Statistical Representation of Distribution System Loads Using Gaussian Mixture Model

Ravindra Singh, *Student Member, IEEE*, Bikash C. Pal, *Senior Member, IEEE*, and Rabih A. Jabr, *Member, IEEE*

Abstract—This paper presents a probabilistic approach for statistical modeling of the loads in distribution networks. In a distribution network, the probability density functions (pdfs) of loads at different buses show a number of variations and cannot be represented by any specific distribution. The approach presented in this paper represents all the load pdfs through Gaussian mixture model (GMM). The expectation maximization (EM) algorithm is used to obtain the parameters of the mixture components. The performance of the method is demonstrated on a 95-bus generic distribution network model.

Index Terms—Expectation maximization (EM) algorithm, Gaussian mixture model, load profile.

I. INTRODUCTION

FOLLOWING the significant development in business regulations, technology evolutions, and various government policies towards low carbon generation technology, the main challenges before distribution companies are to improve their operating efficiencies, develop new tariffs, and offer new services to low voltage (LV) consumers without significant capital burden. Since the loads of the LV consumers are unmetered, the planning and operation of the networks requires to make use of sample customer load profiles and apply modeling techniques for greater asset utilization and increased automation. The randomness of the customer load behavior merely indicates the necessity of statistics-based modeling approach.

The most common technique to model the loads is through Gaussian distribution [1]. However, the single Gaussian assumption is not justified for all the loads [2]. In [2], the research work on statistical methods for load research data analysis concludes that the statistical distribution of electric load variation does not follow any common probability distribution function.

Several attempts have been made to model the loads through various probability distributions [3]–[7]. Irwin *et al.* [3] have fitted the Weibull distribution to consumer billing data. Although the distribution was flexible enough to explain

the distribution of energy consumption of the customers in Northern Ireland, the analysis covers only the customer energy and billing data. Seppala [4] has suggested log-normal distribution models. Reference [4] also proposed a model of customer load confidence interval. The models were verified from hourly load measurement data obtained from a Finnish load research project.

Herman and Kritzinger [5] fitted various distribution functions (Weibull, normal, Erlang, and beta) to grouped domestic loads. As a result, they proposed the beta distribution function. A similar research effort was reported by Ghosh *et al.* [6] in distribution system state estimation problem. They validated various models such as normal, log-normal, and beta distribution through chi-square goodness of fit test. Reference [6] concluded that an appropriate model was system specific with clear preference for the beta distribution because of its flexibility to adapt to the skewness in the distribution. An improved model with beta distribution was also presented in [7].

It can be concluded that there is no unique methodology to model the load pdf. In general, the Gaussian modeling of the load profiles appears to be a natural choice for various applications due to its simplicity as it can be completely described by two moments (mean and variance). Although there are a host of two parameter distributions, the Gaussian is preferable because the analysis based on this distribution is well developed and reported in the literature. Furthermore, many computational tools, such as the state estimation function, can easily incorporate the Gaussian pdfs.

In view of this, in our research, we have modeled the variability in the load distribution through Gaussian mixture model (GMM) approximation. The advantage of GMM approach is that different types of load distributions can be fairly represented as a convex combination of several normal distributions with respective means and variances. The problem of obtaining various mixture components (weight, mean, and variance) is formulated as a parametric estimation problem. The expectation maximization (EM) algorithm [8]–[10] was utilized to obtain the solution. The EM algorithm is a powerful tool in parameter estimation problems. It is a general method of finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data are incomplete or have missing values. The above formalism was applied to a 95-bus UK Generic Distribution System (UKGDS) model. Section II introduces the UKGDS network model and load profile calculations. GMM, EM algorithm, and further refinement through reduction techniques followed by simulation results are discussed in Sections III and IV.

R. Singh and B. C. Pal are with the Department of Electrical and Electronic Engineering, Imperial College, London SW7 2BT, U.K. (e-mail: ravindra.singh@ic.ac.uk; b.pal@ic.ac.uk).

R. A. Jabr is with the Department of Electrical and Computer Engineering, American University of Beirut, Beirut 1107 2020, Lebanon (e-mail: rabih.jabr@aub.edu.lb).

Fig. 1.   UKGDS: 95-bus test system model.

Fig. 2.   Annual half hourly demand profile indices for various consumers.

Fig. 3.   Probability distribution of load at different buses.
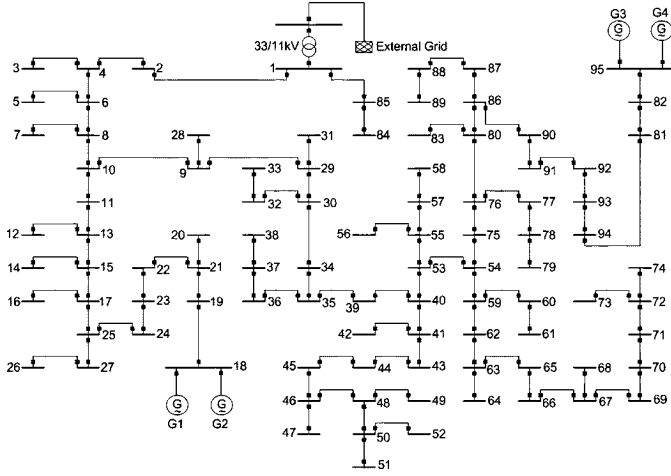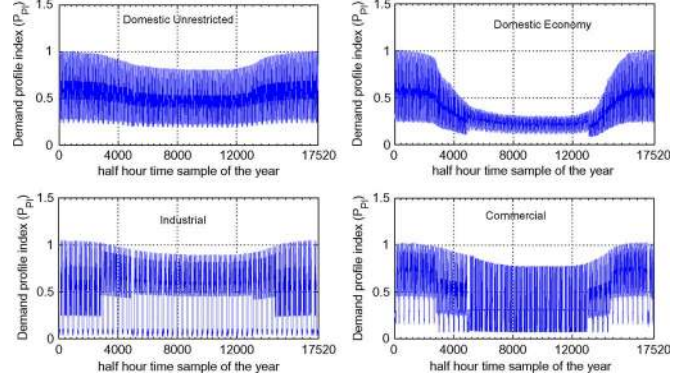
II. LOAD PROFILE CALCULATIONS

Load profiles were generated for a part of UKGDS model. The system model comprises 95 buses, with 55 load buses and two wind farms as sources of distributed generation (DG). Fig. 1 shows the schematic of the test system. The network and load data for UKGDS were obtained from [11].

The U.K. generic distribution network project [11] identified the following four types of consumers for developing generic load profile index (LPI).

1) Domestic-Unrestricted (D/U)
2) Domestic-Economy (D/E)
3) Industrial (I)
4) Commercial (C)

The LPI for a particular class of consumer was defined as the average half hourly power consumptions of several customers across the entire network measured at the same time. This was part of the load survey conducted in the UKGDS project [11]. The power was measured at a feeder of particular class of customer and was normalized with respect to that feeder rating. A half hourly normalized load profile was obtained. This half hourly profile across all the measured feeder of this particular category was averaged to generate a uniform LPI for a particular class of customer. LPIs of four types of consumers were computed in this way. The resulting LPIs are displayed in Fig. 2. The annual maximum demand (kW) information for a type of customer at a bus is known from the maximum demand indicator (MDI) record in each feeder. The annual maximum demand thus maps with the maximum value of the annual LPI.

