Gaussian mixture based probabilistic load flow For LV-network planning

Gaussian mixture based probabilistic load flow for LV-network planning

Michiel Nijhuis, *Member, IEEE,* Madeleine Gibescu, *Member, IEEE,* and Sjef Cobben

***Abstract*—Due the many uncertainties present in the evolution of loads and distributed generation, the use of probabilistic load flow in low voltage (LV) networks is essential for the evaluation of the robustness of these networks from a planning perspective. The main challenge with the assessment of LV-networks is the sheer number of networks which need to be analysed. Moreover, most loads in the LV-network have a volatile nature and are hard to approximate using conventional probability distributions. This can be overcome by the use of a Gaussian mixture distribution in load modelling. Taking advantage of its radial nature and high R/X ratios, the LV-network can be analysed more efficiently from a computation viewpoint. By the application of simplifications defined in this paper, the backwards-forwards load flow can be solved analytically. This allows for the direct computation of the load flow equations with a Gaussian mixture distribution as load. When using this new approach, the required calculation time for small networks can be decreased to 3% of the time it takes to generate a similar accuracy with a Monte Carlo approach. The practical application of this load flow calculation method is illustrated with a case study on PV penetration.**

***Index Terms*—Probabilistic load flow, Power system analysis, Load flow analysis, Distribution network, Gaussian mixture distribution**

I. Introduction

THE upcoming introduction of new technologies on the demand side, like electric vehicles and heat pumps, and the proliferation of distributed generation requires the distribution network operator to evaluate whether its network is still sufficiently strong. The low voltage (LV) network especially will play an important role in the transition towards a more sustainable energy system, where electricity plays a prominent role. To assess whether these LV-networks are robust enough to handle the energy transition, load flow calculations have to be performed. The loads on the LV-network are mainly residential, and detailed information about the load profile is rarely available. Assumptions have to be made regarding the magnitude of the individual residential loads. For LV-networks planning horizons are typically 30-40 years and longer, therefore introducing a large load uncertainty [1]. To include these uncertainties about future residential loads a probabilistic load flow model should be employed. The sheer number of LV-networks and possible future scenarios in combination with an accurate assessment of the probability density function (PDF) for the residential loading would lead to infeasible computational times. Advanced probabilistic load flow methods are therefore necessary to determine the effects of the new load and generation technologies on the LV-network.

M. Nijhuis, M. Gibescu and J.F.G. Cobben are with the Electrical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: m.nijhuis@tue.nl

J.F.G. Cobben is also with Assetmanagement, Liander N.V., Arnhem, The Netherlands

There have been many different probabilistic optimal power/load flow formulations [2]–[5] that are able to accurately and relatively quickly assess the distribution network. These methods generally reduce the required amount of samples in a Monte Carlo (MC) simulation or are based on two-point [6] or multiple point estimates [7]. Gaussian mixture models have already been used to improve the probabilistic load flow computations as well. Application to medium voltage networks with the amount of variable loads limited to mostly the distributed renewable energy resources has been performed [8] [9]. For the DC load flow, a probabilistic load flow based on cumulants and Gram-Charlier expansion has already been applied [10]. For the LV network, the DC simplifications will not hold. The LV-network has in most cases a radial topology combined with a high R/X ratio, hence other simplifications to the load flow can be made to further improve the computational speed. However, for a rough initial network assessment, the trade-off between speed and accuracy is different. With a quick initial assessment, the networks which truly require attention can be identified and more in-depth analysis can be performed on a limited subset of LV-networks.

In this paper, simplifications for the backwards-forwards load flow calculation are presented. By applying these simplifications the load flow calculation is solved analytically employing a Gaussian mixture distribution directly into the load flow formulation. In this way, the probabilistic load flow for LV-networks can be solved in more quickly. First, a Gaussian mixture distribution is applied to model the residential load. Subsequently, based on this probability distribution, the load flow calculation for a radial LV-network is altered to allow for a quick assessment of the adequacy of the network. A conventional load flow would be too computationally intensive, therefore some simplifications are made as discussed in section III. The resulting Gaussian mixture based load flow calculation is shown in the next section. This is followed by a case study on the possible PV penetration levels in a sample network to show the advantages of the altered Gaussian mixture based probabilistic load flow.

II. Gaussian Mixture based load modelling

The first step in assessing the adequacy of an LV-network is to model the residential load. For an efficient estimation of the admissible loading situations in a network, the load should

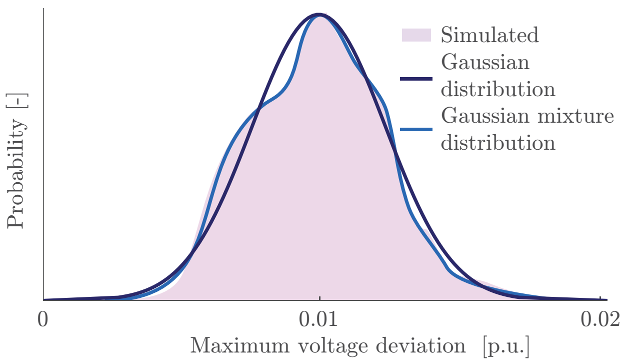

Fig. 1. Probability of the maximum voltage deviation occurring for different combinations of smart meter data

be modelled as a PDF. The loading of an individual household is, however, a stochastic process highly dependent on the switching of high-powered appliances. To accurately model this type of load, a unimodal PDF is not optimal as it cannot always capture the exact distribution of the load which is often multi-modal because of the switching behaviour of household appliances. To illustrate this, the minimum voltage in a simple feeder with 24 households has been calculated based on all possible combinations of smart meter measurements over the households. This results in a voltage distribution as shown in Fig. 1.

