2014 American Control Conference (ACC)
June 4-6, 2014. Portland, Oregon, USA

Risk Adjusted Forecasting of Electric Power Load*

Saahil Shenoy[1] and Dimitry Gorinevsky[2]

Abstract— **Load forecasting of energy demand is usually focused on mean values in related statistical models and ignores rare peak events. This paper provides Extreme Value Theory analysis of the peak events in electrical power load demand. It estimates risk of the peak events by combining forecast of the mean with extreme value modeling of distribution tail. The approach is demonstrated for electric load demand data for a US utility. The problem is to find the forecast margins that keep the risk of demand exceeding forecast plus the margin to one event per year. The long tail model of the peak events is more accurate and yields 50% larger margin compared to the normal distribution model. These results show that the long tail behavior of the forecast errors must be taken into account when trying to keep outage risk low.**

I. INTRODUCTION

Forecasting and management of peak loads is the key to energy efficiency of the power grid and to carbon reduction through renewables. To accommodate for the peak loads, current power grid capacity is three times larger than the average load. The technological and operational changes introduced by the non-traditional generation and distribution increase the load-power balance variability and the risk of large forecast errors.

The load demand substantially exceeding the scheduled supply might lead to an outage. Utilities must schedule supply a day in advance in the electric power markets. The forecasted supply must take into account the risk of exceeding the day-ahead schedule, e.g., see [1], [2]. This paper is focused on forecast adjustment needed to control the risk. The underlying models are trained on historical data. This simplifies practical application of the approach.

Practical load forecasting is usually focused on mean values and relies on least squares or other related statistical methods for fitting predictive models to the historical data. In research literature, several machine learning techniques, such as neural networks, have been tried to forecast the power load [5], [7]. Other techniques used in the literature, including [6], [7], involve versions of ARMAX (auto-regressive moving average with exogenous inputs) model. The mentioned forecasting approaches are based on least squares model fit, which implies normal distribution of the forecast errors. They forecast the distribution mean and they do not address outliers. The training data outliers that do not follow the normal distribution for the model errors can bias the model.

*This work was supported by a Seed Grant from TomKat Center for Sustainable Energy at Stanford University

[1]Saahil Shenoy is a PhD student in the Department of Physics, Stanford University, Stanford, CA 94305, USA saahils@stanford.edu

[2]Dimitry Gorinevsky is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA gorin@stanford.edu

To make the fitted model more accurate, robust statistics approaches could be used in order to eliminate outliers from skewing the model fit to the bulk of the data [9]. This can be done in different ways. One technique is the Huber regression [10], which is a convex optimization approach with greatly reduced outlier sensitivity compared to standard least squares regression. Another technique, which we apply in this paper, is robust statistics update that iteratively removes outliers until convergence is reached [11].

In the day ahead forecast, the outliers create the risk of exceedingly large forecasting errors that could require the utility to buy the missing supply at very high spot prices or, ultimately, to disconnect a part of the customers. The large forecast errors can be described through tail distribution models. The branch of statistics describing rare peak events and the tails of the probability distributions is known as extreme value theory (EVT), see [12]. According to EVT, the distribution tails are often long, follow power laws and decay much slower than the normal distribution. With long tails, the extreme event risk is much higher than would be expected from a normal distribution model.

The risk can be estimated from the past data by combining the forecast of the mean with modeling of the peaks as extreme value forecast errors. To the authors' best knowledge, this has not been done earlier for power load data. In recent paper [13], the long tail distribution is used to analyze the exceedances of the power load demand, but not in conjunction with forecasting.

Finance and actuarial science have many needs for the estimation of risk. For instance, in finance there is interest in forecasting the risk associated with lending money to consumers based on credit [4]. Actuarial applications, such as insurance, involve looking at the long tail distributions of losses [3]. Both finance and actuarial applications are related to this work.

The contributions of this paper are as follows. First, this paper addresses the gap in the power load forecasting literature by developing a method that combines the forecast of the mean with modeling of the distribution tail to evaluate the peak event risk. A robust estimation approach is used for building a regression model for the mean forecast. The tail distribution for the deviation from the mean is modeled in accordance with EVT as Pareto Distribution (power law) that decays much slower than the normal distribution.