Based on this information the real and reactive power load profiles at ith bus were computed as follows:

$$P_i(t) = \sum_{j=1}^{N_c} \frac{P_{i,max}^j}{\max_\tau \left(P_{PI}^j(\tau) \right)} P_{PI}^j(t) \qquad (1)$$

$$Q_i(t) = \sum_{j=1}^{N_c} \frac{P_{i,max}^j}{\max_\tau \left(P_{PI}^j(\tau) \right)} P_{PI}^j(t) \tan(\phi^j) \qquad (2)$$
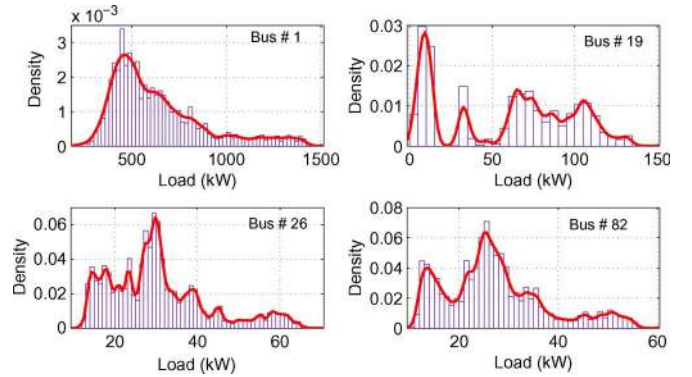
where

t, τ	half hourly time instances of the year;
$P_i(t), Q_i(t)$	real and reactive power loads at the ith bus at time instant t;
$P_{PI}^j(t)$	LPI value of the jth class of consumer at time instant t;
$P_{i,max}^j$	annual maximum demand of the jth class of consumer at bus i;
ϕ^j	angle of average power factor of the jth class of consumer;
N_c	number of consumer classes.

The typical power factors for all four classes of consumers were taken as 0.95, 0.99, 0.98, and 0.90 lagging, respectively.

A. Distribution of Load Profiles

The density histogram of load profiles was generated for a load bus by segmenting the range of the data into various disjoint categories known as bins. The computation of the relative frequencies of each bin is fairly standard and the details can be found in [12]. Some of the load distributions are displayed in Fig. 3. It is clear that no single standard distribution model can fit all of them. Fig. 4 shows that a nonstandard distribution can be composed through a weighted combination of normal distributions. This takes us to the concept of GMM, which is discussed next.
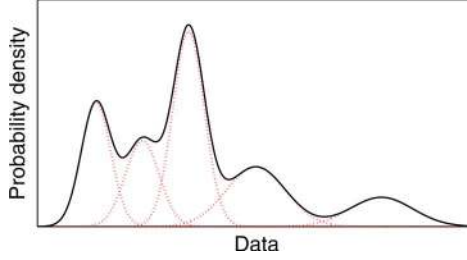
Fig. 4. Gaussian mixture approximation of density: dotted lines represent individual mixture components and solid line represents the resultant density.

III. GAUSSIAN MIXTURE MODEL

A Gaussian mixture (GM) pdf is a weighted finite sum of Gaussian pdfs (Fig. 4). It is characterized by the number of mixture components and the weights, mean, and variance (mean vector and covariance matrix for the multivariate case) of each component. Since a pdf must be nonnegative and the integral of a pdf over the sample space of the random quantity it represents must evaluate to unity, the mixture weights must be nonnegative and the sum of all the weights must equal to one. For the multivariate case, the GM pdf model is given by

$$f(\mathbf{z}|\boldsymbol{\gamma}) = \sum_{i=1}^{M_c} w_i f(\mathbf{z}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad (3)$$

where M_c is the number of mixture components and w_i is the weight of the ith mixture component, subject to $w_i > 0$ and $\sum_{i=1}^{M_c} w_i = 1$. $\boldsymbol{\gamma}$ is chosen from the set of parameters $\boldsymbol{\Gamma} = \{\boldsymbol{\gamma} : \boldsymbol{\gamma} = \{w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^{M_c}\}$, a member of which defines a GM. Given a d-dimensional random variable $\mathbf{z}$, with mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$, the density function of each mixture component $f(\mathbf{z}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is a normal distribution given by

$$f(\mathbf{z}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)(\mathbf{z})$$
$$= \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma}_i)^{1/2}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{z}-\boldsymbol{\mu}_i)}. \quad (4)$$

Example: Mixture of Two Gaussian Components: Consider that a probability density function is represented by a weighted sum of two normally distributed pdfs. The resulting probability density function will be

$$f\left(z|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = w_1 \mathcal{N}\left(\mu_1, \sigma_1^2\right) + w_2 \mathcal{N}\left(\mu_2, \sigma_2^2\right) \quad (5)$$
$$= \frac{w_1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(z-\mu_1)^2}{2\sigma_1^2}}$$
$$+ \frac{w_2}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(z-\mu_2)^2}{2\sigma_2^2}}. \quad (6)$$

Integrating both sides of (6) over the sample space of $z \in (-\infty + \infty)$

$$\int_{-\infty}^{\infty} f\left(z|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) dz = w_1 \int_{-\infty}^{\infty} \frac{e^{-\frac{(z-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi\sigma_1^2}} dz$$
$$+ w_2 \int_{-\infty}^{\infty} \frac{e^{-\frac{(z-\mu_2)^2}{2\sigma_2^2}}}{\sqrt{2\pi\sigma_2^2}} dz. \quad (7)$$

Since integral of any pdf over the entire sample space gives the total probability of distribution which is equal to one, the integral terms on both sides of (7) evaluate to unity. In this case, (7) becomes

$$w_1 + w_2 = 1. \quad (8)$$

Equation (8) shows that the resultant density in (6) is the convex combination of the two Gaussian densities. In this example, in order to express a given probability density as a convex combination of two Gaussian components, we need to know the parameters $\{w_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\}$. The EM algorithm is an efficient tool to obtain these parameters. The details of the EM algorithm to obtain the mixture parameters to fit a distribution are described next.

A. GM Parameter Estimation by EM Algorithm

The EM algorithm to obtain the parameters of the GM, constructed by sampling from a given distribution, has been adopted in various application areas such as target tracking [13], clustering [14], and pattern recognition [15]. In the context of power distribution load modeling, a version of the EM algorithm is used to obtain $\boldsymbol{\gamma}$ in the circumstances when each $f(\mathbf{z}|\boldsymbol{\gamma})$ can be regarded as marginal distribution of a joint distribution $f(\mathbf{z}, \mathbf{y}|\boldsymbol{\gamma})$ involving an additional "auxiliary" variable $\mathbf{y}$, i.e.

$$f(\mathbf{z}|\boldsymbol{\gamma}) = \int f(\mathbf{z}, \mathbf{y}|\boldsymbol{\gamma}) d\mathbf{y}. \quad (9)$$

Let $\mathcal{Z}$ denote a set of N independent and identically distributed data samples, i.e., $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$ and $\mathcal{Y} = \{y_1, y_2, \ldots, y_N\}$ denote a set of hidden random variables y_i, such that $y_i \in \boldsymbol{\Upsilon} = \{y : y = \{1, 2, \ldots, M_c\}\}$. To put it simply, each y_i refers to a mixture component through which an observation $\mathbf{z}_i$ comes from.

The EM algorithm used in this paper maximizes the following log-likelihood expectation:

$$Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s) = E\left[\ln\left(f(\mathcal{Z}, \mathcal{Y}|\boldsymbol{\gamma})|\mathcal{Z}, \boldsymbol{\gamma}^s\right)\right] \quad (10)$$

where $\boldsymbol{\gamma}^s$ are the current parameters estimates that are used to evaluate the expectation and $\boldsymbol{\gamma}$ are the new parameters that are optimized to increase Q.

The EM algorithm recursively generates a sequence of parameters $\boldsymbol{\gamma}^s = \{w_j^s, \boldsymbol{\mu}_j^s, \boldsymbol{\Sigma}_j^s\}_{j=1}^{M_c}$, $s = 1, 2, \ldots$ that, in feasible circumstances, converges to a maximizer of $Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s)$ over $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$. One step of the recursion, yielding formulae for $\boldsymbol{\gamma}^{s+1}$ given $\boldsymbol{\gamma}^s$, is

$$w_j^{s+1} = \frac{1}{N} \sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s) \quad (11)$$

$$\boldsymbol{\mu}_j^{s+1} = \frac{\sum_{i=1}^{N} \mathbf{z}_i f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)}{\sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)} \quad (12)$$

$$\boldsymbol{\Sigma}_j^{s+1} = \frac{\sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)\left(\mathbf{z}_i - \boldsymbol{\mu}_j^{s+1}\right)\left(\mathbf{z}_i - \boldsymbol{\mu}_j^{s+1}\right)^T}{\sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)} \quad (13)$$

where

$$
\begin{aligned}
f(j|\mathbf{z}_i,\boldsymbol{\gamma}^s) &= \frac{w_j^s\, f\left(\mathbf{z}_i|\boldsymbol{\mu}_j^s,\boldsymbol{\Sigma}_j^s\right)}{\sum_{k=1}^{M_c} w_k^s\, f\left(\mathbf{z}_i|\boldsymbol{\mu}_k^s,\boldsymbol{\Sigma}_k^s\right)} \\
&= \frac{w_j^s\,\mathcal{N}\left(\boldsymbol{\mu}_j^s,\boldsymbol{\Sigma}_j^s\right)(\mathbf{z})}{\sum_{k=1}^{M_c} w_k^s\,\mathcal{N}\left(\boldsymbol{\mu}_k^s,\boldsymbol{\Sigma}_k^s\right)(\mathbf{z})}.
\end{aligned} \qquad (14)
$$

The steps to obtain (11)–(13) are briefly explained in Appendix A.

IV. SIMULATION STUDY

The formulation of GMM is general, as it treats $\mathbf{z}$ as a matrix of dimension $d \times N$. We first consider the univariate case in which we have applied the GMM methodology to the UKGDS model to obtain the mixture components for real and reactive demands in all 55 buses. In the univariate distribution $(d=1)$, the load at each bus was modeled as a GMM. This excludes the correlation amongst the loads. We also applied the GMM in the multivariate case by augmenting the loads at different buses in $\mathbf{z}$. The multivariate GMM incorporates the correlation information. The GMM components at every bus were obtained using the EM algorithm. The algorithm was coded in MATLAB and run on a Pentium-IV PC, 2.99-GHz processor with 1 GB of RAM. The EM algorithm was initialized using K-*means* clustering algorithm available in MATLAB [16]. The algorithm was terminated when the relative difference of log-likelihood values [17] in two consecutive iterations is below a threshold. (i.e., $|(L_n - L_{n-1})/L_{n-1}| \le thr$). A threshold value of 0.001 was taken as the termination criterion. In the algorithm, the number of mixture components (M_c) was pre-specified. The basis for choosing the M_c is explained with the help of Table III in Section IV-D.

A. Full Component GMM

Fig. 5 shows the distribution of various mixture components in buses #1, #26, and #82. There can be a situation in which a load can be represented by more than one Gaussian mixture component. For example, a careful look at bus #82 load pdf suggests that a 12-kW value can be solely captured by mixture component #2, whereas a 25-kW value can be represented either by component #1 or component #5 or by the weighted combination of the both. Since a single Gaussian pdf is required in various applications, it is necessary to produce an equivalent Gaussian pdf for the 25-kW load. There are several ways to achieve this. One can select the pdf with the highest weight. However when the weights are comparable with low overlap between the respective pdfs, both of them should be chosen. In the cases such as 25 kW, where the respective pdfs have significant overlap with comparable weights, the equivalent mean and variance can be generated by merging relevant components. A full component merging to produce the overall mean and covariance is based on the following equations:

$$
\boldsymbol{\mu}_c = E[\mathbf{z}] = \sum_{j=1}^{M_c} w_j \boldsymbol{\mu}_j \qquad (15)
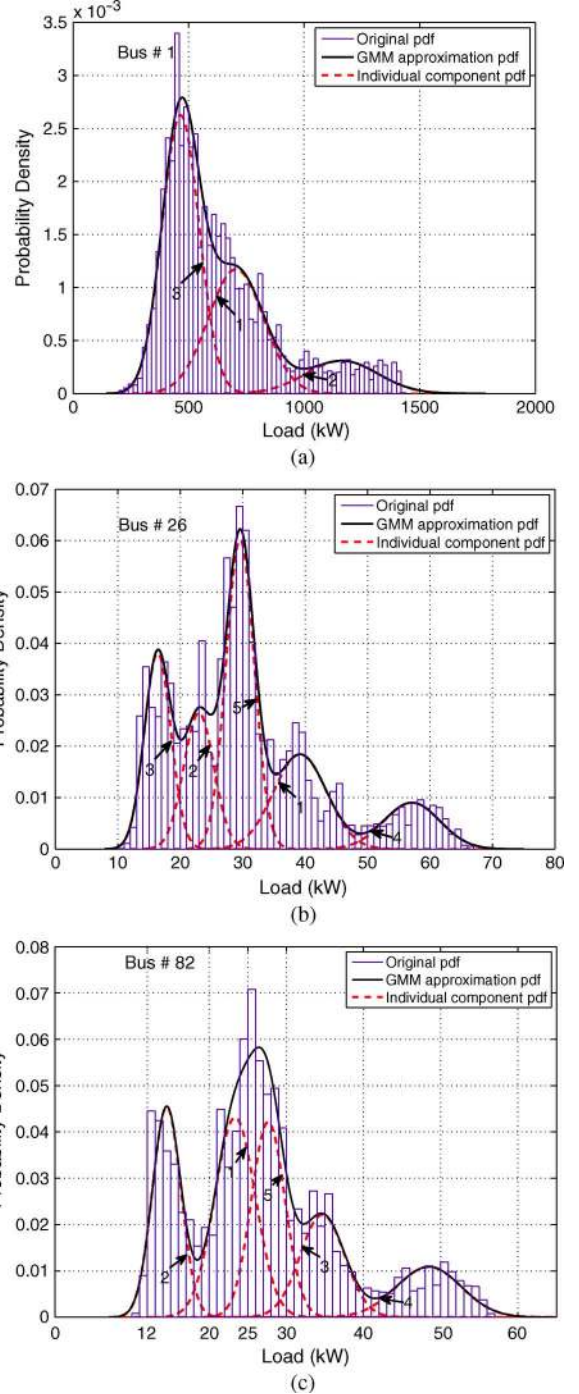$$

Fig. 5. GMM approximation of the load pdf with (a) 3, (b) 5, and (c) 5 Gaussian components.

$$
\begin{aligned}
\boldsymbol{\Sigma}_c &= E\left[(\mathbf{z}-\boldsymbol{\mu}_c)(\mathbf{z}-\boldsymbol{\mu}_c)^T\right] \\
&= \sum_{j=1}^{M_c} w_j\left[\boldsymbol{\Sigma}_j + (\boldsymbol{\mu}_j-\boldsymbol{\mu}_c)(\boldsymbol{\mu}_j-\boldsymbol{\mu}_c)^T\right].
\end{aligned} \qquad (16)
$$

However, this generates a pdf with larger spread and does not seem to offer preferential treatment to high density region. This happens due to the presence of some pdfs having significantly different parameters from the rest. A solution is through an improved analytical approach to merging via *mixture reduction*.

TABLE I
PARAMETERS OF THE GMM COMPONENTS: THE HIGHLIGHTED COMPONENTS
ARE SELECTED FOR DELETION USING DISTANCE MEASURE GIVEN IN (17)

Bus No.	GMM Components		
	Weight	Mean	Variance
1	(0.35 **0.12** 0.53)	(705.39 **1163.28** 465.28)	(14326.16 **24200.85** 6420.11)
3	(0.20 0.71 **0.10**)	(7.90 3.60 **12.68**)	(1.60 1.01 **1.82**)
5	(0.13 **0.08** 0.51 0.28)	(21.17 **30.54** 9.92 16.17)	8.12 **5.50** 2.76 4.99)
11	(0.20 0.71 **0.09**)	(2.63 1.20 **4.23**)	(0.18 0.11 **0.20**)
12	(**0.09** 0.20 0.28 0.43)	(**4.24** 2.63 0.90 1.39)	(**0.20** 0.18 0.04 0.05)
14	(0.31 **0.29** 0.40)	(13.03 **19.37** 2.62)	(2.89 **5.45** 3.80)
22	(0.43 0.20 **0.09** 0.28)	(8.33 15.77 **25.42** 5.38)	(1.92 6.41 **7.06** 1.63)
23	(**0.09** 0.51 0.20 0.21)	(**25.26** 7.32 16.16 10.44)	(**3.61** 1.84 5.72 1.80)
24	(0.26 **0.08** 0.50 0.16)	(28.52 **56.62** 17.91 38.31)	(16.90 **19.78** 8.24 28.36)
26	(0.20 0.16 0.20 **0.10** 0.34)	(39.19 22.97 16.38 **57.11** 29.60)	(18.03 5.87 4.50 **20.08** 5.04)
28	(0.43 0.43 **0.14**)	(7.36 13.13 **23.04**)	(2.44 5.38 **9.80**)
33	(**0.09** 0.71 0.20)	(**21.14** 6.00 13.17)	(**5.07** 2.81 4.45)
34	(0.21 0.20 **0.09** 0.50)	(10.44 16.16 **25.26** 7.32)	(1.80 5.72 **3.61** 1.84)
36	(0.27 0.13 0.51 **0.09**)	(30.79 39.36 18.78 **56.64**)	(16.54 28.18 10.60 **19.35**)
42	(0.56 0.33 **0.11**)	(9.59 16.68 **28.72**)	(3.41 11.10 **13.75**)
43	(0.44 **0.13** 0.43)	(14.71 **46.09** 26.26)	(9.77 **39.22** 21.51)
47	(**0.09** 0.71 0.20)	(**12.68** 3.60 7.90)	(**1.82** 1.01 1.60)
51	(0.33 0.56 **0.11**)	(16.67 9.59 **28.72**)	(11.10 3.41 **13.75**)
54	(0.20 0.21 0.51 **0.08**)	(16.16 10.44 7.32 **25.26**)	(5.72 1.80 1.84 **3.61**)
56	(0.71 0.20 **0.09**)	(1.92 4.21 **6.76**)	(0.29 0.45 **0.52**)
57	(0.20 **0.09** 0.71)	(4.74 **7.61** 2.16)	(0.58 **0.66** 0.36)
60	(**0.09** 0.71 0.20)	(**12.68** 3.60 7.90)	(**1.82** 1.01 1.60)
61	(0.71 0.20 **0.09**)	(3.60 7.90 **12.68**)	(1.01 1.60 **1.82**)
64	(0.48 **0.08** 0.19 0.25)	(8.00 **26.00** 17.05 12.23)	(1.55 **4.70** 6.08 2.93)
65	(0.20 **0.09** 0.71)	(15.81 **25.37** 7.20)	(6.40 **7.30** 4.05)
68	(0.20 **0.09** 0.71)	(18.44 **29.60** 8.40)	(8.72 **9.93** 5.51)
69	(**0.09** 0.71 0.20)	(**11.84** 3.36 7.38)	(**1.59** 0.88 1.40)
73	(**0.09** 0.71 0.20)	(**10.99** 3.12 6.85)	(**1.37** 0.76 1.20)
74	(0.12 0.14 **0.31** 0.43)	(52.00 23.70 **14.82** 34.78)	(19.78 8.83 **9.23** 10.74)
75	(0.71 **0.09** 0.20)	(2.40 **8.46** 5.27)	(0.45 **0.81** 0.71)
77	(**0.09** 0.20 0.71)	(**4.23** 2.63 1.20)	(**0.20** 0.18 0.11)
78	(0.20 **0.09** 0.71)	(6.32 **10.15** 2.88)	(1.02 **1.17** 0.65)
79	(0.48 0.25 **0.19 0.08**)	(15.99 24.46 **34.10 52.01**)	(6.20 11.72 **24.30 18.80**)
82	(0.29 0.21 0.16 **0.11** 0.23)	(23.35 14.49 34.67 **48.47** 27.65)	(6.96 3.45 8.29 **17.60** 4.71)
83	(**0.09** 0.71 0.20)	(**4.23** 1.20 2.63)	(**0.20** 0.11 0.18)
84	(0.45 0.27 0.18 **0.10**)	(389.58 272.70 543.83 **796.50**)	(1626.46 1449.86 3884.44 **6508.93**)

B. Reduced Component GMM

The merging is based on the clustering algorithm that combines mixtures into groups (clusters). The algorithm operates by selecting the component with the largest weight as the principal component for a cluster, and merging all components that are within a certain distance of the principal component. The distance measure is defined by Salmond [18]:

$$D_{p,j} = \frac{w_p w_j}{w_p + w_j}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_j) \qquad (17)$$

where subscript "p" denotes the principal component. All the mixture components, satisfying $D_{p,j} < T$ for $j = 1, 2, \ldots, M_c$, are merged together. The components which do not satisfy $D_{p,j} < T$ are ignored. Threshold "T" is determined by the χ^2-test using a 99% confidence. The equivalent mean and covariance of the merged components are given by

$$w_m = \sum_{j \in \mathcal{I}} w_j \qquad (18)$$

$$\boldsymbol{\mu}_m = \frac{1}{w_m} \sum_{j \in \mathcal{I}} w_j \boldsymbol{\mu}_j \qquad (19)$$

$$\boldsymbol{\Sigma}_m = \frac{1}{w_m} \sum_{j \in \mathcal{I}} w_j \left[\boldsymbol{\Sigma}_j + (\boldsymbol{\mu}_j - \boldsymbol{\mu}_m)(\boldsymbol{\mu}_j - \boldsymbol{\mu}_m)^T \right]. \qquad (20)$$

It is to be noted that because of the exclusion of some of the components based on the above criterion, the summation of the weights of the components within the cluster will not add to unity. This has been accounted for by normalizing the mean and variances of the cluster with w_m.

Table I shows the parameters of GMM components. The buses requiring mixture reduction are shown in Table I. The parameters of the components selected for deletion are highlighted.

The results of the full component merging and reduced component merging to a single Gaussian pdf are displayed in Fig. 6.
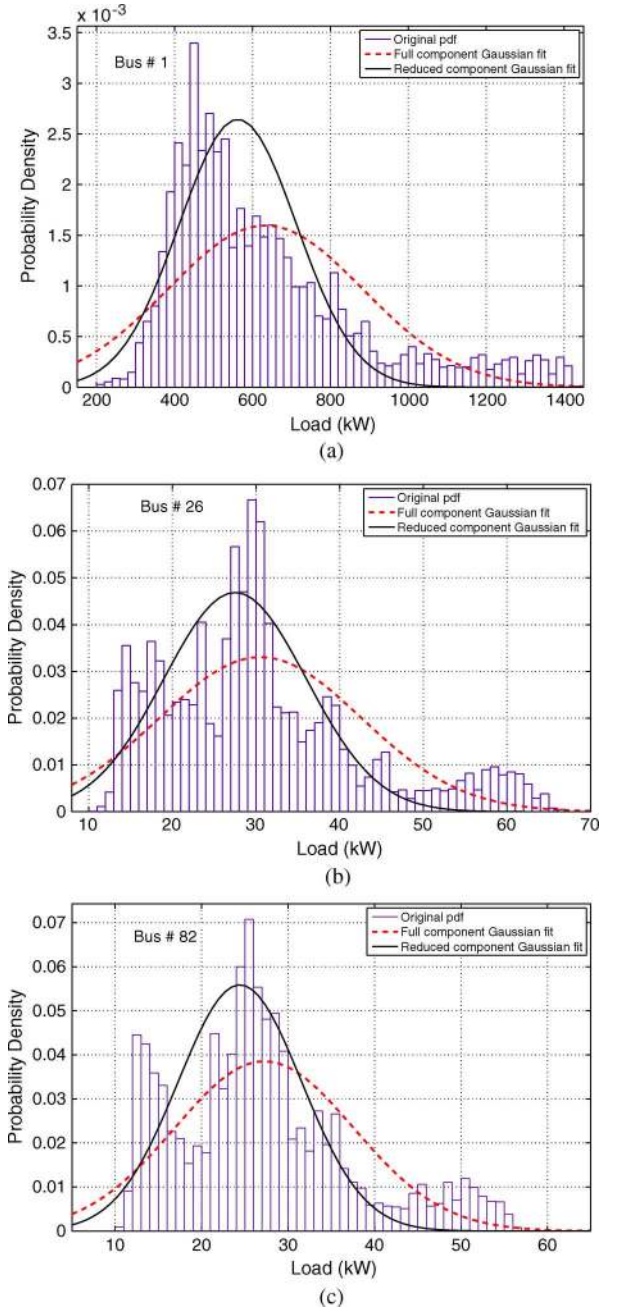

Fig. 6. Single Gaussian approximation of the load pdf with and without mixture reduction.

It is seen that the mixture reduction algorithm provides a better approximation of the original pdf. In Fig. 6, the merging of full and reduced components to a single Gaussian pdf is demonstrated for comparison purposes. In practice, a set of non-overlapping Gaussian pdfs is produced by merging the relevant components and each pdf in the set statistically represents the range of load (low, medium, and high).

C. Performance Comparison of GMM With Other Distributions

To compare the performance of the GMM, a number of distributions (normal, log-normal, beta, gamma) were fitted in the load density histograms. The fitted distributions along with
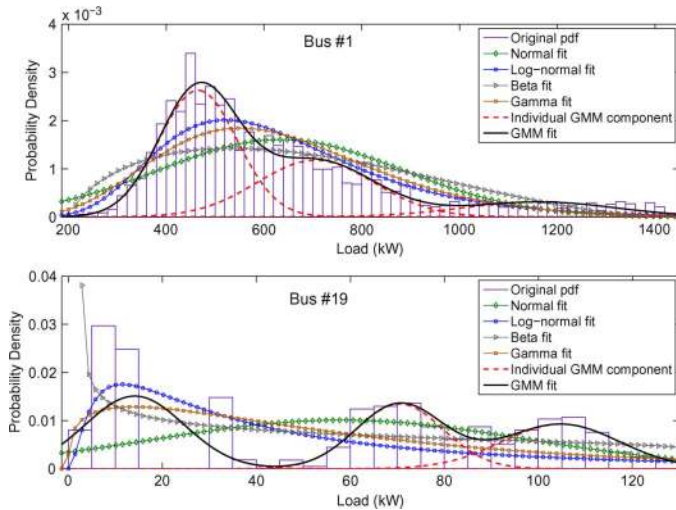
Fig. 7. Probability distribution of load at different buses.

M_c	Chi-Square Goodness-of-Fit	CPU time (s)	Iterations
2	952.65	12.88	4
3	163.11	33.81	8
4	61.13	94.75	18
5	47.17	200.61	26

Fig. 8. Bivariate GMM approximation with four mixture components.

TABLE II
CHI-SQUARE GOODNESS-OF-FIT FOR VARIOUS DISTRIBUTIONS

Bus No.	Chi-Square Goodness-of-Fit				
	Normal	Log-normal	Beta	Gamma	GMM
1	8649.02	1585.31	6610.62	3046.02	163.11
19	8204.74	18275.99	6653.73	12673.63	5093.15

GMM at buses #1 and #19 are displayed in Fig. 7. The GMM offers to provide a better fit for the load density histograms as compared to other distributions. To verify this numerically, the χ^2-statistics which measures the goodness of fit was used. A similar measure is used in [6]. The Chi-square goodness-of-fit was obtained using $chi2gof$ function available in MATLAB [16]. Table II shows the Chi-square goodness-of-fit for various distributions. A smaller value indicates a better fit. At both buses, this value is smallest for the GMM, confirming the GMM as best fit of all those considered.

D. Performance of GMM With Increased Mixture Components

Table III shows the chi-square goodness-of-fit values for the GMM at bus #1 after termination of the algorithm with #2, #3, #4, and #5 Gaussian components. A lower value of the chi-square goodness-of-fit indicates better accuracy. For one Gaussian pdf, the value of the chi-square goodness-of-fit is very high (8649.02 in Table II). As seen from Table III, when the number of components increases from 2 to 3 and 3 to 4, the chi-square goodness-of-fit value decreases significantly. With increase in number of components from 4 to 5, this decrease is not very high, whereas computational efforts in terms of CPU times and number of iterations are very high. Thus as a trade-off, three components were selected for the GMM at bus #1. The components for the GMM at other buses were determined similarly.