In the figure also a Gaussian distribution and a five component Gaussian mixture distribution are plotted for reasons of comparison. Based on the central limit theorem, the distribution of the voltage drops should approximate a Gaussian distribution, for large enough sample size. However, from the figure, it becomes apparent that for this 24-household feeder the voltage drop in the feeder is still significantly different from the Gaussian distribution, as multiple modes are clearly present. The modelling through a combination of Gaussian distributions would generate a more realistic voltage drop profile.

A Gaussian mixture distribution is given by the following formula:

$$f(x) = \sum_{k=1}^{K} \omega_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}} \qquad (1)$$
$$0 \leq \omega_k \leq 1 \text{ and } \sum \omega_k = 1$$

where K is the number of components, ω_k is the weight, σ_k the standard deviation and μ_k the mean of the k-th component. A Gaussian mixture consisting of five components can estimate the empirical distribution of the voltage deviation with greater accuracy than the normal distribution, as shown by Fig. 1. Therefore it is useful to develop a LV load flow algorithm based on Gaussian mixture distributions, as explained in the next section.

A. Load estimation through Gaussian mixture distribution

A Gaussian mixture distribution has already been applied for the modelling of the load. In [11] and [12] methods are shown to model the load through the use of the Expectation-Maximization (EM), either with the use of Akaikes Information Criterion (AIC) to determine the necessary number of components or with a predefined number. In this paper, the use of the approximation of the parameters of the components through the EM-algorithm is also chosen, however for the determination of the number of the components, the Bayesian information criterion (BIC) is applied [13]. This is done to give more emphasis on a lower number of components than prescribed by AIC. From a computational point of view this is favourable for the subsequent load flow calculation and to gain a better estimate of the required number of components with respect to non-machine choice.

The EM algorithm is applied, using a two-step iterative approach to find the maximum likelihood for the parameters of the Gaussian mixture model. Assuming a Gaussian mixture model with K components with the weights ω_k and component distributions $p_k(x_i|\theta_k)$, the likelihood is given by:

$$\mathcal{L}(\Theta|X) = \prod_{i=1}^{N} \sum_{k=1}^{K} \omega_k p_k(x_i|\theta_k) \qquad (2)$$

with N being the number of observations of the empirical data X and θ_k consist of the parameters for the normal distribution $\theta_k = \{\mu_k, \sigma_k\}$. The maximum likelihood estimate $\arg\max_\Theta \mathcal{L}(\Theta|X)$ cannot be determined analytically. The EM algorithm interprets the data X as incomplete and adds a binary vector Y. This vector indicates which observation of X belongs to which component in the Gaussian mixture. This vector Y is determined by using a K-means clustering approach. Resulting in the following likelihood:

$$\mathcal{L}(\Theta|X, Y) = \prod_{i=1}^{N} \sum_{k=1}^{K} y_{k,i} \omega_k p_k(x_i|\theta_k) \qquad (3)$$

The EM algorithm first step is to find the expected value of the complete likelihood given the parametrisation in the previous time step Θ^{p-1} through the Q-function:

$$Q(\Theta, \Theta^{p-1}) = E[\log p(\Theta|X, Y)|(Y, \Theta^{p-1})] \qquad (4)$$

The next step is the maximisation of the expectation of the previous step $\Theta^p = \arg\max_\Theta Q(\Theta, \Theta^{p-1})$. This procedure is iterated until the difference between two iterations is less than 10^{-6}. The EM algorithm needs to be started with an initial value for the first iteration. This initialisation is done with the use of a k-means clustering of the data. For the value of K (the number of clusters) the BIC is calculated through:

$$BIC = K \ln(N) - 2 \ln \arg\max_\Theta \mathcal{L}(\Theta|X) \qquad (5)$$

The BIC is calculated for increasing values of K till the value of $BIC(K) - BIC(K+1)$ becomes smaller than a predefined stopping criterion. This procedure is applied to smart meter measurements of 197 households for the peak hours (from 18:00-20:00) to generate the probability distribution of load values during the peak hours. The loss of accuracy for using of the BIC instead of the AIC criterion can be determined by the integral squared difference [14], which is $3 \cdot 10^{-6}$. The resulting three component Gaussian mixture distribution and the original data are plotted in Fig. 2.

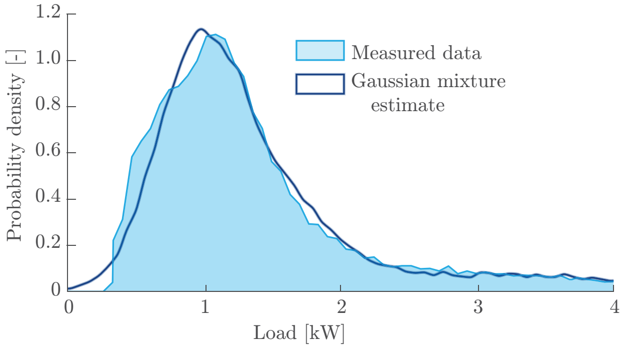

Fig. 2. Probability density function of smart meter data and the estimated Gaussian mixture distribution for the peak hours

TABLE I
CORRELATION AND P-VALUE OF THE ELECTRICITY CONSUMPTION FOR 120 HOUSEHOLDS WITHIN A SINGLE NEIGHBOURHOOD

Time	**Correlation**	**p-value**
00:00-02:00	0.085	0.27
12:00-14:00	0.028	0.36
18:00-20:00	0.012	0.37
00:00-24:00	0.258	0.01

TABLE II
OVERVIEW OF THE THREE FEEDER CHARACTERISTICS

	Bus [#]	**Load [#]**	**Load [kVA]**	**Voltage LN [V]**	**Phases [#]**	**R/X [-]**
US Residential	14	10	37	120	1	3.14
US Commercial	12	8	160	120	3	2.17
EU Residential	906	55	55	240	3	9.71

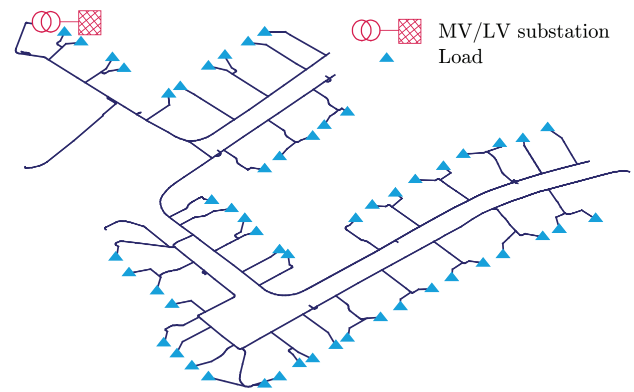

Fig. 3. The European low voltage test feeder

From the figure, it can be seen that for most parts the Gaussian mixture distribution and the empirical probability distribution are close together. One of the main differences is the existence of negative values in the Gaussian mixture distribution, as the Gaussian distribution ranges from $-\infty$ to ∞. Truncating the mixture distribution or individual components would result in a closer matched distribution. From the point of view of the final load flow outcome these negative values have limited effect due to their low probability and truncation is therefore not performed at this stage.

To determine how the household load can be modelled for network planning, the correlation between households within the same snapshot time is computed. This gives an indication of the level of dependence between the electricity consumption of two households. Smart meter data with a time resolution of 15-min from 120 households within the same neighbourhood are used for the analysis. As for network planning, the most important time periods are the hours of maximum and minimum load, and with the introduction of PV the household load at the middle of the day as well. In Table I the correlation coefficients and p-values are given for these 3 times of the day.

From the table, it can be seen that the correlation is low for all the evaluated time frames except when the correlation is calculated over the whole day. This indicates that the households follow a similar profile throughout the day, however, when looking at a shorter time frame this relation is no longer present. For the modelling of the household load through a Gaussian mixture distribution, this implies that the distributions of the different household loads can be considered independent for short time frames.

III. LOAD FLOW APPROXIMATIONS

For probabilistic load flow calculations in the LV-network, all the loads can be represented by PDF. As the load flow calculation needs to be solved in an iterative manner, the probabilistic approach is usually performed through MC simulations. Fortunately, LV-networks have characteristics which allow for some simplifications to generate a probabilistic load flow approach which no longer requires MC sampling. These assumptions are discussed in this section, preceded by a short overview of the characteristics of the feeders used to test the resulting load flow computation.

For the testing of the Gaussian mixture based load flow approach two of the Cigré US LV test feeders [15] (with the voltage level of the US feeder reduced to the more common 120V) and the IEEE European test feeder [16] are used. The main characteristics of these feeders are given in Table II. In the US test feeders the loads are combined at the buses to create a smaller network, therefore the US residential feeder consists of only 14 buses. For the European case the network modelling is much more detailed generating a 906 bus network. The US networks consist of a combination of overhead lines and underground cables, while the European network has only underground cable. For the EU test network the structure is also given in Fig. 3.

These LV-feeders will be used throughout to evaluate the assumptions made and to demonstrate the results of the proposed method. With the differences in the US and EU radial feeder characteristics, the suitability of the proposed method can be determined for different feeder types. The scalability can be assessed through the difference in the number of buses.

The first assumption is to assume balanced operation of the network, although the imbalance in LV-networks is generally higher than the one on higher voltage levels due to the large presence of single phase loads. Through the implementation of new technologies as PV or electric vehicles, the expectancy is that the imbalance will further rise. In most cases at LV levels, there is either a single- or a three-phase network available and the mitigation of imbalance problems in a three-phase

TABLE III
MAXIMUM ABSOLUTE VOLTAGE ERROR ASSUMED CONSTANT δ FOR DIFFERENT LOADING SCENARIOS LEADING TO A CERTAIN MINIMUM VOLTAGE

Minimum Voltage [p.u.]	**0.99**	**0.95**	**0.9**	**0.85**	**0.8**
US Residential [p.u.] $\cdot 10^{-3}$	0.13	0.65	1.33	1.86	2.39
US Commercial [p.u.] $\cdot 10^{-3}$	0.04	0.25	0.53	0.84	1.21
EU Residential [p.u.] $\cdot 10^{-3}$	0.10	0.53	1.08	1.66	2.28

network can often be done by switching a customer from one phase to another. As this will not affect the long-term network planning results, the issue of imbalance in the LV-networks will be neglected.

The second assumption is to assume that the voltage angle is constant. As most LV-networks have high R/X ratios the change in angle throughout the network is usually limited. For the older underground cable networks in Europe, the R/X ratio is usually between 8 and 10. For newer networks the ratio is around 3, equal to the value for networks of overhead lines used in other parts of the world. To assess the difference in load flow outcome when assuming the voltage angle is constant the difference in minimum voltage has been calculated for different loading scenarios. The maximum absolute difference is voltage is calculated and displayed in Table III.

In the table, the maximum absolute error in the minimum voltage is depicted for different loading levels. The first row in the table indicates the minimum voltage which is present in the network. For all the feeders the error increases as the voltage level starts to drop, which is logical as the voltage angle differences in the network increase with an increasing current. The US commercial network has the lowest impedance and thus the smallest error. For all three networks, the voltage magnitude estimated when neglecting the phasor angle will result in an answer which is 99.7% accurate, which is more than sufficient for an initial analysis of the network from a planning perspective.

Most LV networks are radially designed or at least radially operated. This allows for the use of a backwards forwards sweep load flow approach to the probabilistic load flow calculation [17]. In a backwards-forwards sweep the currents are summed moving from the end of the network to the MV/LV substation. Hereafter, the voltages starting from the slack bus towards the end of the feeders are calculated, followed by an adjustment of the load current based on the new voltages and losses. To make a conservative estimate (estimate which leads to the highest currents and largest voltage deviations) of loading of the network a constant power load model is assumed, which leads to the following load flow expression:

$$U(i) = U_0(i) - \sum_{n=i}^{N_b} \left(Z(n) \sum_{m=n}^{N_b} \left(\frac{S(m) + L(m)}{U(m)} \right) \right) \quad (6)$$

where $U(i)$ is the voltage at bus i, bus indexes i,n and m running from the busbar (index 0) to the end of the feeder (index N_b), $U_0(i)$ the base voltage, $Z(n)$ the impedance of the section between node n and $n-1$, $S(m)$ the apparent

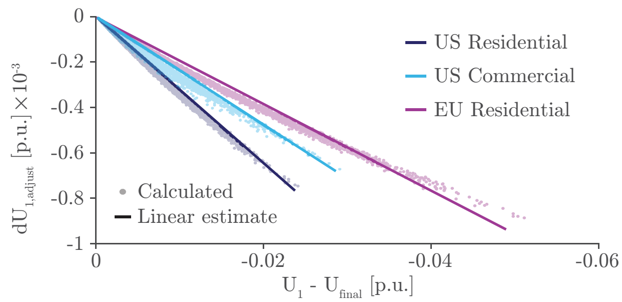

Fig. 4. The difference between the initial voltage drop and final voltage drop versus the voltage drop caused by the adjustment in losses and load current

power loading at bus m and $L(m)$ the loss ($Z \cdot I^2$) between bus m and bus $m+1$ as calculated by:

$$L(m) = (U(m) - U(m+1)) \sum_{k=m+1}^{N_b} \left(\frac{P(k)}{U(k)} \right) \quad (7)$$

Generally this iterative backward-forward procedure is repeated until the change in voltage between the iterations is lower than a certain threshold. To accelerate the computation of the load flow, the voltage drop difference between two iterations can be multiplied with a scalar value. If the appropriate scalar value is used the load flow can be computed with just a single iteration. To calculate this scalar value the relationship between the final voltage deviations and the voltage deviations of the losses and estimated adjusted load is defined in this paper. This relation is described as follows:

$$F_L(i) = \frac{U_1(i) - U_{final}(i)}{dU_1(i)^2 + \sum_{n=i}^{N_b} \left(Z(n) \sum_{m=n}^{N_b} \left(\frac{L(m)}{U_0(m)} \right) \right)} \quad (8)$$

where $dU_1(i)$ is the change in voltage between iterations 0 and 1 at node i. For the three different feeders this relation is plotted in Fig. 4. As can be seen from the figure this relation is almost linear. As the relationship defined in (8) shows a linear relation it can be used to generate a linear estimate of the load flow problem for LV-networks. The load flow computation can be adjusted to include this relation, leading to the following function:

$$U_{final}(i) = U_1(i) - F_L(i) \cdot (U_1(i) - U_0(i))^2 \\ -F_L(i) \sum_{n=i}^{N_b} \left(Z(n) \sum_{m=n}^{N_b} \left(\frac{L(m)}{U_0(m)} \right) \right) \quad (9)$$

with U_1 calculated by using flat start voltages and no losses. This generates a non-iterative procedure for the computation of the load flow. The main difficulty with this load flow computation is the lack of knowledge about the value of $F_L(i)$ and the error introduced by this approximation. To evaluate the applicability of this load flow computation simplification, this method is applied to the three test feeders. For the load flow computation, the loading of the network is assumed to consist half of a constant load and half of a uniformly distributed variable load. This in order to generate significant differences between the instances of the calculation of the load flow. The linear relation $F_L(i)$ is approximated based on a low loading and a high loading point estimate. Subsequently, with

TABLE IV
AVERAGE MAXIMUM ABSOLUTE VOLTAGE ERROR FOR USING F_L

Minimum Voltage [p.u.]	**0.99**	**0.95**	**0.9**	**0.85**	**0.8**
US Residential [p.u.] $\cdot 10^{-3}$	0.25	1.86	5.16	6.62	5.10
US Commercial [p.u.] $\cdot 10^{-3}$	0.06	1.21	2.77	2.72	9.74
EU Residential [p.u.] $\cdot 10^{-3}$	0.06	1.06	2.78	2.96	2.21

TABLE V
AVERAGE MAXIMUM ABSOLUTE VOLTAGE ERROR FOR COMBINED ASSUMPTIONS

Minimum Voltage [p.u.]	**0.99**	**0.95**	**0.9**	**0.85**	**0.8**
US Residential [p.u.] $\cdot 10^{-3}$	0.37	2.21	5.72	7.71	6.23
US Commercial [p.u.] $\cdot 10^{-3}$	0.22	1.93	4.10	3.43	7.46
EU Residential [p.u.] $\cdot 10^{-3}$	0.14	1.19	2.41	1.74	3.88

this value for $F_L(i)$ 10^4 load flow calculations have been performed to estimate the induced error.

In Table IV the resulting errors have been depicted. From the table, it becomes clear that the single iteration approximation induces a larger error than the previous assumption of the constant voltage angle. Note that the accuracy decreases to 98.8% for the commercial US feeder. The error also increases significantly as the minimum voltage in the system goes down. This is caused by the increasing influence of the part of the voltage reduction which is based on the factor F_L. The magnitude of the error is still within acceptable levels for an initial network adequacy assessment, especially if the voltage drop stays within the limit of $\pm 10\%$.

In the final load flow calculations, both the assumption of the single iteration and the assumption of the constant voltage angle will be used. Therefore the combined error of these two assumptions is shown in Table V. When comparing the table to Tables IV and III it can be seen that the combined error of the two assumptions is less than the addition of the two individual errors. Also for some cases the error value is actually lower than the maximum error encountered when looking at the assumptions individually. The total introduced error is therefore low enough to allow the proposed load flow computation to be used for the initial adequacy assessment of a LV-feeder from a planning perspective.

A. Convolution of Gaussian mixture distributions

Let $f(x)$ be a Gaussian distribution with mean μ_f and variance σ_f^2 and $g(y)$ a Gaussian distribution with mean μ_g and variance σ_g^2 and the correlation between the two distributions is given by ρ. The resulting PDF of the addition of two random variables from these distributions $\mathbb{N}_x + \mathbb{N}_y$ can be computed through the use of the characteristic function of the Gaussian distribution [18].

The convolution of two independent Gaussian distributions (F_x and F_y) can be computed through their combined characteristic function $\varphi_{x+y}(t) = \varphi_x(t)\varphi_y(t)$

$$\begin{aligned}\varphi_{x+y}(t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(\mu_x+\mu_y)-\frac{\left(\sigma_x^2+\sigma_y^2+2\sigma_{xy}\right)t^2}{2}}\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} e^{it\mu_z-\frac{\sigma_z^2t^2}{2}}\mathrm{d}z\end{aligned} \tag{10}$$

with $\mu_z = \mu_x + \mu_y$ and $\sigma_z = \sqrt{\sigma_x^2+\sigma_y^2+2\sigma_{xy}}$ and σ_{xy} is the covariance of x and y. For a mixture of independent (σ_{xy}=0) Gaussian distributions this can be expanded to:

$$f(x+y) = \sum_{k_x=1}^{K_x}\sum_{k_y=1}^{K_y}\frac{\omega_{k_x}\omega_{k_y}}{\sqrt{2\pi(\sigma_{k_x}^2+\sigma_{k_y}^2)}}e^{-\frac{\left(x+y-(\mu_{k_x}+\mu_{k_y})\right)^2}{2(\sigma_{k_x}^2+\sigma_{k_y}^2)}} \tag{11}$$

This will increase the number of components in the mixture to $n_x \times n_y$ components. The number of components increases rapidly if all of the loads are represented as Gaussian mixture distributions. For a simple network with only 15 loads which all are represented by Gaussian mixture distributions with three components, the number of resulting components from summing these loads leads to over 14 million (3^{15}) components. To ensure this method remains computationally feasible, a component reduction needs to be applied. The most common way to reduce the components in the new Gaussian mixture distribution is to apply the EM algorithm with a reduced number of components. When assessing a LV-network many of the loads will have a similar probability distribution, this leads to equal or near equal terms in the Gaussian mixture. These equal terms can be compounded into one, reducing the number of components based on the following rules:

$$\begin{aligned}\omega_{1,2} &= \omega_1 + \omega_2 \\ \mu_{1,2} &= \frac{\mu_1\omega_1 + \mu_2\omega_2}{\omega_1+\omega_2} \\ \sigma_{1,2}^2 &= \frac{\sigma_1^2\omega_1 + \sigma_2^2\omega_2 + \frac{(\mu_1-\mu_2)^2\omega_1\omega_2}{\omega_1+\omega_2}}{\omega_1+\omega_2}\end{aligned} \tag{12}$$

In the computation of the load flow, the multiplication of two dependent Gaussian mixture distributions is required. The computation of this multiplication can be done with the assumption that the standard deviation of the components is much lower than the mean. For the random variables x and y their multiplication can be defined as:

$$\begin{aligned}f_{xy} &= (\mu_x + \mathcal{N}[0,\sigma_x])(\mu_y + \mathcal{N}[0,\sigma_y]) \\ &= \mu_y\mu_x + \mu_y\mathcal{N}[0,\sigma_x] + \mu_x\mathcal{N}[0,\sigma_y] + \mathcal{N}[0,\sigma_x]\mathcal{N}[0,\sigma_y])\end{aligned} \tag{13}$$

Since in our case, the mean is much higher than the standard deviation the term $\mathcal{N}[0,\sigma_y]\mathcal{N}[0,\sigma_y]$ can be neglected, yielding the following Gaussian distribution:

$$\mathcal{N}\left[\mu_y\mu_x, \sqrt{(\mu_y\sigma_x)^2 + (\mu_x\sigma_y)^2 + \mu_y\mu_x\sigma_x\sigma_y}\right] \tag{14}$$

B. Gaussian mixture based load flow algorithm

As an illustration, consider a radial network with loads at each node except for the slack bus at node 1. The voltage can also be calculated by matrix multiplication through the use of the adjacency matrix extended with the shortest path between the node and the slack bus. For the network the extended adjacency matrix is given by:

$$\mathbf{A_j} = \begin{bmatrix}1 & 1 & \dots & 1\\ 0 & 1 & \dots & 1\\ \vdots & & & \vdots\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{15}$$

Based on the extended adjacency matrix, the extended ZBus matrix C_j can be defined:

$$\mathbf{C_j} = ((\mathbf{A_j}^T(\mathbf{1}^T Z) \circ \mathbf{I})\mathbf{A_j}) \tag{16}$$

The initial iteration of the load flow computation becomes:

$$U_1 = \frac{S\mathbf{C_j}}{U_0^2} \tag{17}$$

where S is the vector of loads, which are represented by the σ, μ and ω of their respective Gaussian mixture distribution. The resulting voltage deviation U_1 can be squared to obtain the approximate voltage deviation through the change in load current as the in the first iteration the voltage is equal to 1[p.u.] so the additional load current can be estimated by the voltage deviation for a constant power load. The additional change in voltage caused by the losses can be calculated by using the following formula:

$$dV_L = \frac{Z\left(\frac{\mathbf{A_j}S}{U_0}\right)^2\mathbf{C_j}}{U_0^2} \tag{18}$$

The losses and the estimate voltage deviation from the change in load current need to be multiplied by the factor F_L as defined in equation (8). F_L can be estimated by performing two ordinary load flow computations, one with a relatively low load and one with a high load. The voltage deviations calculated over each branch need to be added to obtain the total voltage deviation. The branch currents can sequentially be calculated from the voltage deviations. For most networks the use of a component reduction step, as described in equation (12), is needed after each of the steps to keep the memory usage in check.

