Energy 119 (2017) 152–166

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand

Caston Sigauke*, Alphonce Bere

Department of Statistics, University of Venda, Private Bag X5050, Thohoyandou, 0950, Limpopo, South Africa

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form
2 December 2016
Accepted 6 December 2016

Keywords:
Extreme value theory
Non stationary time series
Peak electricity demand
Penalized smoothing splines
Time varying threshold

ABSTRACT

Long term peak electricity demand forecasting is a crucial step in the process of planning for power transmission and new generation capacity. This paper discusses an application of the Generalized Pareto Distribution to the modelling of daily peak electricity demand using South African data for the period 2000 to 2010. The main contribution of this paper is in the use of a cubic smoothing spline with a constant shift factor as a time varying threshold. An intervals estimator method is then used to decluster the observations above the threshold. We explore the influence of temperature by including it as a covariate in the Generalized Pareto Distribution parameters. A comparative analysis is done using the block maxima approach. The GPD model showed a better fit to the data compared to the GEVD model. Key findings from this study are that the Weibull class of distributions best fits the data which is bounded from above for both stationary and non-stationary models. Another key finding is that for different values of the temperature covariate the shape parameter is invariant and the scale parameter changes for different values of heating degree days.

1. Introduction

Planning for new generating capacity and also for reserve margins requires accurate long term peak electricity demand forecasts. Long-term peak electricity demand forecasting has not received the same attention in literature compared to short-term peak electricity demand forecasting. However there are a few notable contributions. A semi-parametric additive model for long term peak electricity demand is developed in Ref. [21]. The authors use the model to forecast the probability distribution of peak electricity demand. In Ref. [23] a regression based model for long-term peak load forecasting for a small utility in Cyprus is presented. The developed model incorporates important drivers of electricity demand. Results from this study show that the developed model has high predictive power. The use of high resolution electricity demand data in long term probabilistic forecasting is proposed in Ref. [33]. Using a case study it is shown from empirical results that high resolution models outperform low resolution models for both probabilistic monthly peak and energy load forecasting. A detailed discussion of some guidelines on methods for probabilistic forecasting with applications to electricity price forecasting, energy, wind and solar forecasting are presented in Ref. [26]. In Ref. [26] a wide range of tests for assessing reliability which they define to be “the statistical consistency between the distributional forecasts and the observations” (some of the tests being the Kupiec, Christoffersen, probability integral transform, Berkowitz tests among others), and methods for measuring sharpness which they define to be “the concentration of the predictive distributions” (some of these methods are the proper scoring rules, pinball loss function, Winkler score, continuous ranked probability score, among others).

One of the major drivers of electricity demand is temperature and is known to cause seasonal variation in electricity demand [21]. Modelling extreme temperature using extreme value theory (EVT) is discussed in detail in literature, but the use of EVT in modelling peak electricity demand has not received much attention. A few notable contributions in this area include [19] who discuss an application of EVT in modelling maximum load forecast errors which they use for given acceptable levels of risk to predict electricity demand and also [27] who use of the Generalized Pareto Distribution (GPD) for predicting extreme daily increases in peak electricity demand. The authors carry out a comparative analysis

* Corresponding author.
E-mail address: csigauke@gmail.com (C. Sigauke).

http://dx.doi.org/10.1016/j.energy.2016.12.027
0360-5442/

with a Generalized Single Pareto Distribution (GSPD). The results show a good fit for both models. Policy implications are then discussed. In a related study [9], model the effect of temperature below a reference temperature which separates the weather neutral period from the winter sensitive period on average daily electricity demand using the generalized extreme value distribution (GEVD). Empirical results from this study show that the gradual decrease of temperature from 18° C converges to 4.6° C while the marginal increases in average daily electricity demand converges to 1.58 MW. The authors argue that this modelling approach assists system operators in scheduling and dispatching of electrical energy. A stationary GEVD was used in this study.

Modelling of non-stationary sequences using extreme value theory with time varying parameters and thresholds is discussed and applied in the field of environmetrics [3]. A point process approach to modelling non-stationary series with an application to ozone data is used in Ref. [28]. Empirical results in this study give values of the shape parameter which are positive and not close to zero as is the general case for modelling such data. The study did not incorporate covariates. An application of non-stationary extremes using temperature is discussed in Ref. [25]. In this study, block maxima and peaks over threshold (POT) models are used. A time varying threshold which depends on both the scale and shape parameters is used in the POT model. The temperature data is initially centered and normed which enables the retrieval of the underlying trend. In a similar study [6] critically discuss the relevance of asymptotic theory to applications. They argue that for near-independent models, the extremal index requires a detailed investigation. However they agree that for asymptotically dependent data the extremal index beyond time series is well covered in the literature.