E. Load Correlation Through Multivariate GMM

The results presented in the previous sections were obtained by considering the load profile at an individual bus. This leads to the univariate GMM, and thus, no information about the correlations was obtained. However, the GMM algorithm presented in
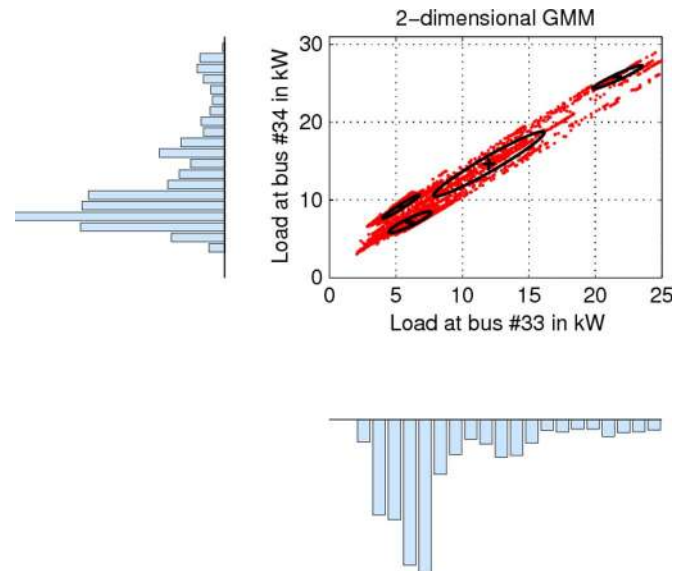
this paper is generic and also incorporates the multivariate distributions. In the multivariate case, the load profiles at different buses are augmented together in $\mathbf{z}$ in order to capture the correlations. In this case, the off-diagonal elements of the covariance matrix $\mathbf{\Sigma}_i$ (corresponding to the ith component) represent the correlations. In this research, we augmented the real and reactive power load profiles at all 55 buses and obtained the GMM components. The GMM was represented by ten mixture components. The covariance matrix of each component has dimension of 110×110 which is difficult to show in this paper. However, we have demonstrated the correlation between the real power loads by considering the correlations in two and three loads in buses which are close to each other. Fig. 8 shows the two-dimensional GMM which is obtained from the load profiles at buses #33 and #34, respectively. The bivariate distribution was modeled through four Gaussian components which in two dimensions are represented by black ellipses. Table IV shows the parameters of GMM components. In the table, off-diagonal elements of each of the $\{\mathbf{\Sigma}_i\}_{i=1}^4$ represent the correlation between the loads at buses #33 and #34. Similarly the load correlations between the buses #22, #23, and #24 were obtained, and results are presented in Table V. In the three-dimensional case, a Gaussian component is represented by an ellipsoid.

V. DISCUSSION

The GMM technique and statistical representation of the load based on the consumer load profiles can be very useful for various distribution system applications such as distribution network planning [1], probabilistic load flow [19], load forecasting

TABLE IV
MIXTURE COMPONENTS WITH CORRELATION AT BUSES #33 AND #34

CPU time: 168.72s, Number of iterations: 35, Likelihood value: 4.0225×10^5

$$w_1 = 0.3832, \mu_1 = \begin{bmatrix} 6.0586 \\ 7.1458 \end{bmatrix}, w_2 = 0.3233, \mu_2 = \begin{bmatrix} 12.0030 \\ 14.6331 \end{bmatrix}, w_3 = 0.0670, \mu_3 = \begin{bmatrix} 21.6750 \\ 25.8169 \end{bmatrix}, w_4 = 0.2265, \mu_4 = \begin{bmatrix} 5.3738 \\ 9.1446 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 2.5526 & 2.0222 \\ 2.0222 & 1.9093 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 17.3008 & 16.7542 \\ 16.7542 & 17.4402 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 3.4871 & 2.7186 \\ 2.7186 & 2.3094 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 1.9364 & 2.0616 \\ 2.0616 & 2.3052 \end{bmatrix}$$

TABLE V
MIXTURE COMPONENTS WITH CORRELATION AT BUSES #22, #23, AND #24

CPU time: 17.58 s, Number of iterations: 3, Likelihood value: 3.5617×10^5

$$w_1 = 0.2188, \mu_1 = \begin{bmatrix} 8.5481 \\ 9.0205 \\ 27.4467 \end{bmatrix}, w_2 = 0.4871, \mu_2 = \begin{bmatrix} 6.5988 \\ 7.7851 \\ 17.8099 \end{bmatrix}, w_3 = 0.2075, \mu_3 = \begin{bmatrix} 16.0242 \\ 16.0006 \\ 36.4170 \end{bmatrix}, w_4 = 0.0866, \mu_4 = \begin{bmatrix} 25.8241 \\ 25.2209 \\ 55.6186 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 2.3407 & 2.0645 & 2.3764 \\ 2.0645 & 3.3919 & 1.9322 \\ 2.3764 & 1.9322 & 11.7299 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 3.6267 & 2.2787 & 2.2465 \\ 2.2787 & 3.2504 & 2.3928 \\ 2.2465 & 2.3928 & 7.7021 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 7.2342 & 5.8072 & 10.7451 \\ 5.8072 & 5.9938 & 11.3317 \\ 10.7451 & 11.3317 & 31.4828 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 5.6938 & 3.8896 & 12.2212 \\ 3.8896 & 3.8227 & 7.6954 \\ 12.2212 & 7.6954 & 30.2096 \end{bmatrix}$$

[20], customer billing [3], load management [21], state estimation [6], system restoration, and distribution automation. For example, most of the probabilistic load flow techniques are developed based on the assumption that the distribution of the load is Gaussian. This assumption works well in transmission networks, whereas in distribution networks, different probabilistic load distributions exist and representing them through single Gaussian pdf cannot be justified. Furthermore, the large size of the distribution network having various probability distributions at different buses makes accommodating them in a single load flow formulation impractical. On the other hand, representation of loads through GMM provides a unique framework to model the variability of distribution functions while retaining the Gaussian distribution assumption.

The GMM-based technique finds another important application in distribution system state estimation (DSSE). In distribution systems, measurements are limited, and hence, a large number of pseudo measurements are introduced in order to run the state estimation algorithms. The most common algorithm for the DSSE is weighted least squares (WLS) [22]. The WLS formulation is based on the maximum-likelihood estimation theory and utilizes the fact that the measurements are normally distributed. The GMM-based technique can be applied to model the loads as pseudo measurements and the Gaussian components of the pseudo measurements can be accommodated in WLS formulation.

In the load forecasting based on the ARMA model [20], the error component of the model can be represented through GMM. In reliability studies, various probability of failure functions irrespective of their distributions can be fairly represented by GMM. In the billing process, the GMM can be utilized to explain the energy consumptions of various class of customers. In load management, where the load modeling methodology allows the independent consideration of individual load component use and response model, the proposed technology can be adapted to the study of the load response evaluation for demand side management control actions, cold load pick-up, etc.

In general, the proposed technique can be applied in hosts of probabilistic-based power system analysis. However, our recent research efforts have targeted its application in distribution system state estimation [23]. In DSSE, the loads are modeled as pseudo measurements and load information is obtained from the load profile. If we choose a snapshot of load to perform the

state estimation, we also need the information about its variance. This information is obtained from the mixture component corresponding to this load. A load can be associated with a particular Gaussian component in the mixture through data association by relative marginal density given in (14). This is done by computing the relative marginal density of the load snapshot with respect to each component. The component with maximum density is identified as the representative of the load. If several components have comparable marginal densities, they can be merged together using the mixture reduction technique described in Section IV-B. As seen from (14), the computation of marginal densities with respect to each component requires all the parameters of the GMM.