IV. Results

The accuracy of the Gaussian mixture based load flow is evaluated with respect to the conventional approach of using a Latin hypercube sampling based MC simulation [19] in combination with load flow calculations using a full backwards-forwards sweep. The load flows are calculated by using scaled versions of the distribution as shown in Fig. 2 for each load. In the Fig. 5 the resulting PDF of the minimum voltage for the US residential feeder is plotted for the Gaussian mixture approach as well as for the MC approach with increasing sample sizes.

From the figure, it becomes clear that the PDF for the US residential case presents many modes. This can be explained by the low number of buses in the system, too low for the central limit theorem to apply. The MC sampling results with only 10^4 and to a lesser extent with 10^5 samples, show erratic behaviour. The MC method with 10^6 samples and the Gaussian mixture approach all are close together. A more detailed analysis is required to judge the applicability of the Gaussian mixture-based load flow in comparison with the established MC sampling.

The result of a probabilistic load flow is a histogram of the possible outcomes. To assess to what degree two histograms are derived from the same underlying distribution, a number of statistical tools can be used. The three tests used will generate

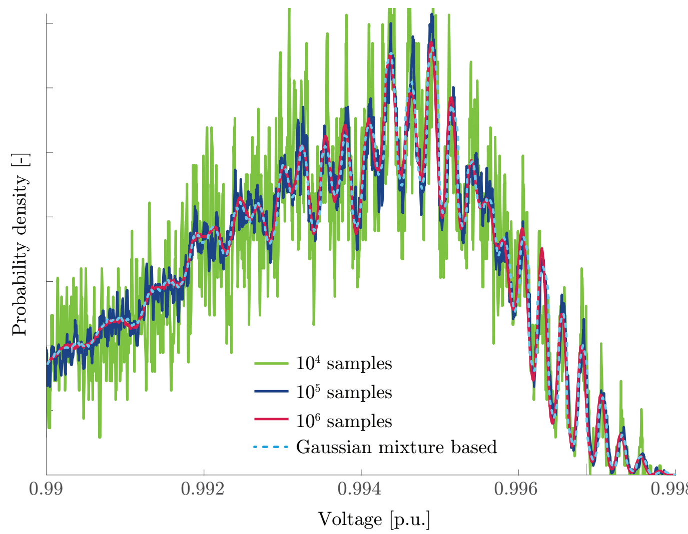

Fig. 5. Probability density function of the minimum voltage for the US residential feeder for the MC method and the proposed Gaussian mixture based load flow

an outcome of one if the histograms are exactly the same, while if there are discrepancies the result will be lower. To give a combined overview of the fit of histograms H and G the following compounded measure $D_{tot}(G,H)$ is defined for this paper, which is the product of three individual distance measures:

$$D_{tot}(G,H) = D_W(G,H) \cdot D_B(G,H) \cdot D_{KL}(G,H) \tag{19}$$

where $D_W(G,H)$ is the Wasserstein metric or earth movers distance [20], defined as follows:

$$D_W(G,H) = \frac{\min_{f_{ij}} \sum_{i,j} |f_{ij}| \cdot d_{ij}}{\sum_{i,j} f_{ij}} \tag{20}$$

where f_{ij} is a function for which the following holds: $G_i + f_{ij} = H_i$ and $G_j - f_{ij} = H_j$ the and d_{ij} the distance between i and j. $D_W(G,H)$ is the Bhattacharyya distance [21], defined as follows:

$$D_B(G,H) = -\ln \sum_i^n \sqrt{G_i \cdot H_i} \tag{21}$$

where G_i is the bin count of bin i in histogram G and H_i is the bin count of bin i in histogram H. $D_{KL}(G,H)$ is the Kullback-Leiber divergence [20] as defined by:

$$D_{KL}(G,H) = \sum_i^n G_i \ln \frac{G_i}{H_i} \tag{22}$$

By combining these three metrics both the absolute and relative error in bin count as well as the total variation distance will be assessed. The results for the fit of the histogram with respect to a distribution generated with a MC simulation with 10^7 samples and the required computational times (using an Intel i7-3770 3.4 GHz processor, 8GB ram computer running Matlab 2015a) are given in Table VI.

From the table, it can be seen that the Gaussian mixture based method performs comparable to an MC simulation with 10^7 samples. The computational time is however significantly

TABLE VI
EVALUATION OF THE MC AND GAUSSIAN MIXTURE-BASED (GM) LOAD FLOW METHODS FOR THREE FEEDERS BASED ON THE CALCULATION TIME AND THE DISCREPANCIES WITH A PROBABILITY DISTRIBUTION GENERATED WITH MC SIMULATION WITH 10^7 SAMPLES

US Residential [p.u.]	**MC-10^4**	**MC-10^5**	**MC-$5 \cdot 10^5$**	**GM**
D_{tot} [-]	0.869	0.961	0.994	0.996
Calculation time [s]	6.17	49.3	221	4.58

US Commercial	**MC-10^4**	**MC-10^5**	**MC-$5 \cdot 10^5$**	**GM**
D_{tot} [-]	0.879	0.965	0.995	0.992
Calculation time [s]	6.68	49.0	217	4.45

EU Residential	**MC-10^4**	**MC-10^5**	**MC-$5 \cdot 10^5$**	**GM**
D_{tot} [-]	0.868	0.969	0.995	0.996
Calculation time [s]	7.98	64.8	307	51.5

lower with only 2.1% of the computational time needed for the US test feeders, while the Gaussian mixture approach required 16.8% of the time for the EU case. The difference in calculation time between US and EU test feeders indicates the scalability of the Gaussian mixture based is slightly lower, due to the need for additional component reduction steps. The computational time to obtain a similar level of accuracy is however still much lower compared to a conventional MC approach.