Second, the paper applies the developed method to forecasting of energy load demand using IEEE data set for a US utility, see [15]. We formulate a linear regression model of the hourly load demand with 58 nonlinear regressors that depend on temperature, load values, and time. We fit a long

978-1-4799-3274-0/$31.00 ©2014 AACC

tail model of the peaks to the data for modeling the risk of extreme deviation from the regression forecast.

Third, the paper presents the results of applying the developed model to the risk adjusted power load forecasting. We show that when the long tails are taken into account, the risk of extreme events, such as unusually high loads leading to power outages, is much higher than with models based on normal distribution. Taking the long tails into account, load forecast margins might need to be larger by as much as 50% compared to the normal distribution model.

II. General Problem Formulation

We will start by formulating the problem in general terms. The following sections apply this general formulation to the power load data forecasting. We use log power for forecasting, hence, the Pareto distribution (power law) distribution of the tails becomes exponential distribution.

A. Main Problem

We consider a dataset

$$D = \{X_i, y_i\}_{i=1}^{N}, \tag{1}$$

where scalars y_i are dependent variables (the load) and vectors $X_i \in \Re^n$ are explanatory variables (regressors). Index i describes the time sample and N is the number of samples available.

The forecasting model assumes that data (1) was generated by an underlying conditional distribution with known parametric form

$$y \sim p(\theta|X) \tag{2}$$

The forecasting problem is then to estimate distribution (2) (parameter θ) from data (1) and compute a risk adjusted forecast $y_m = \phi(\theta, r, X)$ such that

$$P(y > \phi(\theta, r, X)) = r, \tag{3}$$

We will call r the exceedance risk probability or just 'risk'.

This paper considers a model of distribution (2) that is a mixture of a normal distribution, with probability $1-q$, and an exponential distribution with probability q

$$\begin{aligned} y &= \beta^T X + (1-z)v_n + z v_e, & (4)\\ z &\sim B(1,q), & (5)\\ v_n &\sim N(0,\sigma^2), & (6)\\ v_e &\sim Exp(\lambda), & (7) \end{aligned}$$

where $B(1,q)$ is the binomial distribution with $\{0,1\}$ outcomes, $\beta \in \Re^n$ is regression parameter vector, σ is the standard deviation of the normal distribution, and λ is the rate parameter of the exponential distribution.

Exponential Gaussian mixture distribution (4)–(7) describes $p(\theta|X)$ in (2) with parameter vector $\theta = \text{col}\{\beta, \sigma, q, \lambda\}$. Distribution (4)–(7) models both the normal and tail behavior of observed data. Section III of this paper discusses how the electric load forecasting problem could be described by the model (4)–(7).

In what follows, we assume that the tail intensity q is a small parameter, $q \ll 1$. Then, for $v = y - \beta^T X$ small, we approximately have $v \sim N$ (central distribution (6)). For v large, the exponent dominates the gaussian and we approximately have $v \sim Exp$ (tail distribution (7)). The described approximations separately consider the central and the tail distribution parts and are the basis of the proposed approach to estimating distribution (2). Subsection IV-C discusses how the risk adjusted forecast $y_m = \phi(\theta, r, X)$ in (3) can be computed from the estimated model.

B. Estimating Mean

For a moment, let us assume zero tail intensity, $q = 0$. With $q \ll 1$, this approximation closely holds in the central part of the distribution (4)–(7). Consider a subset of the data points belonging to the central part of the distribution

$$\mathcal{C} = \{j_1, \ldots, j_K\} \tag{8}$$

In that case, we need to estimate the model $y \sim N(\beta^T X, \sigma^2)$ from data (1). By introducing data matrices

$$\begin{aligned} \bar{y}_{\mathcal{C}} &= [y_{j_1} \ \ldots \ y_{j_K}] \in \Re^{1,K} & (9)\\ \bar{X}_{\mathcal{C}} &= [X_{j_1} \ \ldots \ X_{j_K}] \in \Re^{n,K} & (10) \end{aligned}$$

the well known Maximum Likelihood Estimates (MLE) for regression parameter vector β and standard deviation σ are

$$\begin{aligned} \hat{\beta} &= (\bar{X}_{\mathcal{C}} \bar{X}_{\mathcal{C}}^T)^{-1} \bar{X}_{\mathcal{C}} \bar{y}_{\mathcal{C}}^T & (11)\\ \hat{\sigma} &= K^{-\frac{1}{2}} \|\bar{y}_{\mathcal{C}} - \hat{\beta}^T \bar{X}_{\mathcal{C}}\|_2 & (12) \end{aligned}$$

where $\|\cdot\|_2$ is the Euclidean norm of a vector. The regression parameter estimates (11), (12) depend on the data subset (8).

The mean forecast $\hat{y}$ can be computed from the estimated regression parameter. For a given regressor vector X, the mean forecast is

$$\hat{y} = \hat{\beta}^T X \tag{13}$$

The described regression model and the forecast (13) are how the forecasting is done in most of the existing literature on time series forecasting in general and power load in particular. The regressor choices may vary. The described forecasting approach is oriented towards the mean load and does not specify the risk that actual load will exceed the forecasted value.

C. Estimating Tail

As discussed in Section II-A, for large $v = y - \beta^T X$, the exponential (long tail) distribution v_e dominates the normal distribution v_n. To model the tail, we consider a subset of the data points $\mathcal{T}$ that belong to the tail.

$$\mathcal{T} = \{j_1, \ldots, j_M\} \tag{14}$$

Selecting the subset $D_{\mathcal{T}}$ is discussed in Section IV-B.

Our approach to modeling the tail, follows the Peaks Over Threshold Method of EVT, see [12], in spirit. Consider the exceedance data set

$$\mathcal{T} = \{j : y_j - \beta^T X_j > a\}, \tag{15}$$

where a is the exceedance threshold.

In accordance with (4), $y - \beta^T X = (1-z)v_n + zv_e$. For a high enough, the probability of $v_n > a$ is much lower than the probability of $v_e > a$. Therefore, it is a reasonable approximation to assume that for each $j \in \mathcal{T}$, we have $z_j = 1$ and $y_j - \beta^T X_j = v_{e,j}$. Since $v_e \sim \exp$ in (7), the exceedances $v_{e,j} - a$ follow the same exponential distribution.

The above reasoning suggests that an MLE estimate of λ from the data samples $\mathcal{T}$ is given by

$$\hat{\lambda} \quad = \quad \left[\frac{1}{M} \sum_{j \in \mathcal{T}} (y_j - \beta^T X_j - a) \right]^{-1}, \tag{16}$$

where $M = \text{card}\,\mathcal{T}$ is the cardinality of (number of unique elements in) the set $\mathcal{T}$. In the special case of $\beta^T X_j = 0$, $y = \log L$, and $a = \min_{j \in \mathcal{T}} y_j$, the MLE estimate (16) yields the well-known Hill's estimator described in [12].

Once the estimate of λ is available, q in (4) can be estimated from the survival function of the tail distribution

$$F(m;\theta) = P(y - \beta^T X > m) \approx q e^{-\lambda m}, \tag{17}$$

This formula is approximate because it ignores the tail of the central (normal) distribution. It is obtained by integrating distribution density (7) from m to infinity. With N total data points, and M points where $y_j - \hat{\beta}^T X_j$ exceeds the threshold a, we can estimate $F(m;\theta)$ as M/N. By substituting this estimate into (17), given $\hat{\lambda}$, we can estimate q as

$$\hat{q} = \frac{M}{N} e^{\hat{\lambda} a} \tag{18}$$

The proposed long tail model can be used for estimation and control of the outage risk. As a baseline for comparison, we will also consider risk modeling through the tail of normal distribution (6) by assuming $q = 0$ in (4). Similar to (17), the tail of the normal distribution can be described by a survival function $F_n(m;\theta) = P(y - \beta^T X > m)$

$$F_n(m;\theta) = \frac{1}{2} - \frac{1}{2} \,\text{erf}\, \frac{m}{\sqrt{2}\sigma}, \tag{19}$$

where erf is the Error Function. The normal distribution is a special case of (4) with $q = 0$ and any λ; therefore, in (19) we have $\theta = \text{col}\{\beta, \sigma, 0, 0\}$. The tail of normal distribution is commonly used for Value At Risk modeling in finance.

III. Power Load Forecasting Problem

We used data set from [15]. It includes hourly loads and ambient temperature data for an anonymous US utility. The data in the set are for 20 zones served by the utility. The methodology described above was applied to the aggregate load across all these zones. The range of the aggregate load is 0.8 to 3.2GW, with the average value being 1.6GW. The data covers a time range of approximately 4 years with sampling interval of one hour. Figure 1 shows a 601-hour segment of the 38,070 hour aggregated load data set.

A. Linear Regression Forecast Model

Let $L(t)$ be the load demand at time sample t. The data is sampled every hour and t is the number of hours elapsed since the start of the data collection. We use logarithmic load, normalized by L_0 =1GW, as dependent variable y

$$y_t \quad = \quad \log(L(t)/L_0), \tag{20}$$

The regressor set that we use in the forecasting model is a modified version of the regressors described in [14]. The regressors can be described as follows

$$X_t = [1, t, DD(t), C(t), HOL(t), L(t-24)/L_0]^T \tag{21}$$

The 58 components of the regressors vector X_t are

$$\begin{aligned} DD(t) &= [HDD(t), CDD(t)] & (22)\\ C(t) &= [D(t), M(t), H(t)] & (23)\\ HOL(t) &= [CH(t), CH_1(t), PP(t)] & (24)\\ HDD(t) &= [HDD_0(t), ..., HDD_5(t)] & (25)\\ CDD(t) &= [CDD_0(t), ..., CDD_5(t)] & (26)\\ HDD_i(t) &= \max\{F_{ref}(t) - F_i(t), 0\} & (27)\\ CDD_i(t) &= \max\{F_i(t) - F_{ref}(t), 0\} & (28) \end{aligned}$$

The regressors are defined as follows:

Regressors $F_i(t)$ in (27) and (28) have the meaning of average temperatures over the previous 2^{i-1} hours, for $i = 1, ..., 5$, and for the current hour, $i = 0$. More precisely

$$F_i(t) = \begin{cases} F_{temp}(t) & i = 0 \\ F_{temp}(t-1) & i = 1 \\ \frac{1}{2^{i-1}} \sum_{j=t-2^i+1}^{t-2^{i-1}} F_{temp}(j) & i = 2, ..., 5 \end{cases}, \tag{29}$$

where $F_{temp}(t)$ is the ambient temperature in degrees F at time t. We take $F_{temp}(t) = 0$ for $t < 0$. Using (29) reduces the number of regressors by averaging contributions of increasingly dated inputs $F_{temp}(t)$ over increasingly longer intervals.

Regressors $HDD_i(t)$ in (27) and $CDD_i(t)$ in (28) are known as the heating and cooling degree days, respectively. F_{ref} is the reference temperature where the temperature influence in minimized. To find F_{ref}, a cubic polynomial is fitted to load demand data taken vs temperature. The minimum of this polynomial is acieved at F_{ref}, see [14].

Regressor $D(t)$ is a vector with components $D_i(t)$. Index i is the day of the week and t is the hour. One day is left

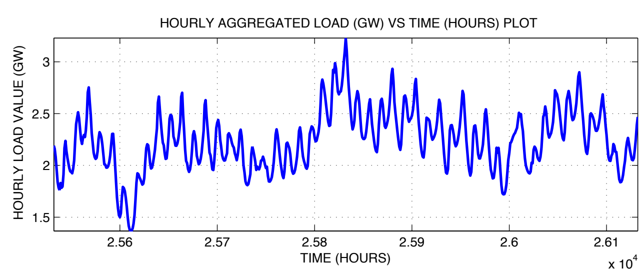

Fig. 1. Aggregate load value (GW) for a 601 hour interval.

out as a baseline and $i = 1, ..., 6$; $D_i(t) = 1$ if hour t is in day i and 0 otherwise.

Regressor vector $M(t)$ has components $M_i(t)$, ($i = 1, ..., 11$), where i is the month number. One month is left out; $M_i(t) = 1$ if hour t is in month i and 0 otherwise.

Regressor vector $H(t)$ has components $H_i(t)$, ($i = 1, ..., 23$), where i is the hour of the day. One hour has been left out.

$$H_i(t) = \begin{cases} 1 & \text{if } t \bmod 24 = i \\ 0 & \text{otherwise} \end{cases} \qquad (30)$$

The regressors in (24) are: $CH(t) = 1$ if hour t is in a fixed holiday (e.g. Christmas, Independence Day, etc.) and 0 otherwise; $CH_1(t) = 1$ if hour t is in the day after a fixed holiday and 0 otherwise; $PP(t) = 1$ if hour t is in the last day of the year and 0 otherwise.

Regressor $L(t - 24)$ is the energy load a day before the prediction time; we assume that $L(t) = 0$ for $t < 0$. Using $L(t - 24)$ and assuming that an accurate 24-hour-ahead forecast of ambient temperature $F_{temp}(t)$ is available provides the day ahead forecasting.

B. Outage Risk

The outage risk r in (3) is a small probability parameter. A more meaningful interpretation of r follows by relating it to n_{out}, the expected number of outliers (outages) that will occur in a year time frame.

In our probabilistic model, the residuals $y_t - \beta^T X_t$ are independent and identically distributed. Therefore the expected number of outliers can be computed as

$$n_{out} = r \times N_h, \qquad (31)$$

where $N_h = 8760$ is the number of hours in a year. In the power load data set example, we assume $n_{out} = 1$, i.e., one outage per year, and get the risk parameter $r = 1.14 \cdot 10^{-4}$.

IV. Algorithm

This section presents an algorithm for estimating parameters of the mixture model (4)–(7) from data (1) and for computing the risk adjusted forecast (3). The algorithms are illustrated by applying them to the logarithmic power load data set described in the previous section.

The algorithms outlined here start from the central (normal) part of the data. We use a robust statistics approach for fitting a model to the data while ignoring outliers. After the central part of the distribution is fitted, the tail model is fitted to the threshold exceedance data.

A. Robust Regression Estimation

Method for implementing linear regression that mitigates effect of outliers is an iterative approach. We initialize (8) to $\mathcal{C} = \{1, \ldots, N\}$. At each algorithm iteration, the model fit parameters $\hat{\beta}$, $\hat{\sigma}$ are computed from (8)–(12). Then, the set $\mathcal{C}$ (8) is updated by rejecting the model fit outliers as $\mathcal{C} = \{t : |y_t - \hat{\beta}^T X_t| \leq c \cdot \hat{\sigma}\}$. The iterations have converged when the set $\mathcal{C}$ stops changing.

The described robust regression estimation algorithm was applied to the logarithmic power load data set described in the previous section. Figure 2 shows the distribution of the residuals $v_i = y_i - \hat{\beta}^T X_i$ obtained after fitting the linear regression described in the previous section. The estimated 58-component regression parameter vector $\hat{\beta}$ is not presented here because of the space limitations. The estimated standard deviation is $\hat{\sigma} = 0.0584$ and describes the width of the bell-shaped curve in the histogram of Figure 2.

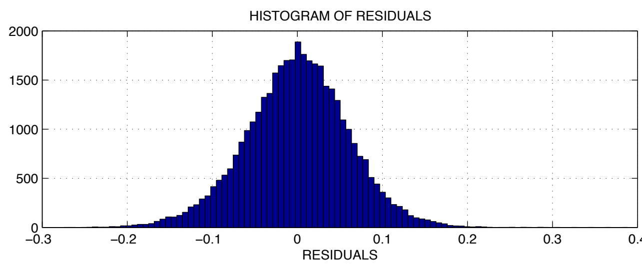

Fig. 2. Histogram of residuals for the robust regression fit.

B. Tail Model Estimation

In accordance with the model (4)–(7), the exceedances of the residual $v_i = y_i - \hat{\beta}^T X_i$ over a large threshold a are described by the exponential distribution (7). (The exceedances of the exponential distribution are distributed exponentially.) This distribution describes the logarithmic load data (20). By exponentiating and returning to the raw load data we get a Pareto Distribution. This is consistent with long tail model for the load demand data considered in [13] and is as prediced by EVT.

The tail model is fitted to the tail data set D indexed by $\mathcal{T}$, in accordance with (16), where β is replaced by the robust regression estimate $\hat{\beta}$.

The fit of the exponential distribution is illustrated in Figure 3. The tail threshold was selected as $a = 4\hat{\sigma}$. The plot of the normal distribution part of the mixture (4)–(7) in Figure 3 is within the line thickness of the horizontal axis. This justifies the assumption of the exponential distribution dominance used in the tail modeling.

C. Risk Adjusted Forecast

Model (2) described by (4)–(7) can be used to solve problem (3) by computing m such that $P(y - \beta^T X > m) = r$. Margin value m describes how much extra power on average needs to be reserved for the next day above the

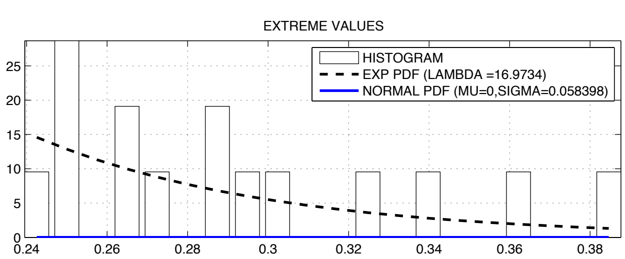

Fig. 3. Tail model fit for the histogram of residuals.

forecast (13). Once m is found then the risk adjusted forecast is $y_m = \phi(r, X, \theta) = \beta^T X + m$. For given risk probability r, margin m can be computed from (17) as $m = F^{-1}(r; \theta)$.

By solving $F(m; \theta) = r$ for m in (17), we get $m = \lambda^{-1}(\log q - \log r)$. The risk adjusted forecast (3) can be computed from this and (13) as

$$y_m = \hat{\beta}^T X + \hat{\lambda}^{-1}(\log \hat{q} - \log r) \tag{32}$$

The right hand side of (32) defines function $\phi(r, X, \hat{\theta})$ in (3). This expression is only valid for small risk r, specifically for $r \ll q$. In the load data example of the previous section this holds since $r = 1.14 \cdot 10^{-4}$ and $\hat{q} = 0.02$.

As baseline for comparison of the risk adjusted forecast (32) we will consider the risk based on the normal distribution tail model. By inverting the survival function in (19) we get the baseline risk adjusted forecast for the normal model

$$y_n = \hat{\beta}^T X + \sqrt{2}\hat{\sigma} \cdot \mathrm{erf}^{-1}(1 - 2r) \tag{33}$$

D. Algorithm Verification

Data set (1) with $N = 40,000$ data points was generated by simulating a mixture of Gaussian and exponential distribution (4)–(7). In the simulation, the regression vector X was empty and we assumed the zero mean, $\beta^T X = 0$. Simulation parameters were picked roughly corresponding to the real power load data example in this paper: standard deviation of the normal distribution $\sigma = 0.05$, exponential tail inverse length $\lambda = 15$, and tail intensity $q = 0.05$.

Parameter σ was estimated from robust regression with the threshold parameter $c = 3$. To estimate λ, (16) was used with the threshold $a = 4\hat{\sigma}$. To estimate q, (18) was used. Table I shows these estimates versus the ground truth parameters in the simulation. The accuracy is quite good.

V. Power Load Forecasting Results

The algorithms of Section IV were applied to the power load data and the forecasting problem described in Section III. A linear regression model was fitted to the data as described in Section IV-A. The time series plot comparing actual load demand data (in GW) with load obtained from the regression model fit (13) is shown in Figure 4 and covers about 300 hours of the data. In accordance with (20), the modeling used logarithmic loads. The plotted data are converted back to the engineering units (exponentiated).

To fit the tail model, the methodology of Section IV-B was used. Parameter estimates are given in Table II. Figure 3 shows the exponential distribution fit for the tail; the normal distribution tail does not match the data. The exponential and normal distribution fit are further compared in the QQ plot of Figure 5. A QQ (Quantile-Quantile) plot is a statistics tool commonly used in EVT to evaluate the extrapolation power of the statistical model for the tail. The theoretical inverse cumulative distribution function (CDF) predicted by the model is plotted vs the empirical inverse CDF. The theoretical inverse CDF for the long tail distribution in Figure 5 is related to survival function (17), which is complementary to the CDF. The theoretical CDF for the normal distribution is complementary to survival function (19). The long tail model data are much closer to the diagonal (empirical data) than the normal distribution tail data. This is yet another indicator of the long tail nature of the peak load demand events.