The performance of the DSSE using GMM is reported in [23] where two approaches are demonstrated. One approach is based on measurement correlation factor and the other is based on GMM. The results in the paper confirm that the GMM performance is better than that of the correlation-based approach.

The effect of correlation between the loads on the accuracy of the estimates is significant [24]. In view of this, the multivariate GMM must be used to generate the parameters of the Gaussian components, which are useful for state estimation.

VI. CONCLUSION

An efficient approach based on the EM algorithm, to represent the loads statistically, is presented and theoretically justified. The advantage of the approach is that all the load pdfs irrespective of their distributions are represented by GMM approximation followed by the appropriate reduction. The performance comparison of the GMM with various distribution indicates its effectiveness for load modeling in various distribution system applications.

APPENDIX A
EM ALGORITHM AND PROBABILITY DENSITY ESTIMATION

Given a measure space $\mathcal{Z}$ of *incomplete data*. Assume that a complete data set exists $\mathcal{D} = (\mathcal{Z}, \mathcal{Y})$. An element of $\mathcal{Y}$ is a vector which contains the labels of each element of $\mathcal{Z}$ (i.e., which mixture component spawned which samples). In view of Section III-A assume a joint density function

$$f(\mathbf{z}, \mathbf{y}|\boldsymbol{\gamma}) = f(\mathbf{y}|\mathbf{z}, \boldsymbol{\gamma}) f(\mathbf{z}|\boldsymbol{\gamma}). \tag{21}$$

Now, define a complete data likelihood as a function of the joint density:

$$\mathcal{L}(\boldsymbol{\gamma}|\mathcal{Z}, \mathcal{Y}) = f(\mathcal{Z}, \mathcal{Y}|\boldsymbol{\gamma}). \tag{22}$$

The EM algorithm first finds the expected value of complete-data log-likelihood $\ln f(\mathcal{Z}, \mathcal{Y}|\boldsymbol{\gamma})$ with respect to the unknown data $\mathcal{Y}$ given observation $\mathcal{Z}$ and current parameter estimates $\boldsymbol{\gamma}^s$. That is, we define

$$Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s) = E\left[\ln(f(\mathcal{Z}, \mathcal{Y}|\boldsymbol{\gamma})|\mathcal{Z}, \boldsymbol{\gamma}^s)\right]$$
$$= \int_{\mathbf{y} \in \boldsymbol{\Upsilon}} \ln\left(f(\mathcal{Z}, \mathbf{y}|\boldsymbol{\gamma})\right) f(\mathbf{y}|\mathcal{Z}, \boldsymbol{\gamma}^s) d\mathbf{y}. \tag{23}$$

$f(\mathbf{y}|\mathcal{Z}, \boldsymbol{\gamma}^s)$ is the marginal distribution of the unobserved data and is dependent on both the observed data $\mathcal{Z}$ and current parameters.

The evaluation of the expectation is called the E-step of the EM algorithm. The second step of the algorithm is called the M-step, in which the expectation computed in the E-step is maximized by choosing

$$\boldsymbol{\gamma}^{s+1} \in \arg\max_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s). \tag{24}$$

$\arg\max_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s)$ denotes the set of values $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ which maximize $Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s)$ over $\boldsymbol{\Gamma}$. This set must be nonempty for the M-step to be well defined. If this set is a singleton, then we have

$$\boldsymbol{\gamma}^{s+1} = \arg\max_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s). \tag{25}$$

These two steps are repeated in recursion until convergence. Now consider a Gaussian mixture model defined in (3). Also define the vector $\mathbf{y} = (y_1, y_2, \ldots, y_N)$. If we know $\mathbf{y}$, we have

$$\ln\left(f(\mathcal{Z}, \mathbf{y}|\boldsymbol{\gamma})\right) = \sum_{i=1}^{N} \ln\left(w_{yi} f(\mathbf{z}_i|\boldsymbol{\mu}_{yi}, \boldsymbol{\Sigma}_{yi})\right) \tag{26}$$

$$f(\mathbf{y}|\mathcal{Z}, \boldsymbol{\gamma}^s) = \prod_{i=1}^{N} f(y_i|\mathbf{z}_i, \boldsymbol{\gamma}^s). \tag{27}$$

By Bayes's rule

$$f(y_i|\mathbf{z}_i, \boldsymbol{\gamma}^s) = \frac{f(\mathbf{z}_i, y_i|\boldsymbol{\gamma}^s)}{f(\mathbf{z}_i|\boldsymbol{\gamma}^s)} = \frac{w_{yi}^s f\left(\mathbf{z}_i|\boldsymbol{\mu}_{yi}^s, \boldsymbol{\Sigma}_{yi}^s\right)}{\sum_{j=1}^{M_c} w_j^s f\left(\mathbf{z}_i|\boldsymbol{\mu}_j^s, \boldsymbol{\Sigma}_j^s\right)}. \tag{28}$$

Equation (23) in discrete domain takes the form

$$Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s) = \sum_{\mathbf{y} \in \boldsymbol{\Upsilon}} \ln\left(f(\mathcal{Z}, \mathbf{y}|\boldsymbol{\gamma})\right) f(\mathbf{y}|\mathcal{Z}, \boldsymbol{\gamma}^s)$$
$$= \sum_{\mathbf{y} \in \boldsymbol{\Upsilon}} \sum_{i=1}^{N} \ln\left(w_{yi} f(\mathbf{z}_i|\boldsymbol{\mu}_{yi}, \boldsymbol{\Sigma}_{yi})\right) \prod_{i=1}^{N} f(y_i|\mathbf{z}_i, \boldsymbol{\gamma}^s)$$
$$= \sum_{y_1} \sum_{y_2} \cdots \sum_{y_N} \sum_{i=1}^{N} \ln\left(w_{yi} f(\mathbf{z}_i|\boldsymbol{\mu}_{yi}, \boldsymbol{\Sigma}_{yi})\right)$$
$$\times \prod_{i=1}^{N} f(y_i|\mathbf{z}_i, \boldsymbol{\gamma}^s). \tag{29}$$

After some complex algebraic manipulation, the above equation takes the following simplified form:

$$Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^s) = \sum_{j=1}^{M_c} \sum_{i=1}^{N} \ln\left(w_j f(\mathbf{z}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\right) f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)$$
$$= \sum_{j=1}^{M_c} \sum_{i=1}^{N} \ln(w_j) f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)$$
$$+ \sum_{j=1}^{M_c} \sum_{i=1}^{N} \ln\left(f(\mathbf{z}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\right) f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s). \tag{30}$$

The parameters of the mixture components can be obtained by maximizing the above expression using first-order (Kuhn Tucker) conditions. Now we can introduce the Lagrange multiplier λ with the constraint that $\sum_{j=1}^{M_c} w_j = 1$ and find w_j as follows:

$$\frac{\partial}{\partial w_j}\left[\sum_{j=1}^{M_c} \sum_{i=1}^{N} \ln(w_j) f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s) + \lambda\left(\sum_{j=1}^{M_c} w_j - 1\right)\right] = 0. \tag{31}$$