V. CASE STUDY

One of the applications of the probabilistic load flow is to assess the effects of PV generation on the distribution network [22]. The use of a Gaussian mixture based load flow has additional benefits compared to a normal probabilistic load flow. As the loads are modelled through the use of components within a Gaussian mixture distribution the increase or decrease in the probability of certain components can relate to an increase or decrease in penetration level of a certain technology. This is illustrated with a case study wherein the European LV test feeder is used with an increasing amount of PV generation. The PV generation is modelled based on a model for the creation of PV time series [23]. The generated PV time series for the times between 12:00 and 14:00 is added to the distribution of the household load between these two times to create the Gaussian mixture distribution of the load as illustrated in Fig. 6.

The European LV test feeder which is used in the analysis is a reasonably strong feeder, under normal load conditions the absolute voltage deviation is only 0.02. For the sake of the analysis, the loads on the feeder have been doubled to ensure some voltage violations will arise during the analysis. The maximum current of the cables within the network is based on their resistance with the main cable type 4c_70 having a maximum current rating of 160A per phase. The analysis is performed for the installation of rooftop PV at 0-100% of the households. By using the proposed approach the whole calculation for each percentile value took 272s. The resulting voltage at the end of the feeder is used as the main characteristic during the analysis as the expected voltage deviations are highest at that point and no data is available on the maximum loading of the different cable parts. In Fig. 7 the resulting voltage PDFs are plotted.

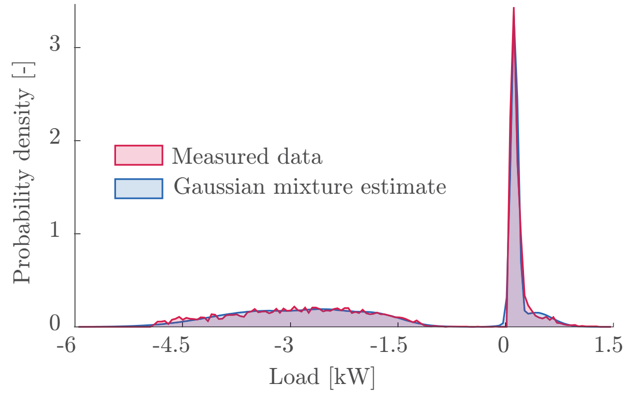

Fig. 6. Gaussian mixture distribution of residential load at midday with a PV penetration rate of 50%

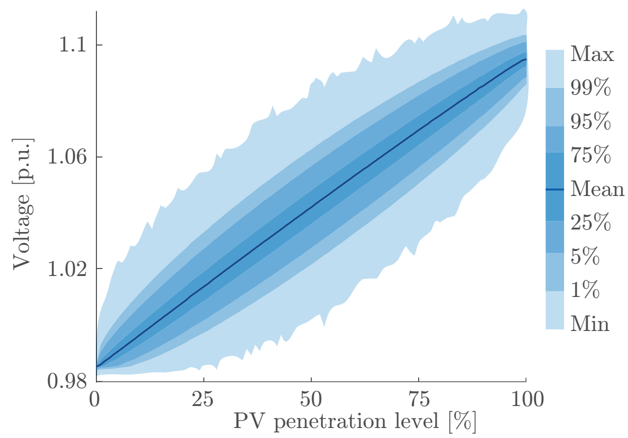

Fig. 7. Voltage distribution at the end of the feeder for different levels of PV penetration with a darker colour indicating a higher probability

In the figure, a higher saturation indicates a higher probability of a certain voltage value occurring at the end of the feeder. From the figure, it can be seen that at a penetration level of 0% or 100% the PDF is narrow while around the 50% penetration rate the uncertainty about the actual voltage deviation is much larger.

For the planning of the reinforcements of the LV-network, it is important to know what the risk of overloading or the occurrence of voltage violations in the LV-network is. In order to gain insight into this statistic the chance overloading and of violating the voltage limits of the cables is plotted as a function of the PV penetration level for two different limits in Fig. 8.

In the figure the probability of exceeding the chosen maximum allowable current (0.7 or 1 [p.u.]) or the maximum allowable voltage deviation (± 0.02 or ± 0.05 [p.u.]) is shown. From the figure, it can be seen that all of these functions follow an S-curve. In the network the overloading of the cables will occur before the voltage limits are violated, only a very strict voltage limit of ± 0.02 would lead to an overvoltage before overloading problems start. The S-curves of these limits vary, with the ± 0.02 rising the fastest and ± 0.05 having the smallest slope for the voltage. The S-curve shows that the chance of an overloading or overvoltage can rise quickly with a relatively

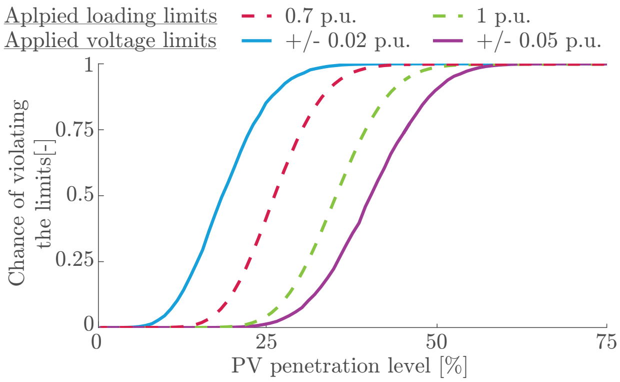

Fig. 8. The chance of violating different voltage and loading limits for different PV penetration levels

small increase in PV penetration level. Within these ranges of PV penetration levels, additional care must be taken when planning network reinforcements as a small additional amount of PV can introduce a violation of the loading or voltage limits.

VI. Conclusions

By the application of simplifications to the backwards-forwards load flow applicable to most (radial) LV-networks it becomes possible to analytically solve the load flow equations with a Gaussian mixture distribution for the load as input. By utilising this approach the required calculation time for small networks can be decreased to 2.1% with respect to the time it takes to generate a similar accuracy with a Latin hypercube sampling MC approach.

The backwards forwards sweep for LV-networks has been simplified. The voltage angle can be neglected as the R/X ratios in LV-networks tends to be between 2 and 10. In addition, the acquired voltage difference from the initial iteration is shown to be almost linearly related to the difference between the initial and final iteration voltage deviation. The application of these two simplifications still allows for a load flow computation accuracy of 99.0%. These simplifications allow for to use the PDF of a Gaussian mixture distribution directly as an input for the load flow computation. Based on these computations the distribution of the voltage at each bus and power at each branch can be analytically determined. This allows for much faster computations of probabilistic load flow. The use of a Gaussian mixture distribution as input allows for a large variation in load distributions. To illustrate how this method could be applied, the case of assessing which level of PV-penetration a test feeder can handle has been performed. Within 5 minutes for each percentage point of PV penetration the associated risk has been determined, showing that the method can analyse the risks due to load changes in LV-networks accurately and efficiently.

References

[1] H. L. Willis, *Power distribution planning reference book*. CRC Press, Mar 2004, vol. 2.

[2] T. Williams and C. Crawford, "Probabilistic load flow modeling comparing maximum entropy and gram-charlier probability density function reconstructions," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 272–280, Feb 2013.

[3] P. Siano and G. Mokryani, "Probabilistic assessment of the impact of wind energy integration into distribution networks," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4209–4217, Nov 2013.

[4] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Probabilistic load flow for photovoltaic distributed generation using the cornish–fisher expansion," *Electric Power Systems Research*, vol. 89, pp. 129–138, Apr 2012.

[5] J. Liang, D. D. Molina, G. K. Venayagamoorthy, and R. G. Harley, "Two-level dynamic stochastic optimal power flow control for power systems with intermittent renewable generation," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2670–2678, Jul 2013.

[6] G. Verbic and C. A. Canizares, "Probabilistic optimal power flow in electricity markets based on a two-point estimate method," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1883–1893, Nov 2006.

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

[8] G. Valverde, A. T. Saric, and V. Terzija, "Probabilistic load flow with non-gaussian correlated random variables using gaussian mixture models," *IET Generation, Transmission & Distribution*, vol. 6, no. 7, p. 701, Jan 2012.

[9] F. J. Ruiz-Rodriguez, J. C. Hernandez, and F. Jurado, "Probabilistic load flow for radial distribution networks with photovoltaic generators," *IET Renewable Power Generation*, vol. 6, no. 2, p. 110, Mar 2012.

[10] D. Villanueva, A. E. Feijóo, and J. L. Pazos, "An analytical method to solve the probabilistic load flow considering load demand correlation using the dc load flow," *Electric Power Systems Research*, vol. 110, pp. 1–8, Jan 2014.

[11] C. Carmona-Delgado, E. Romero-Ramos, and J. Riquelme-Santos, "Probabilistic load flow with versatile non-gaussian power injections," *Electric Power Systems Research*, vol. 119, pp. 266–277, Oct 2014.

[12] 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.

[13] G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, Inc., Jan 2000.

[14] J. L. Williams and P. S. Maybeck, "Cost-function-based gaussian mixture reduction for target tracking," in *Proceedings of the sixth international conference of Information fusion*, vol. 2, 2003, pp. 1047–1054.

[15] K. Strunz, R. H. Fletcher, R. Campbell, and F. Gao, "Developing benchmark models for low-voltage distribution feeders," in *2009 IEEE Power & Energy Society General Meeting*, Jul 2009, pp. 1–3.

[16] I. PES. (11, Jan) Distribution test feeders. [Online]. Available: http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html

[17] E. Janecek and D. Georgiev, "Probabilistic extension of the backward/forward load flow analysis method," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 695–704, May 2012.

[18] O. Knill, "Probability and stochastic processes with applications," pp. 7–14,116–122, Jun 2011. [Online]. Available: www.math.harvard.edu/~knill/teaching/math144_1994/probability.pdf

[19] H. Yu, C. Y. Chung, K. P. Wong, H. W. Lee, and J. H. Zhang, "Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 661–667, May 2009.

[20] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International Statistical Review*, vol. 70, pp. 419–435, Dec 2002. [Online]. Available: http://adsabs.harvard.edu/abs/2002ISRv...70..419G

[21] A. Bhattacharyya, "On a measure of divergence between two multinomial populations," *Sankhya: The Indian Journal of Statistics*, vol. 7, no. 4, pp. 401–406, Jul 1946.

[22] S. Conti and S. Raiti, "Probabilistic load flow using monte carlo techniques for distribution networks with photovoltaic generators," *Solar Energy*, vol. 81, pp. 1473–1481, Dec 2007. [Online]. Available: http://adsabs.harvard.edu/abs/2007SoEn...81.1473C

[23] Z. Ren, W. Yan, X. Zhao, W. Li, and J. Yu, "Chronological probability model of photovoltaic generation," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1077–1088, May 2014.