TABLE I

TRUE AND ESTIMATED PARAMETERS FOR SIMULATED DATA

	σ	λ	q
Estimated Value	0.0501	17.3584	0.0812
True Value	0.05	15	0.05

TABLE II

DISTRIBUTION PARAMETER ESTIMATES FOR THE MIXTURE MODEL

Distribution	Normal σ_{est}	Exp λ_{est}	Binomial q_{est}
	0.0584	16.9743	0.0208

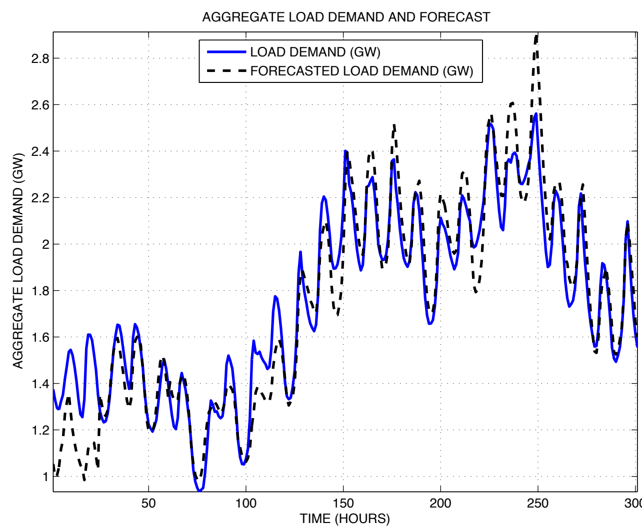

Fig. 4. Actual load demand data (GW) and regression model forecast.

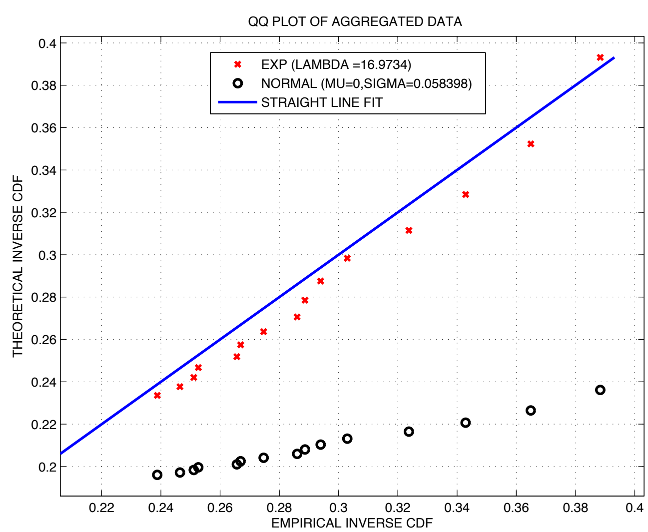

Fig. 5. QQ plot for the exponential and normal distribution tail data fit.

When simulating the forecast action of the model, it is assumed that an accurate forecast of the weather is available.

The day-ahead ambient temperature is needed to determine some of the regressors described in Section III-A. With the regression and long tail model estimates available, we can compute the risk adjusted forecast (32) for the long tail model. We also compute risk adjusted forecast (33) for the normal risk model. The margins are computed in accordance with risk value $r = 1.14 \cdot 10^{-4}$, see Section III-B.

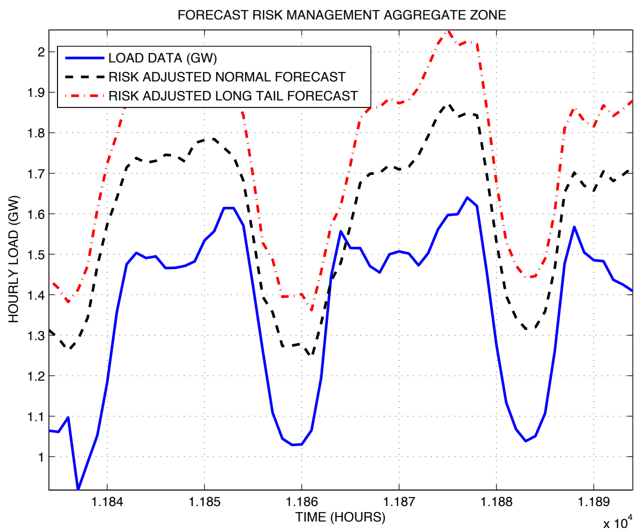

Fig. 6. Load data time with the two risk adjusted forecasts.