Using the fact that $\sum_{j=1}^{M_c} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s) = 1$, the above expression results in

$$w_j = \frac{1}{N} \sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s). \tag{32}$$

Taking the log of (4), ignoring any constant term (since they disappear after taking the derivatives), the second summation on the right-hand side of (30) becomes

$$\sum_{j=1}^{M_c} \sum_{i=1}^{N} \ln\left(f(\mathbf{z}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\right) f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)$$
$$= \sum_{j=1}^{M_c} \sum_{i=1}^{N} \left(\frac{1}{2}\ln\left(\det\left(\boldsymbol{\Sigma}_j^{-1}\right)\right)\right.$$
$$\left. - \frac{1}{2}(\mathbf{z}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_j)\right) f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s). \tag{33}$$

Let A be a square matrix, and then from matrix algebra, we have $\sum_i \mathbf{x}_i^T A \mathbf{x}_i = \text{tr}(A \sum_i \mathbf{x}_i \mathbf{x}_i^T)$. Defining $(\mathbf{z}_i - \boldsymbol{\mu}_j)(\mathbf{z}_i - \boldsymbol{\mu}_j)^T = P_{i,j}$, the right-hand side of (33) can be written as

$$\sum_{j=1}^{M_c} \left[\frac{1}{2}\ln\left(\det\left(\boldsymbol{\Sigma}_j^{-1}\right)\right) \sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)\right.$$
$$\left. - \frac{1}{2}\sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)\text{tr}\left(\boldsymbol{\Sigma}_j^{-1} P_{i,j}\right)\right]. \tag{34}$$

Taking the derivatives of (33) and (34) with respect to $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j^{-1}$, respectively, and setting them equal to zero, we get

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^{N} \mathbf{z}_i f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)}{\sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)} \tag{35}$$

$$\boldsymbol{\Sigma}_j = \frac{\sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)(\mathbf{z}_i - \boldsymbol{\mu}_j)(\mathbf{z}_i - \boldsymbol{\mu}_j)^T}{\sum_{i=1}^{N} f(j|\mathbf{z}_i, \boldsymbol{\gamma}^s)}. \tag{36}$$

Now the algorithm proceeds recursively from an initial starting point. One step of the recursion of (32), (35), and (36) results in

(11), (12), and (13), respectively. It should be noted that (36) is derived using the facts $\partial \ln(\det(A))/\partial A = 2A^{-1} - diag(A^{-1})$ and $\partial \mathrm{tr}(AB)/\partial A = B + B^T - diag(B)$.

REFERENCES

[1] Z. Fikri, "Statistical load analysis for distribution network planning," Ph.D. dissertation, Dept. Elect. Power Syst. Eng., Royal Inst. Technol., Stockholm, Sweden, 1975.

[2] Selected Statistical Methods for Analysis of Load Research Data, EPRI Rep. EA-3467, May 1984.

[3] G. W. Irwin, W. Monteith, and W. C. Beattle, "Statistical electricity demand modelling from consumer billing data," *Proc. Inst. Elect. Eng., Gen., Transm., Distrib.*, vol. 133, no. 6, pp. 328–335, 1986.

[4] A. Seppala, "Statistical distribution of customer load profiles," in *Proc. IEEE Int. Conf. Energy Management and Power Delivery*, Nov. 21–23, 1995, vol. 2, pp. 696–701.

[5] R. Herman and J. J. Kritzinger, "The statistical description of grouped domestic electrical load currents," *Elect. Power Syst. Res. J.*, vol. 27, pp. 43–48, 1993.

[6] A. K. Ghosh, D. L. Lubkeman, M. J. Downey, and R. H. Jones, "Distribution circuit state estimation using a probabilistic approach," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 45–51, Feb. 1997.

[7] S. W. Heunis and R. Herman, "A probabilistic model for residential consumer loads," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 621–625, Aug. 2002.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[9] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, Apr. 1984.

[10] J. A. Bilmes, A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, International Computer Science Institute, Tech. Rep. ICSI-TR-97-021, Apr. 1998.

[11] United Kingdom Generic Distribution Network (UKGDS). [Online]. Available: http://monaco.eee.strath.ac.uk/ukgds/.

[12] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.

[13] H. Gauvrit, J. P. Le Cadre, and C. Jauffret, "A formulation of multi-target tracking as an incomplete data problem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 4, pp. 1242–1257, Oct. 1997.

[14] G. McLachlan and K. Basford, *Mixture Models: Interference and Applications to Clustering*. New York: Marcel Dekker, 1988.

[15] N. Kehtarnavaz and E. Nakamura, "Generalization of the EM algorithm for mixture density estimation," *Pattern Recognit. Lett.*, vol. 19, no. 2, pp. 133–140, Feb. 1998.

[16] *"Matlab Statistical Toolbox 6, User's Guide."* [Online]. Available: http://www.mathworks.com.

[17] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. New York: Academic, 1979.

[18] D. J. Salmond, Mixture Reduction Algorithms for Uncertain Tracking, Royal Aerospace Establishment, Farnborough, U.K., Tech. Rep. 88004, Jan. 1988.

[19] N. D. Hatziargyriou, T. S. Karakatsanis, and M. Papadopoulos, "Probabilistic load flow in distribution systems containing dispersed wind power generation," *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 159–165, Feb. 1993.

[20] S. J. Huang and K. R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, May 2003.

[21] C. Alvarez, R. P. Malhame, and A. Gabaldon, "A class of models for load management application and evaluation revisited," *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1435–1443, Nov. 1992.

[22] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. New York: Marcel Dekker, 2004.

[23] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, "Modelling of pseudo measurements for distribution system state estimation," in *Proc. CIRED*, Frankfurt, Germany, 2008.

[24] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 911–916, May 1996.

Ravindra Singh (S'07) received the B.Tech. (with honors) degree from HBTI, Kanpur, India, and the MSc. (Engg.) degree from the Indian Institute of Science, Bangalore, India, in 2002, and 2004, respectively. Currently, he is pursuing the Ph.D. degree at the Imperial College London, London, U.K.

From 2004 to 2006, he was with Honeywell Research and Technology Group, Bangalore, India. His current research interests include distribution system state estimation.

Bikash C. Pal (M'00–SM'02) received the B.E.E. (with honors) degree from Jadavpur University, Calcutta, India, the M.E. degree from the Indian Institute of Science, Bangalore, India, and the Ph.D. degree from the Imperial College London, London, U.K., in 1990, 1992, and 1999, respectively, all in electrical engineering.

Currently, he is a Reader in the Department of Electrical and Electronic Engineering, Imperial College London. His current research interests include state estimation, power system dynamics, and flexible ac transmission system controllers.

Rabih A. Jabr (M'02) received the B.E. degree in electrical engineering (with high distinction) from the American University of Beirut, Beirut, Lebanon, in 1997, and the Ph.D. degree in electrical engineering from Imperial College London, London, U.K., in 2000.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering at the American University of Beirut, Beirut, Lebanon. His research interests are in mathematical optimization techniques, design optimization, optimal power flow, and state estimation.

Dr. Jabr received the Eryl Cadwaladar Davies Prize (2000) from the Imperial College and the Sebastien De Ferranti Premium Award (2003) from the Institution of Electrical Engineers (U.K.) for his work on optimal security dispatch.