Non-stationary time series models for predicting inflows of water into Dez dam reservoir are presented in Ref. [32]. A comparative analysis of the models shows that the dynamic autoregressive artificial neural network outperforms all the other models used in the study. The modelling of non-stationary extremes with an application to surface level ozone is discussed in Ref. [10]. The authors propose a two stage approach to modelling the non-stationary data. Initially they use preprocessing methods which are then followed by standard methods of extreme value theory to the preprocessed data. Results from this study show that using preprocessing methods gives a better model fit compared to direct use of extreme value theory models.

In a recent study [17] use GEVD with time varying parameters to model non-stationary extreme wave heights in some selected locations of the Greek sea. The GEVD parameters are estimated using the conditional density network. Results from this case study show that the Greek coastal areas are at high risk of extreme high waves. An application of the GPD with time varying parameters to the modelling of electricity demand in the United Kingdom is given in Ref. [5]. The authors estimate the value at risk of electricity demand.

It is important to carry out regular assessments of the frequency of occurrence of extreme peak electricity demand for ensuring the stability of the grid [19]. The use of regression quantile models in the modelling of extreme wave height distributions is proposed in Ref. [24]. Data from two Portuguese locations is used. The proposed regression quantile models used are the 3-parameter Weibull, the generalised extreme value (GEV) and the generalised Pareto (GP). Results from this study show that the GP regression quantile model gives the best results when estimating the 50 and 100 year return levels.

Modelling of South African daily peak electricity demand data is discussed in literature [7,27]. This study focuses on the modelling of daily peak electricity demand (DPED) data from South Africa's power utility company, Eskom using extreme value theory distributions. The GPD is fitted to DPED above a sufficiently high time varying threshold after which we estimate extreme peak demand using the k-period return level quantile function. The thrust in this paper is in modelling time series extremes which requires the modelling of the upper tail of a distribution. Modelling extreme peak electricity demand helps decision makers in power utility companies in planning the scheduling and dispatching of electrical energy including long term planning for capacity expansion. The main contribution of this paper is in the use of a cubic smoothing spline with a positive shift factor in determining a sufficiently high time varying threshold. We estimate the shift factor using the extremal mixture models discussed in Ref. [31]. The excesses above the threshold are declustered using the intervals estimator method discussed in Ref. [11]. A GPD is then fitted to cluster maxima. This is followed by the inclusion of temperature as a covariate in the GPD parameters. The effectiveness of this modelling approach is shown through a simulation study. A comparative analysis is done with a GEV regression quantile model.

The rest of the paper is organized as follows. A discussion of the modelling of non-stationary time series using the generalised Pareto family of distributions is given in Section 2. Section 3 presents a discussion of the generalised extreme value family of distributions. The predictive performance of the EVT models is discussed in Section 4 while a simulation study is presented in Section 5. Empirical results are presented in Section 6 while Section 7 concludes.

2. Modelling non-stationary time series using the generalised Pareto distribution

The generalized Pareto distribution (GPD) is a peaks-over-threshold model normally used in modelling exceedances above a sufficiently high threshold. If the time series data is available, the GPD is usually used instead of the generalised extreme value (GEV) distribution [4] which uses only one observation in a block. This results in loss of important information from other observations in the block.

Let Y_t denote daily peak electricity demand (DPED) on day t with associated time varying covariates X_t, for $t = 1, \ldots, n$, where n is the number of observations. The objective is to estimate extreme quantiles of Y_t above a high time varying threshold, $\tau(t)$. These extreme quantiles are denoted by $y_{k,t}$ and are conditional on the covariates X_t and $P(Y_t > y_{k,t} | X_t = x_t)$ is the tail probability above the quantile $y_{k,t}$. On average the high quantile $y_{k,t}$ is exceeded approximately once every $\frac{1}{k}$ observations [10]. The observations y_t above $\tau(t)$ are assumed to follow a GPD, i.e.

$$Y_t \sim \mathrm{GPD}(\sigma(x_t), \xi(x_t)) \tag{1}$$

where $\sigma(x_t)$ and $\xi(x_t)$ are the scale and shape parameters respectively which depend on the time varying covariate x_t. The distribution function is given by

$$W(y_t) = 1 - \left(1 + \frac{\xi(x_t)(y_t - \tau(t))}{\sigma(x_t)}\right)^{-\frac{1}{\xi(x_t)}} \tag{2}$$

where $\xi(x_t) \neq 0$, y_t is the DPED data and the parameters are modelled as functions of the covariate x_t. In order to ensure that the scale parameter is always positive we reparametrise it as $\theta = \log(\sigma(x_t))$.

Electricity demand data exhibit strong seasonality and are highly sensitive to temperature fluctuations. To account for the seasonality and temperature effects the model specification is given as: