

# Comparison of Gaussian Mixture Reductions for Probabilistic Studies in Power Systems

G. Valverde, *Member, IEEE*, J. Quirós Tortós, *Student Member, IEEE*, V. Terzija, *Senior Member, IEEE*

**Abstract**—This paper presents the comparison of three pair-merging methods to reduce the number of Gaussian mixture components used to model non-Gaussian Probabilistic Density Function (PDF) of random power system variables such as power demands, wind power outputs or other intermittent power sources. It also introduces a fine-tuning algorithm to improve the solution of the pair-merging methods to better approximate the original Gaussian mixture. A Gaussian mixture distribution with seven components is used to validate and demonstrate the algorithms.

**Index Terms**—Gaussian mixture model, probabilistic density function, probabilistic power flow, wind power.

## I. INTRODUCTION

GAUSSIAN mixture distributions can be used to model any random variable in power systems for a certain period. This is of great interest because modern power systems are operated under high level of uncertainties, due to large incorporation of intermittent generation, the proliferation of storage devices, and the variability of aggregated loads.

In order to analyse all the possible combinations of uncertainties, probabilistic studies have been widely used to analyse the interaction of random variables for different scenarios of load and generation.

An example of these studies is the Probabilistic Power Flow (PPF) analysis. These studies determine the most likely power flows and bus voltages in the network given a set of random input variables. The advantage of the PPF methods is the simplicity to determine the possible ranges of the power flows by running a finite number of simulations [1].

During the last few decades, the PPF studies were used to study the variability of aggregated loads modelled by Gaussian distributions. With the insertion of intermittent generation, the probabilistic studies have gained more attention due to the need of modelling the intermittent power output as random variables that are typically non-Gaussian distributed [2], [3].

Because of the proliferation of these renewable sources, the representation of these non-Gaussian PDFs is an open field of research.

Different approximations have been developed to model non-Gaussian input random variables in power systems. For instance, in order to model the variability of wind power output, an indirect algorithm based on the *Beta* distribution was proposed in [4] and later considered in [2].

The probability distribution of the wind speed is typically non-Gaussian and it has been modelled by the *Gamma*, *Weibull* or the *Rayleigh* distributions [5]. The *Weibull* distribution has demonstrated better results because of its two flexible parameters  $k$  and  $c$  [6]. Nonetheless, since the wind speed PDF cannot be always approximated as a *Weibull* distribution, a mixed *Gamma-Weibull* distribution and a mixed truncated *Normal* distribution was introduced in [6].

The PDFs of power demands of aggregated loads can be also non-Gaussian distributed. For example, the *Normal*, *log-normal* and *Beta* distributions were used to evaluate their effectiveness to model the load uncertainty [7]. Because of its flexibility to adapt to the skewness of the distribution, the *Beta* distribution was found to be the most appropriate. This distribution was also used in [8] to model the variability of load demand.

Recently, a more accurate approximation of the marginal distribution of any power demand was introduced in [9]. As the PDF of load demands cannot be represented by a specific distribution, the authors proposed the use of the Gaussian mixture distribution. Although the work in [9] concentrates on the probability distribution of power demands, Gaussian mixtures can be used to represent the variability of any non-Gaussian variable.

This paper starts from the latter affirmation: it assumes that the marginal distribution of any wind farm power output, wind speed or power demands can be represented by Gaussian mixtures.

As the Gaussian mixtures are made of a finite number of components, this paper explores three different methods to reduce the number of components while keeping a good level of accuracy of the original non-Gaussian PDF. The proposed reductions are used to minimize the data used to represent the random input variables or when the number of Gaussian components becomes a computational constraint [10].

The reduction algorithms presented in this paper replace any GMM by another with fewer components. In order to improve the accuracy of the approximations, this paper also presents a fine-tuning algorithm based on optimization theory.

Section II introduces the Gaussian mixture distribution. Section III presents the different Gaussian mixture reduction methods. Later, in order to demonstrate the capabilities of the

---

This work was supported in part by the Engineering and Physical Science Research Council (EPSRC) in the UK and The University of Costa Rica.

G. Valverde, J. Quirós Tortós and V. Terzija are with the School of Electrical & Electronic Engineering, The University of Manchester, Manchester, UK (e-mail: gvalverde@ieee.org, jairoquiortortos@ieee.org, terzija@ieee.org).

reduction methods, a seven component Gaussian mixture is reduced in Section IV. Finally, the conclusions drawn from the study are presented in Section V.

## II. GAUSSIAN MIXTURE DISTRIBUTION

The mixture of  $L$  weighted Gaussian distributions creates a Gaussian mixture distribution  $f(y)$ . The probability density function  $f_Y(y)$ , for a given random variable  $Y$ , is defined as [11]:

$$f_Y(y) = \sum_{i=1}^L \omega_i f_{N(\mu_i, \sigma_i^2)}(y) \quad (1)$$

where  $0 \leq \omega_i \leq 1$ ,  $\mu_i$  and  $\sigma_i^2$  are the proportion, mean and variance of the  $i$ -th component of the Gaussian mixture, respectively. The distribution of the  $i$ -th Gaussian component is [12]:

$$f_{N(\mu_i, \sigma_i^2)}(y) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-\mu_i)^2}{2\sigma_i^2}} \quad (2)$$

where the mean and variance of the random variable  $Y$  are [13]:

$$\mu_Y = \sum_{i=1}^L \omega_i \mu_i \quad (3)$$

$$\sigma_Y^2 = \sum_{i=1}^L \omega_i (\sigma_i^2 + (\mu_i - \mu_Y)^2) \quad (4)$$

Consider a Gaussian mixture distribution ( $L = 7$ ) with the parameters given in Table I. The PDF of the Gaussian mixture is obtained by evaluating  $f_Y(y)$  for  $-\infty < y < \infty$ , as presented in Fig. 1.

TABLE I

PARAMETERS OF A GAUSSIAN MIXTURE DISTRIBUTION WITH 7 COMPONENTS

| Component    | $i=1$ | $i=2$ | $i=3$ | $i=4$ | $i=5$ | $i=6$ | $i=7$ |
|--------------|-------|-------|-------|-------|-------|-------|-------|
| $\omega_i$   | 0.20  | 0.20  | 0.20  | 0.20  | 0.10  | 0.05  | 0.05  |
| $\mu_i$      | 43.0  | 50.0  | 52.0  | 58.0  | 62.0  | 72.0  | 80.0  |
| $\sigma_i^2$ | 9.0   | 16.0  | 9.0   | 64.0  | 4.0   | 4.0   | 25.0  |

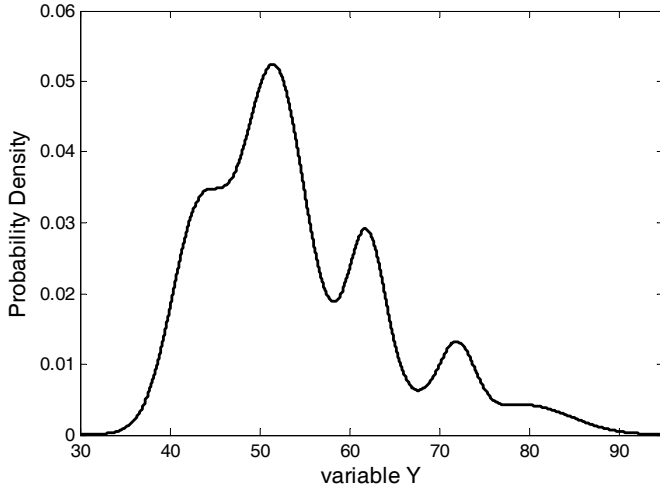


Fig. 1. Gaussian mixture distribution with 7 Gaussian components

The Gaussian mixture distribution is created by summing the individual weighted Gaussian components. This generic variable  $Y$  represents any random input variable (wind speed, generation power output or power demand). As it is presented

in Fig. 1, the PDF of the variable  $Y$  cannot be approximated by any other typical marginal distribution.

The possibility to approximate any marginal distribution with a finite number of components is the main advantage of the Gaussian mixture distribution, usually referred as the Gaussian Mixture Model (GMM).

The Expectation Maximization (EM) algorithm is the most effective methodology to determine the GMM that best approximates the distribution of the samples of  $Y$  [9].

The input of the EM algorithm is the set of incomplete samples  $\gamma = [y_1, y_2, \dots, y_s]$  of the random variable  $Y$  and the desired number of Gaussian components of the GMM. Given  $\gamma$  and the initial (or updated) Gaussian mixture parameters  $\eta$ , the algorithm computes the expectation of the log-likelihood of the complete data with respect to the unknown samples. Later,  $\eta$  is updated to maximize the log-likelihood expectation previously determined. This procedure is iteratively executed until convergence is achieved. The Statistics Toolbox of MATLAB offers the function *gmdistribution.fit* to estimate  $\eta$  using the EM algorithm given the incomplete samples  $\gamma$  and the desired number of Gaussian components [14].

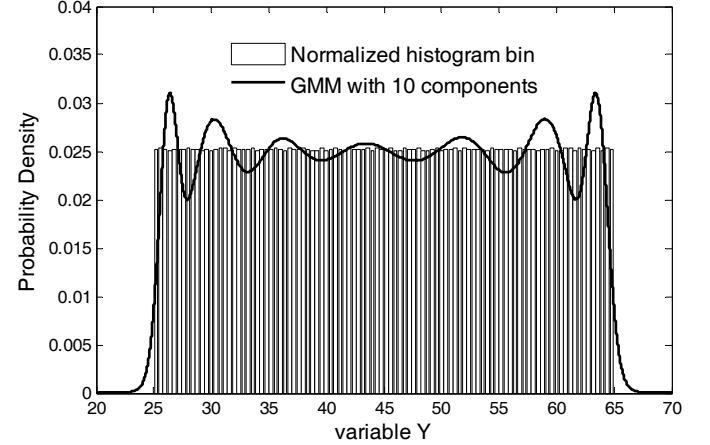


Fig. 2. Uniform distributed random variable modelled by GMM

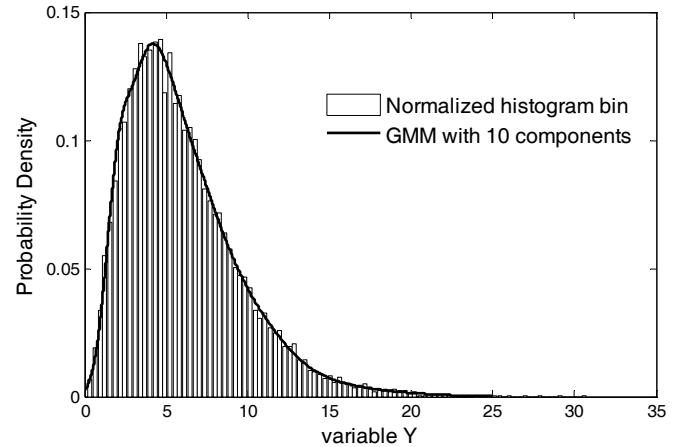


Fig. 3. Gamma distributed random variable modelled by GMM

Fig. 2 presents the resulting GMM to approximate a Uniform random variable using the EM algorithm. In fact, the Uniform distribution is one of the most difficult distributions

to approximate. As it can be seen, the approximation could be improved with a higher number of Gaussian components.

Fig. 3 presents the GMM approximation of a Gamma distributed random variable with ten components. In this case, the mismatch between the GMM and the random variable's density is negligible.

The required level of approximation determines the number of components needed to approximate any marginal distribution. In order to quantify the degree of fitness of the set of samples and the GMM, the Chi-Square goodness of fit test can be used [9].

### III. REDUCTION OF GAUSSIAN MIXTURES

As stated above, the number of components of the GMM affects the approximation of the non-Gaussian PDF. Although using many components appears to be the best option, reduced GMMs (with fewer components) are usually desired in order to limit the size and complexity of the problem, particularly if the study involves many GMMs at the same time.

The idea of the Gaussian mixture components reduction is to approximate the original GMM in (1), as a new GMM with fewer components:

$$g_Y(y) = \sum_{j=1}^M \tilde{\omega}_j f_{N(\tilde{\mu}_j, \tilde{\sigma}_j^2)}(y) \quad (5)$$

where  $M < L$  and  $\tilde{\omega}_j$ ,  $\tilde{\mu}_j$  and  $\tilde{\sigma}_j^2$  are the proportion, mean and variance of the  $j$ -th component, respectively. The main advantage of using these reduction algorithms is that it is not required to run the EM algorithm again. The reduction methods use the parameters of the original GMM and do not consider the original samples of  $Y$  any longer.

Diverse algorithms to reduce the number of components are found in the literature. The goal is to maintain the mean and variance of the original mixture. If not, the deviation should be negligible and the resulting GMM should properly represent the structure of the original mixture [13].

To accomplish these requirements, one single algorithm might not be enough. Thus, it is common to employ a mixture of algorithms that can achieve the highest number of requirements.

The simplest method to reduce Gaussian mixture components is based on merging pairs of components. This approach starts by defining  $g_Y(y) = f_Y(y)$ . The merging is applied to the pair of components of  $g_Y(y)$  that produce the minimum difference between  $g_Y(y)$  and  $f_Y(y)$  [15], [16].

Once the pair of components  $i$  and  $j$  are identified, they are merged together to obtain a new resulting component  $ij$  with the following component parameters [16]:

$$\tilde{\omega}_{ij} = \tilde{\omega}_i + \tilde{\omega}_j \quad (6)$$

The mean and the variance of the new Gaussian component are:

$$\tilde{\mu}_{ij} = \frac{1}{\tilde{\omega}_i + \tilde{\omega}_j} (\tilde{\omega}_i \tilde{\mu}_i + \tilde{\omega}_j \tilde{\mu}_j) \quad (7)$$

$$\tilde{\sigma}_{ij}^2 = \frac{1}{\tilde{\omega}_i + \tilde{\omega}_j} \left( \tilde{\omega}_i \tilde{\sigma}_i^2 + \tilde{\omega}_j \tilde{\sigma}_j^2 + \frac{\tilde{\omega}_i \tilde{\omega}_j}{\tilde{\omega}_i + \tilde{\omega}_j} (\tilde{\mu}_i - \tilde{\mu}_j)^2 \right) \quad (8)$$

The resulting component  $ij$  replaces the  $i$ -th and  $j$ -th components, which have been previously identified. Although  $g_Y(y)$  has lost one component, the mean and the variance are the same from the original mixture. The iterative process stops when the desired number of components is reached.

The identification of the  $i$ -th and  $j$ -th components of  $g_Y(y)$  depends on the selection criteria. The Salmond method identifies the pair of components  $i$  and  $j$  that produce the minimum increase of the first summand in (4) [17]. The increase of the first summand in (4) is related to the Square Distance (SD) measure:

$$d_{ij}^2 = \frac{\tilde{\omega}_i \tilde{\omega}_j}{\tilde{\omega}_i + \tilde{\omega}_j} \cdot \frac{(\tilde{\mu}_i - \tilde{\mu}_j)^2}{\sigma_Y^2} \quad (9)$$

The smallest squared distance identifies the merged components  $i$  and  $j$  that produce the minimum value of the cost function. The main disadvantage of this methodology is that it merges the pair of components with the closest means even if their variances are very different [16].

As an alternative, the Williams method uses the Integral Square Difference (ISD or  $J_s$ ) to evaluate the difference between the original  $f_Y(y)$  and the reduced Gaussian mixture  $g_Y(y)$  [15]:

$$J_s = \int (f_Y(y) - g_Y(y))^2 dy \quad (10)$$

Equation (10) is used for all combinations of pairs of components of  $g_Y(y)$ . The components  $i$  and  $j$  to be merged are those that when merged produce the minimum  $J_s$ . The iterative process stops when  $g_Y(y)$  is reduced to the desired number of components. The ISD defined in (10) is calculated by:

$$\begin{aligned} J_s = & \sum_{i=1}^L \sum_{j=1}^L \omega_i \omega_j f_{N(\mu_j, \sigma_i^2 + \sigma_j^2)}(\mu_i) \\ & + \sum_{i=1}^L \sum_{j=1}^M \omega_i \tilde{\omega}_j f_{N(\tilde{\mu}_j, \sigma_i^2 + \tilde{\sigma}_j^2)}(\mu_i) \\ & + \sum_{i=1}^M \sum_{j=1}^M \tilde{\omega}_i \tilde{\omega}_j f_{N(\tilde{\mu}_j, \tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)}(\tilde{\mu}_i) \end{aligned} \quad (11)$$

The ISD selection method takes into account the whole Gaussian mixture to decide which components to merge. However, it is more time consuming than the Salmond method as more calculations are required.