Figure 6 shows a 61 hour segment of the data along with the long tail risk adjusted forecast (larger margin) and the normal distribution risk adjusted forecast. As Figure 6 shows, the normal distribution risk adjusted forecast is surpassed by load demand at a few points. The long tail forecast is almost never exceeded by the demand as predicted by the model.

For better engineering insight, we compute the risk margin as an additive value in GW. This is how much extra power the utility must order to mitigate the risk. Since y_t in (20) is the logarithmic load, the additive margin (last term in (32) or in (33)) corresponds to a multiplier factor for the load. As a result, the power load margin GW values change with the load. Table III shows the additive margin values averaged over the entire 4-year dataset. The table shows over 50% difference in the margin values between the tail distribution and normal distribution.

Table IV is a description of the number of outliers that will occur per year depending on which risk adjustment margin and model is being used. The number of outliers predicted by the long tail model is always higher. If the margin is computed using the normal model, the long tail model shows that about 16 outliers per year would in fact occur instead of the desired single outlier per year. This is consistent with what is actually observed in the data.

TABLE III
MARGIN VALUES

	Forecast Margin
Normal Margin	0.3872 GW
Exp Margin	0.5785 GW

TABLE IV
OUTLIERS PER YEAR

	Normal Model	Long Tail Model
Normal Margin	1	15.7287
Exp Margin	$6.672 \cdot 10^{-4}$	1

VI. CONCLUSIONS

This paper shows that peak electrical power loads cannot be accurately forecasted with help of standard normal distribution models. We combined a robust statistics approach to modeling of the central part of the power load distribution with a long tail model of the extreme events, the data points that have large deviation from the regression model. Such modeling provides a very good description of the real utility power load data and allows to compute the forecast margin for a given level of exceedance risk. A simpler standard normal distribution (least squares) model underestimates the forecast margin value by more than 50% compared to the long tail model. For the utility data, an extra margin on the order of 200MW is necessary to avoid the risk unaccounted by the normal distribution.

REFERENCES

[1] P. Varaiya et al., "Smart operation of smart grid: Risk limiting dispatch," *Proc. IEEE*, vol. 99, no. 1, pp. 40–57, 2001.
[2] S. J. Beuning et al., *Variable Generation Power Forecasting for Operations*, NERC Report, 2010.
[3] R. V. Hogg and S. A. Klugman, "On the estimation of long tailed skewed distributions with actuarial applications," *Journal of Econometrics*, vol. 23, pp. 91–102, 1983.
[4] L. C. Thomas, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers," *International Journal of Forecasting*, vol. 16, pp. 149–172, 2000.
[5] G. Li, C. Liu, C. Mattson, and J. Lawarre, "Day-ahead electricity price forecasting in a grid environment," *IEEE Trans. Power Syst.*, vol. 22, pp. 266–274, February 2007.
[6] R. C. Garcia, J. Contreras, M. van Akkeren, and J. B. C. Garcia, "A garch forecasting model to predict day-ahead electricity prices," *IEEE Trans. Power Syst.*, vol. 20, pp. 867–874, May 2005.
[7] S. Subbayya, J.G. Jetcheva, and W.-P. Chen, "Model selection for short-term microgrid-scale electricity load forecasts," *IEEE PES Conf. on Innovative Smart Grid Technologies (ISGT)*, Washington, D.C., February 2013
[8] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and arima models," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1035–1042, 2005.
[9] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. New Jersey: Wiley, 2nd ed., 2009.
[10] P. J. Huber, "Robust regression: Asymptotics, conjectures and monte carlo," *Ann. Statist.*, vol. 1, no. 5, pp. 799–821, 1973.
[11] D. J. Cummins and C. W. Andrews, "Iteratively reweighted partial least squares: A performance analysis by Monte Carlo simulation," *Journal of Chemometrics*, vol. 9, pp. 489–507, 1995.
[12] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*. New York: Springer, 2006.
[13] C. Sigauke, A. Verster, and D. Chikobvu, "Extreme daily increase in peak electricity demand: Tail-quantile estimation," *Energy Policy*, vol. 53, pp. 90–96, 2013.
[14] S. Mirasgedis et al., "Models for mid-term electricity demand forecasting incorporating weather influences," *Energy*, vol. 31, pp. 208–227, 2006.
[15] Kaggle.com, "Global energy forecasting competition 2012 load forecasting," 2012, Available: `https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting`.