Online]. Available: http://www.pscc-central.org/uploads/tx_ethpublications/pscc2008_571.pdfs

[6] C. Sigauke, A. Verster, and D. Chikobvu, "Extreme daily increases in peak electricity demand: Tail-quantile estimation," *Energy Policy*, vol. 53, pp. 90–96, Feb. 2013.

[7] C.-L. Hor, S. J. Watson, and S. Majithia, "Daily load forecasting and maximum demand estimation using ARIMA and GARCH," in *Proc. Int. Conf. Probabilistic Methods Appl. to Power Syst.*, Jun. 2006, pp. 1–6.

[8] S. Chan and S. Nadarajah, "Extreme value analysis of electricity demand in the UK," *Appl. Econ. Lett.*, vol. 22, no. 15, pp. 1246–1251, Oct. 2015.
[9] C. Sigauke, A. Verster, and D. Chikobvu, "Tail quantile estimation of heteroskedastic intraday increases in peak electricity demand," *Open J. Statist.*, vol. 2, no. 4, pp. 435–442, 2012.
[10] J. Pickands III, "The two-dimensional Poisson process and extremal processes," *J. Appl. Probab.*, vol. 8, no. 4, pp. 745–756, Dec. 1971.
[11] R. L. Smith, "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone," *Stat. Sci.*, vol. 4, no. 4, pp. 367–377, Nov. 1989.
[12] S. Coles, *An Introduction to Statistical Modelling of Extreme Values*. London, U.K.: Springer-Verlag, 2001.
[13] F. Beichelt, *Stochastic Processes in Science, Engineering and Finance*. London, U.K.: Chapman & Hall, 2006.
[14] C. Sigauke, S. Kumar, N. Maswanganyi, and E. Ranganai, "Reliable predictions of peak electricity demand and reliability of power system management," in *Book Chapter in System Reliability Management Solutions and Technologies*. Boca Raton, FL, USA: CRC Press, 2019.
[15] M. Jacob, C. Neves, and D. V. Greetham, *Forecasting and Assessing Risk of Individual Electricity Peaks*. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-28669-9.
[16] Y. Li and B. Jones, "The use of extreme value theory for forecasting long-term substation maximum electricity demand," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 128–139, Jan. 2020, doi: 10.1109/TPWRS.2019.2930113.
[17] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Math. Proc. Cambridge Phil. Soc.*, vol. 24, no. 2, pp. 180–190, Apr. 1928.
[18] M. R. Leadbetter, "Extremes and local dependence in stationary sequences," *Zeitschrift Für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 65, no. 2, pp. 291–306, 1983.
[19] R. Karr, *Point Processes and their Statistical Inference*. New York, NY, USA: Marcel Dekker, 1991.
[20] S. I. Resnick, *Extreme values, Regular Variation and point Processes*. New York, NY, USA: Springer, 2013.
[21] A. MacDonald, C. J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell, "A flexible extreme value mixture model," *Comput. Statist. Data Anal.*, vol. 55, no. 6, pp. 2137–2157, Jun. 2011.
[22] C. Sigauke and A. Bere, "Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand," *Energy*, vol. 119, pp. 152–166, 2017.
[23] Y. Wang, *Smoothing Splines: Methods and Applications*. London, U.K.: Chapman & Hall, 2011.
[24] C. A. T. Ferro and J. Segers, "Inference for clusters of extreme values," *J. Roy. Stat. Soc., Ser. B, Stat. Methodol.*, vol. 65, no. 2, pp. 545–556, May 2003.
[25] R. L. Smith, "Statistics of extremes with application in environment, insurance and finance," in *Extreme Values Finance, Telecommunications, and the Environment*. Boca Raton, FL, USA: CRC Press, 2004, pp. 12–89.



JERRY BOANO-DANQUAH received the Diploma degree in education from the University of Cape Coast, Ghana, the B.Sc. degree in mathematics from the Kwame Nkrumah University of Science and Technology, Ghana, and the B.Sc. (Hons.) and M.Sc. degrees in statistics from the University of Venda. He is currently pursuing the Ph.D. degree in mathematical statistics with the University of the Witwatersrand, Johannesburg. His research interests include statistics of extreme, forecasting, and time series analysis. He is a member of the South African Statistical Association (SASA).

CASTON SIGAUKE received the M.Sc. degree in operations research from the National University of Science and Technology, Zimbabwe, and the Ph.D. degree in statistics from the University of the Free State. He has been a Chartered Statistician with the Institute of Certificated and Chartered Statisticians of South Africa (ICCSSA), since 2013. He is currently a Senior Lecturer with the Department of Statistics, University of Venda, South Africa. His research interests include forecasting and time series, statistics of extremes, statistical learning and modeling, and probabilistic load and renewable energy (solar and wind) forecasting, including optimization of grid integration of renewable energies. He is a member of the International Statistical Institute, the International Institute of Forecasters (IIF), the Operations Research Society of South Africa (ORSSA), and the South African Statistical Association (SASA).

KWABENA A. KYEI received the B.Sc. degree in mathematics and statistics from the Kwame Nkrumah University of Science and Technology, the M.Sc. degree in demography from the Universite Catholique de Louvain, Belgium, the Ph.D. degree in actuarial/biostatistics from the University of Louisville, USA, and the Ph.D. degree in demography from the University of Pretoria. He was a Manager with the Demographic and Health Survey, Human Sciences Research Council, from 1992 to 1994, a large survey that covered over 30 000 households countrywide. He is currently an External Examiner with some universities. As a Statistician/Demographer, he was involved with research projects in primary health care and population issues. He has published about 40 articles in accredited international journals and over 15 articles in non-accredited journals. His research interests include econometrics and banking. He serves as an Associate Editor for some journals. He has reviewed articles of some international journals.