An alternative measure of similarity between two probability densities  $f_1(y)$  and  $f_2(y)$  is the Kullback-Leibler (KL) divergence  $D(f_1 \| f_2)$  [16]:

$$D(f_1 \| f_2) = \int f_1(y) \ln \frac{f_1(y)}{f_2(y)} dy \quad (12)$$

Contrary to the ISD measure, there is no closed form expression for the KL divergence measure when  $f_1(y)$  and  $f_2(y)$  are Gaussian mixtures. Because of this limitation, an upper bound on the discrepancy of the reduced mixture from the mixture before the merge was proposed in [16]. The upper bound measure of discrepancy is defined as follows:

$$B_{ij} = \frac{\tilde{\omega}_i + \tilde{\omega}_j}{2} \cdot \ln \left[ \tilde{\omega}_{|ij} \left( \frac{\tilde{\sigma}_i^2}{\tilde{\sigma}_j^2} \right)^{\tilde{\omega}_{|ij}} \right. \\ \left. + \tilde{\omega}_{j|ij} \left( \frac{\tilde{\sigma}_j^2}{\tilde{\sigma}_i^2} \right)^{\tilde{\omega}_{j|ij}} + \tilde{\omega}_{|ij} \cdot \tilde{\omega}_{j|ij} \frac{(\tilde{\mu}_i - \tilde{\mu}_j)^2}{\tilde{\sigma}_i^2 \tilde{\omega}_{|ij} \tilde{\sigma}_j^2 \tilde{\omega}_{j|ij}} \right] \quad (13)$$

where

$$\tilde{\omega}_{i|ij} = \frac{\tilde{\omega}_i}{\tilde{\omega}_i + \tilde{\omega}_j}; \quad \tilde{\omega}_{j|ij} = \frac{\tilde{\omega}_j}{\tilde{\omega}_i + \tilde{\omega}_j} \quad (14)$$

This KL based selection method identifies the merged components  $i$  and  $j$  that produce the minimum upper bound discrepancy between the mixture after the merge and the mixture before the merge. Equation (13) takes into account the means, weights and variances of the Gaussian components. In addition, it requires fewer computations as compared to the ISD based method.

In order to compare the accuracy of the approximations, the ISD is now normalized into  $0 \leq J_s^N \leq 1$ , as defined in (15) [18]:

$$J_s^N = \frac{\int (f_Y(y) - g_Y(y))^2 dy}{\int f_Y(y)^2 dy + \int g_Y(y)^2 dy} \quad (15)$$

when  $J_s^N = 0$ , the reduced Gaussian mixture perfectly matches the original density. On the contrary,  $J_s^N = 1$  when there is zero overlapping between the GMMs.

Although pair-merging algorithms are easy to execute and they maintain the mean and variance of the original mixture, the structure of the resulting mixture is usually not optimal. The following Section presents an optimal based solution to improve the accuracy of the reduced GMMs.

#### A. Fine Tuning of GMM Reductions

The parameter obtained from pair-merging methods can be fine-tuned to achieve better approximations of the original Gaussian mixture [19]. The purpose is to correct the set of parameters  $\eta$  of  $g_Y(y)$  to minimise the ISD cost function (10) as a function of  $\eta$ :

$$J_s(y, \eta) = \int (f_Y(y) - g_Y(y, \eta))^2 dy \quad (16)$$

where:

$$g_Y(y, \eta) = \sum_{j=1}^M \tilde{\omega}_j^2 f_{N(\tilde{\mu}_j, \tilde{\sigma}_j^2)}(y) \quad (17)$$

$$\eta = [\eta_1^T \quad \eta_2^T \quad \dots \quad \eta_M^T]^T \quad (18)$$

$$\eta_j = [\tilde{\omega}_j \quad \tilde{\mu}_j \quad \tilde{\sigma}_j^2]^T \quad \forall j=1, \dots, M \quad (19)$$

In (17),  $\tilde{\omega}_j^2 \in [0, 1]$  is introduced to guarantee positive proportion of the components [20].

By introducing a perturbation  $\eta = \bar{\eta} + \Delta\eta$  around the original point  $\bar{\eta}$  and by considering only the linear term of the Taylor's series [19]:

$$J_s(y, \eta) = \int \left[ \begin{matrix} f_Y(y) - g_Y(y, \bar{\eta}) \\ - \left( \frac{\partial g_Y(y, \bar{\eta})}{\partial \eta} \right)^T (\eta - \bar{\eta}) \end{matrix} \right]^2 dy \quad (20)$$

it is possible to find the optimal set of parameters  $\eta$  when the first derivative of (20) with respect to  $\eta$  is equal to zero:

$$\underbrace{\int (f_Y(y) - g_Y(y, \bar{\eta})) \frac{\partial g_Y(y, \bar{\eta})}{\partial \eta} dy}_{\mathbf{h}(\bar{\eta})} = \underbrace{\int \frac{\partial g_Y(y, \bar{\eta})}{\partial \eta} \left( \frac{\partial g_Y(y, \bar{\eta})}{\partial \eta} \right)^T dy}_{\mathbf{P}(\bar{\eta})} \cdot \underbrace{(\eta - \bar{\eta})}_{\Delta\eta} \quad (21)$$

Or in more compact form:

$$\mathbf{h}(\bar{\eta}) = \mathbf{P}(\bar{\eta}) \Delta\eta \quad (22)$$

Vector  $\mathbf{h}(\cdot)$  contains  $M$  sub-vectors. Each  $3 \times 1$  sub-vector  $\mathbf{h}_j(\cdot)$  corresponds to one  $\eta_j$ , as follows [20]:

$$\mathbf{h}_j(\bar{\eta}) = \int (f_Y(y) - g_Y(y, \bar{\eta})) \cdot g_j(y, \bar{\eta}_j) \begin{bmatrix} \frac{2}{\tilde{\omega}_j} \\ \frac{y - \tilde{\mu}_j}{\tilde{\sigma}_j^2} \\ \frac{(y - \tilde{\mu}_j)^2 - \tilde{\sigma}_j^2}{\tilde{\sigma}_j^3} \end{bmatrix} dy \quad (23)$$

with  $g_j(y, \bar{\eta}_j) = \tilde{\omega}_j^2 \cdot f_{N(\tilde{\mu}_j, \tilde{\sigma}_j^2)}(y)$ . By solving the integral in (23), equation (24) can be obtained (see at the bottom of the page). The elements in (24) are:

$$\alpha_{ij} = f_{N(\tilde{\mu}_j, \tilde{\sigma}_j^2 + \tilde{\sigma}_i^2)}(\mu_i) \quad (25)$$

$$\tilde{\alpha}_{ij} = f_{N(\tilde{\mu}_j, \tilde{\sigma}_j^2 + \tilde{\sigma}_i^2)}(\tilde{\mu}_i) \quad (26)$$

$$\sigma_{ij}^2 = (1/\sigma_i^2 + 1/\tilde{\sigma}_j^2)^{-1} \quad (27)$$

$$\tilde{\sigma}_{ij}^2 = (1/\tilde{\sigma}_i^2 + 1/\tilde{\sigma}_j^2)^{-1} \quad (28)$$

$$\mu_{ij} = \sigma_{ij}^2 (\mu_i / \sigma_i^2 + \tilde{\mu}_j / \tilde{\sigma}_j^2) \quad (29)$$

$$\tilde{\mu}_{ij} = \tilde{\sigma}_{ij}^2 (\tilde{\mu}_i / \tilde{\sigma}_i^2 + \tilde{\mu}_j / \tilde{\sigma}_j^2) \quad (30)$$

---


$$\mathbf{h}_j(\bar{\eta}) = \begin{bmatrix} \frac{2}{\tilde{\omega}_j} & 0 & 0 \\ \frac{-\tilde{\mu}_j}{\tilde{\sigma}_j^2} & \frac{1}{\tilde{\sigma}_j^2} & 0 \\ \frac{\tilde{\mu}_j^2 - \tilde{\sigma}_j^2}{\tilde{\sigma}_j^3} & \frac{-2\tilde{\mu}_j}{\tilde{\sigma}_j^3} & \frac{1}{\tilde{\sigma}_j^3} \end{bmatrix} \begin{bmatrix} \tilde{\omega}_j^2 \sum_{i=1}^L \omega_i \alpha_{ij} - \tilde{\omega}_j^2 \sum_{i=1}^M \tilde{\omega}_i \tilde{\alpha}_{ij} \\ \tilde{\omega}_j^2 \sum_{i=1}^L \omega_i \alpha_{ij} \mu_{ij} - \tilde{\omega}_j^2 \sum_{i=1}^M \tilde{\omega}_i \tilde{\alpha}_{ij} \tilde{\mu}_{ij} \\ \tilde{\omega}_j^2 \sum_{i=1}^L \omega_i \alpha_{ij} (\mu_{ij}^2 + \sigma_{ij}^2) - \tilde{\omega}_j^2 \sum_{i=1}^M \tilde{\omega}_i \tilde{\alpha}_{ij} (\tilde{\mu}_{ij}^2 + \tilde{\sigma}_{ij}^2) \end{bmatrix} \quad (24)$$

Returning to (22), Matrix  $\mathbf{P}(\cdot)$  is:

$$\mathbf{P}(\bar{\eta}) = \begin{bmatrix} \mathbf{P}^{(1,1)} & \dots & \mathbf{P}^{(1,M)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}^{(M,1)} & \dots & \mathbf{P}^{(M,M)} \end{bmatrix} \quad (28)$$

and each sub-matrix  $\mathbf{P}^{(i,j)}$  in (28) is defined by:

$$\mathbf{P}^{(i,j)} = \tilde{\omega}_i \tilde{\omega}_j f_{N(\tilde{\mu}_j, \tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)}(\tilde{\mu}_i) \begin{bmatrix} \mathbf{P}_{1,1}^{(i,j)} & \mathbf{P}_{1,2}^{(i,j)} & \mathbf{P}_{1,3}^{(i,j)} \\ \mathbf{P}_{2,1}^{(i,j)} & \mathbf{P}_{2,2}^{(i,j)} & \mathbf{P}_{2,3}^{(i,j)} \\ \mathbf{P}_{3,1}^{(i,j)} & \mathbf{P}_{3,2}^{(i,j)} & \mathbf{P}_{3,3}^{(i,j)} \end{bmatrix} \quad (29)$$

where the elements in (29) are defined according to (30), (31) and (32) (see at the bottom of the page).

$$\begin{aligned} \mathbf{P}_{1,1}^{(i,j)} &= 4, \\ \mathbf{P}_{1,2}^{(i,j)} &= 2\tilde{\omega}_j \frac{\tilde{\mu}_i - \tilde{\mu}_j}{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2}, \\ \mathbf{P}_{1,3}^{(i,j)} &= 2\tilde{\omega}_j \tilde{\sigma}_i^2 \frac{(\tilde{\mu}_i - \tilde{\mu}_j)^2 - (\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)}{(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)^2}, \end{aligned} \quad (30)$$

$$\begin{aligned} \mathbf{P}_{2,1}^{(i,j)} &= 2\tilde{\omega}_i \frac{\tilde{\mu}_j - \tilde{\mu}_i}{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2}, \\ \mathbf{P}_{2,2}^{(i,j)} &= \tilde{\omega}_i \tilde{\omega}_j \frac{(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2) - (\tilde{\mu}_i - \tilde{\mu}_j)^2}{(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)^2}, \end{aligned} \quad (31)$$

$$\mathbf{P}_{2,3}^{(i,j)} = \tilde{\omega}_i \tilde{\omega}_j \tilde{\sigma}_j \frac{(\tilde{\mu}_j - \tilde{\mu}_i) \cdot ((\tilde{\mu}_i - \tilde{\mu}_j)^2 - 3(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2))}{(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)^3},$$

By using LU decomposition of  $\mathbf{P}$ , the set of linear equations in (22) is solved. From (33), the new set of parameters is obtained at the  $k$ -th iteration:

$$\eta^{k+1} = \eta^k + \Delta \eta \quad (33)$$

If the initial set of parameters  $\eta^0$  is far from the optimal solution, the method may converge to a local minimum or it may not converge at all. In order to avoid these problems, the solution obtained from the pair-merging algorithms can be used as the initial guess.

#### IV. RESULTS

This Section compares the performance of the reduction methods presented above, by reducing the 7-component Gaussian mixture presented in Fig. 1.

##### A. Reduction Results

The 7-components Gaussian Mixture is approximated by five, four and three component mixtures, as presented in Figs. 4, 5 and 6, respectively. The merged components were selected from the SD (9), the ISD (10) and the KL (12) upper bound measures.

As presented in Fig. 4, the reduced Gaussian mixture obtained from the ISD measure is the same as the resulting Gaussian mixture obtained from the KL upper bound measure.

If the original Gaussian mixture is approximated by four Gaussian components, as presented in Fig. 5, the SD and the KL upper bound selection methods obtained the same reduced distribution.

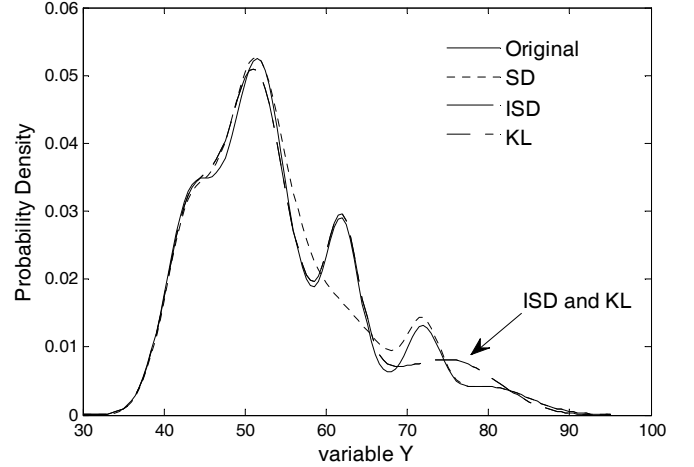


Fig. 4. GMM reduction using five components

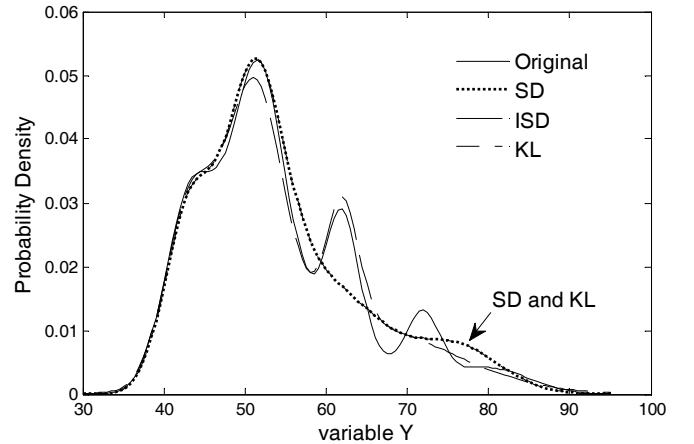


Fig. 5. GMM reduction using four components

As it is presented in Fig. 6, the reduced mixture obtained from the ISD and the KL upper bound discrimination methods

$$\begin{aligned} \mathbf{P}_{3,1}^{(i,j)} &= 2\tilde{\omega}_i \tilde{\sigma}_i \frac{(\tilde{\mu}_i - \tilde{\mu}_j)^2 - (\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)}{(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)^2}, \\ \mathbf{P}_{3,2}^{(i,j)} &= \tilde{\omega}_i \tilde{\omega}_j \tilde{\sigma}_i \frac{(\tilde{\mu}_i - \tilde{\mu}_j) \cdot ((\tilde{\mu}_i - \tilde{\mu}_j)^2 - 3(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2))}{(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)^3}, \\ \mathbf{P}_{3,3}^{(i,j)} &= \tilde{\omega}_i \tilde{\omega}_j \tilde{\sigma}_i \tilde{\sigma}_j \frac{(\tilde{\mu}_i - \tilde{\mu}_j)^4 + 3(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)((\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2) - 2(\tilde{\mu}_i - \tilde{\mu}_j)^2)}{(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)^4} \end{aligned} \quad (32)$$

are exactly the same. Although it is possible to visualize the resulting Gaussian mixtures from Figs. 4-6, the methods have not been quantitatively compared against each other.

Table II presents the comparison of the  $J_s^N$  for reduced components with  $M$  components obtained by merging pair of components based on the SD, ISD and KL upper bound measures. When fewer components are used to model the original density  $f_Y(y)$ , the approximation becomes less accurate and  $J_s^N$  approaches 1.

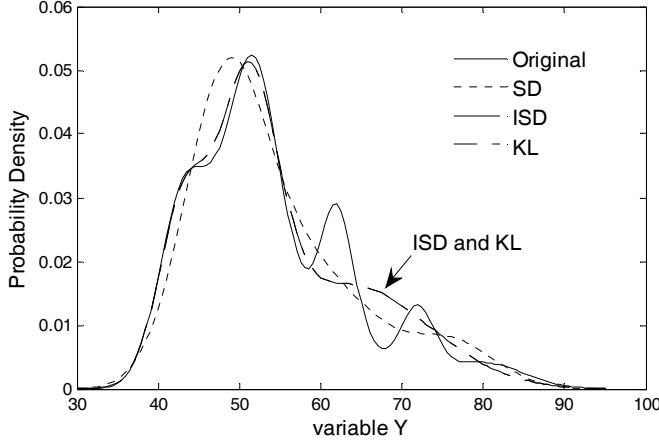


Fig. 6. GMM reduction using three components

As it is presented in Table II, the minimum  $J_s^N$  is always obtained when the ISD discrimination measured is used. It is important to mention that the KL upper bound measure selected the same pair of components as the ISD measure did, excepted when  $M = 4$ .

TABLE II

$J_s^N$  FOR RESULTING  $g_Y(y)$  AND THE ORIGINAL MIXTURE  $f_Y(y)$

| Method     | $M = 6$         | $M = 5$       | $M = 4$       | $M = 3$       | $M = 2$       | $M = 1$       |
|------------|-----------------|---------------|---------------|---------------|---------------|---------------|
| SD         | 6.93E-04        | 0.0094        | 0.0106        | 0.0201        | 0.0231        | 0.0589        |
| <b>ISD</b> | <b>6.93E-04</b> | <b>0.0027</b> | <b>0.0039</b> | <b>0.0118</b> | <b>0.0231</b> | <b>0.0589</b> |
| KL         | 6.93E-04        | 0.0027        | 0.0106        | 0.0118        | 0.0231        | 0.0589        |

Table III presents the required computation time, in milliseconds. For all the cases, the KL upper bound discrimination method required less time to find the pair of components to merge. The higher processing times of the ISD measure can be a constraint if the number of components is large e.g. 100 components reduced into 25 components. In this case, the KL upper bound algorithm will be much more efficient.

TABLE III

COMPARISON OF COMPUTER TIME REQUIREMENTS IN MILLISECONDS

| Method    | $M = 6$       | $M = 5$       | $M = 4$       | $M = 3$       | $M = 2$       | $M = 1$       |
|-----------|---------------|---------------|---------------|---------------|---------------|---------------|
| SD        | 17.200        | 21.953        | 21.861        | 20.796        | 20.154        | 21.869        |
| ISD       | 46.300        | 70.542        | 79.429        | 85.214        | 87.209        | 90.478        |
| <b>KL</b> | <b>11.400</b> | <b>14.359</b> | <b>14.363</b> | <b>14.214</b> | <b>13.905</b> | <b>13.639</b> |

### B. Fine tuning of GMM reductions

Fig. 7 presents the optimal based reduction method compared to the original GMM. In this algorithm, the ISD reduction method was used to calculate the initial set of parameters  $\eta^0$ .

As it is presented in Fig. 7, the optimized reduction is more accurate and the resulting  $g_Y(y)$  better approximates the

original  $f_Y(y)$ . Table IV presents the  $J_s^N$  before and after the fine tuning of the Gaussian mixture reduction.

For all the cases, the fine tuning method makes  $J_s^N$  closer to zero. In the case when  $M = 5$ , the optimization method reduced  $M$  to 4 components. Such elimination is performed when the ratio  $\hat{\sigma}_j^2 / \bar{\sigma}_j^2$  of the  $j$ -th component of  $g_Y(y)$  tends to zero [20]. This is the reason why  $J_s^N$  for  $M = 5$  and  $M = 4$  are the same.

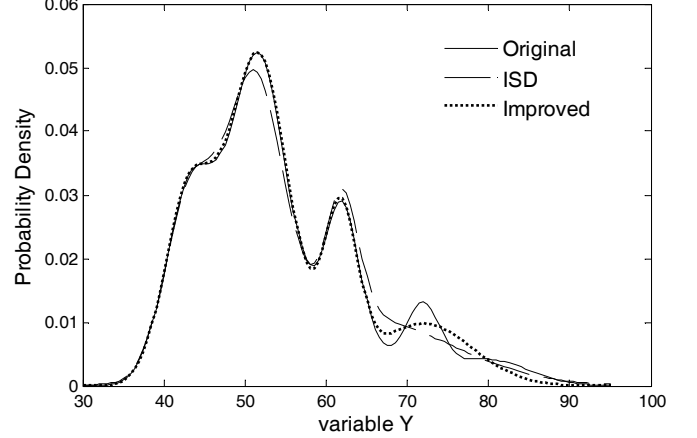


Fig. 7. Original GMM reduced to 4 components using the optimal based method

TABLE IV

$J_s^N$  FOR THE RESULTING IMPROVED  $g_Y(y)$  AND THE ORIGINAL MIXTURE  $f_Y(y)$

| Method          | $M = 6$         | $M = 5$       | $M = 4$       | $M = 3$ | $M = 2$ | $M = 1$ |
|-----------------|-----------------|---------------|---------------|---------|---------|---------|
| ISD             | 6.93E-04        | 0.0027        | 0.0039        | 0.0118  | 0.0231  | 0.0589  |
| <b>Improved</b> | <b>8.59E-07</b> | <b>0.0011</b> | <b>0.0011</b> | 0.0085  | 0.0187  | 0.0321  |

## V. CONCLUSION

This paper explores and compares different methods to simplify Gaussian Mixture Models (GMMs) to represent non-Gaussian power system random variables. These methods reduce the number of components of the GMMs and keep a good approximation of the original distributions.

It was found that the Integral Squared Difference (ISD) discrimination method always identified the pair of components that when merged produced the minimum difference between the original and the reduced Gaussian mixture. However, this method was found to be time consuming if compared to the KL upper bound or the Squared Distance (SD) algorithms.

The KL upper bound algorithm was found to be very efficient in terms of computational demands and accuracy. This can be the best choice if the number of Gaussian components to merge is high.

The fine tuning method obtains better reductions than any pair-merging method. However, the pair merging methods can be used as an initial guess to ensure the convergence of the optimal based method.

## VI. REFERENCES

- [1] J.M. Morales and J.Pérez-Ruiz, "Point estimate schemes to solve the probabilistic power flow," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1594-1601, Nov. 2007.

- [2] J. Usaola, "Probabilistic load flow with wind production uncertainty using cumulants and Cornish-Fisher expansion," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 9, pp. 474-481, Oct. 2009.
- [3] M. Matos and R. J. Bessa, "Setting the operating reserve using probabilistic wind power forecasts," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 594-603, May 2011.
- [4] H. Bludszweit, J.A. Dominguez-Navarro and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983-991, Aug. 2008.
- [5] D. Villanueva, J.L. Pazos and A. Feijoo, "Probabilistic load flow including wind power generation," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1659-1667, Aug. 2011.
- [6] T.P. Chang, "Estimation of wind energy potential using different probability density functions," *Applied Energy*, vol. 88, no. 5, pp. 1848-1856, May 2011.
- [7] A.K. Ghosh, D.L. Lubkeman, M.J. Downey, and R.H. Jones, "Distribution circuit state estimation using a probabilistic approach," *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 45-51, Feb. 1997.
- [8] S.W. Heunis and R. Herman, "A probabilistic model for residential consumer loads," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 621-625, Aug. 2002.
- [9] R. Singh, B.C. Pal and R.A. Jabr, "Statistical representation of distribution system loads using Gaussian Mixture Model," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29-37, Feb. 2010.
- [10] G. Valverde, S. Saric and V. Terzija, "Probabilistic load flow with non-Gaussian correlated random variables using Gaussian Mixture Models," *IET Generation, Transmission and Distribution*, to be published.
- [11] H.R. Sirisena and E.P.M Brown, "Representation of non-Gaussian probability distributions in stochastic load-flow studies by the method of Gaussian sum approximations," *IEE Proceedings Generation, Transmission and Distribution*, vol. 130, no. 4, pp. 165-171 Jul. 1983.
- [12] D.C. Montgomery and G.C. Runger, *Applied Statistics and Probability for Engineers*, 3rd ed. New York: John Wiley & Sons. Inc, 2002.
- [13] D.J. Salmond, "Mixture reduction algorithms for point and extended object tracking in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 667-686, Apr. 2009.
- [14] "Matlab Statistical Toolbox User's Guide," available at: <http://www.mathworks.com>, consulted in Jun. 2011.
- [15] J.L. Williams and P.S. Maybeck, "Cost-function-based Gaussian mixture reduction," in *Proc. 2003 6th International Conference on Information Fusion*, pp. 1047-1054.
- [16] A. Runnalls, "A Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989-999, Jul. 2007.
- [17] D.J. Salmond, "Mixture reduction algorithms for target tracking in clutter," *Proceedings of SPIE*, vol. 1305, pp. 434-445, 1990.
- [18] D. Schieferdecker and M.F. Huber, "Gaussian mixture reduction via clustering," in *Proc. 2009 12th International Conference on Information Fusion*, pp. 1536-1543.
- [19] U.D. Hanebeck, K. Briechle and A. Rauh, "Progressive Bayes: A new framework for nonlinear state estimation," *Proceedings of SPIE*, vol. 5099, pp. 256-267, May 2003.
- [20] M.F. Huber and U.D. Hanebeck, "Progressive Gaussian mixture reduction," in *Proc. 2008 11th International Conference on Information Fusion*, pp. 1-8.

## VII. BIOGRAPHIES

**Gustavo Valverde** (S'08, M'12) obtained the B.Sc. degree in electrical engineering from the University of Costa Rica, San Pedro, in 2005, the M.Sc. and Ph.D. degree in electrical power systems from the University of Manchester, Manchester, U.K., in 2008 and 2012. His research interests include monitoring of power systems, probabilistic analysis, control of distribution networks and voltage stability assessment.

**Jairo H. Quirós Tortós** (S'08) was born in Limón, Costa Rica. He obtained the B.Sc. and Licentiate degree with honours in Electrical Engineering from the University of Costa Rica, San Pedro, Costa Rica in 2008 and 2009 respectively. Currently he is a Ph.D. student at The University of Manchester, working on power system restoration and intelligent controlled islanding for power system. His main research interests are application of intelligent methods to power system restoration, and controlled islanding, power system monitoring, protection and control, voltage stability assessment, and power system dynamics. At the end of his Ph.D. studies, he hopes to return to the University of Costa Rica as a Professor and Researcher.

**Vladimir Terzija** (M'95, SM'00) is the EPSRC Chair Professor in Power System Engineering in the School of Electrical and Electronic Engineering, The University of Manchester, where he has been since 2006. From 1997 to 1999, he was an Assistant Professor at the University of Belgrade. In 1999, he was awarded a prestigious Humboldt Research Fellowship. From 2000 to 2006, he was with ABB AG, Germany, working as an expert for switchgear and distribution automation. His main research interests are application of intelligent methods to power system monitoring, control, and protection, switchgear and fast transient processes, as well as DSP applications in power systems.