Discrete-Continuous Gaussian Mixture Models for Wind Power Generation

Tianyang Yi*, Shibshankar Dey†, D. Adrian Maldonado‡, Sanjay Mehrotra†, Anirudh Subramanyam*
*Pennsylvania State University, University Park, PA, USA
†Northwestern University, Evanston, IL, USA
‡Argonne National Laboratory, Lemont, IL, USA
Email: subramanyam@psu.edu

2024 IEEE Power & Energy Society General Meeting (PESGM) | 979-8-3503-8183-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/PESGM51994.2024.10761067

***Abstract*—Gaussian Mixture Models (GMM) are an effective representation of resource uncertainty in power systems planning, as they can be tractably incorporated within stochastic optimization models. However, the skewness, multimodality, and bounded physical support of long-term wind power forecasts can entail requiring a large number of mixture components to achieve a good fit, leading to complex optimization problems. We propose a probabilistic model for wind generation uncertainty to address this challenge, termed Discrete-Gaussian Mixture Model (DGMM), that combines continuous Gaussian components with discrete masses. The model generalizes classical GMMs that have been widely used to estimate wind power outputs. We employ a modified Expectation-Maximization algorithm (called FixedEM) to estimate the parameters of the DGMM. We provide empirical results on the ACTIVSg2000 synthetic wind generation dataset, where we demonstrate that the fitted DGMM is capable of capturing the high frequencies of time windows when wind generating units are either producing at maximum capacity or not producing any power at all. Furthermore, we find that the Bayesian Information Criterion of the DGMM is significantly lower compared to that of existing GMMs using the same number of Gaussian components. This improvement is particularly advantageous when the allowed number of Gaussian components is limited, facilitating the efficient solution to optimization problems for long-term planning.**

I. Introduction

Wind and solar generation resources will be essential to the decarbonization of the electrical power grid. Due to their variability and dependence on weather conditions, these resources are subject to greater prediction uncertainty than conventional generating units. In particular, the integration of wind generation into long-term expansion planning presents challenges due to the variability and unpredictability of weather, and the resulting unavailability of accurate month-ahead or year-ahead forecasts.

Owing to this increase in variable energy resources, the power system community has developed efficient optimization algorithms that account for planning uncertainties in an accurate and scalable fashion [1]. These algorithms trade off the complexity of the represented distributions with the tractability of the overall optimization model. In one extreme of this spectrum, models typically assume small-magnitude Gaussianity to offer computational efficiency albeit at the expense of losing representation flexibility [2], [3].

This material is based upon work supported by the National Science Foundation under Grant No. DMS-2229408.

In the long term, however, the distribution of wind power generation can display multimodality and skewness, reflecting the combined nonlinear effects of wind speed and turbine physics. The need to capture these distributional characteristics while also curbing the complexity of the optimization has led to the analytical reformulation of stochastic planning problems based on Gaussian Mixture Models (GMMs) [4]–[6]. GMMs are universal approximators capable of representing complex non-Gaussian distributions with a sufficiently large number of Gaussian components, and thus provide a viable pathway toward incorporating complex uncertainty sources in planning.

It is therefore important to develop efficient model fitting methods that can build accurate GMM representations using historical wind generation data. Recently, a new Expectation-Maximization (EM) algorithm was proposed to fit GMMs to wind data [7]. However, while the resulting GMMs offer additional expressivity when compared with Gaussians, they consist of a large number (between 20–100) of Gaussian components. Such large component numbers not only increase model complexity but also translate to computational inefficiencies in the solution of the optimization models in which they are used.

Our paper addresses these issues with the following contributions:

- We propose a novel multivariate Discrete-Gaussian Mixture Model (DGMM), where some of the mixture components are allowed to be discrete masses.
- We propose a modified Expectation-Maximization algorithm (called FixedEM) to fit the DGMM to given wind power data.
- We show empirically that the estimated DGMM is able to better fit wind power data with a fewer number of Gaussian components when compared with the classical GMM, potentially leading the path to more efficient stochastic planning algorithms.

In Section II, we introduce the DGMM and the FixedEM algorithm to estimate its parameters. In Section III, we present a series of experiments and empirical results with time series data of wind power generation. We offer concluding remarks in Section IV.

II. Proposed Mixture Distribution Model

A mixture model is a convex combination of two or more probability distributions. It has the flexibility of approximating a wide variety of distribution shapes, thus applicable in modeling complex data distributions, where a single standard distribution may fail to capture key characteristics of the data. Each probability distribution in a mixture model is called a mixture component.

In this work, we propose a Discrete-Gaussian Mixture Model (DGMM), where some of the mixture components are continuous multivariate Gaussian distributions and all remaining components are discrete masses. Specifically, we propose distributions that have K Gaussian components and J discrete masses. When the random vector is d-dimensional, these distributions have the following density function:

$$f(x|w,\mu,\Sigma) = \sum_{k=1}^{K} w_k \mathcal{N}(x|\mu_k,\Sigma_k) + \sum_{j=K+1}^{K+J} w_j \delta(x-\mu_j), \tag{1}$$

where $\mathcal{N}(\cdot|\mu,\Sigma)$ denotes the density function of a d-variate normal random vector with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, $\delta : \mathbb{R}^d \mapsto \mathbb{R} \cup \{\infty\}$ denotes the Dirac delta function, and $w \in \mathbb{R}^K$ is the weight vector satisfying:

$$\sum_{k=1}^{J+K} w_k = 1, \quad w_k > 0, \quad \forall k = 1,2,\ldots,J+K. \tag{2}$$

In addition to the K continuous Gaussian components, the proposed model (1) allows the random vector to realize any of J discrete values $(\mu_{K+1},\mu_{K+2},\ldots,\mu_{K+J})$ with strictly positive probability. Sampling from the proposed model is thus relatively straightforward. Also, note that the proposed DGMM is a generalization of the classical GMM and reduces to the latter when the number of discrete masses, $J=0$.

From a statistical viewpoint, the discrete masses provide the flexibility of capturing sharp peaks, such as those that occur at the minimum and maximum production levels (0 and 1 per unit, respectively), in the histogram of wind power outputs (e.g., see Figs. 1, 2 in Section III). Perhaps surprisingly, we find that this flexibility also comes with the benefit of having to use a much smaller number K of continuous Gaussian components to achieve a similar level of statistical fit, when compared with the classical GMM.

This has two important consequences. On the one hand, a fewer number of model parameters have to be estimated, thus reducing overall model complexity. On the other hand, the reduction in K allows for the efficient solution of optimization problems for long-term planning, where statistical models of wind generation are used to enforce chance constraints for reliability purposes [8], [9]. Indeed, whereas each Gaussian component entails the introduction of a nonlinear constraint (see [4]–[6], [10] for details), each discrete mass only requires the introduction of a single binary variable and a linear constraint [11], [12]. Therefore, it is important to reduce the number K of Gaussian components as much as possible without sacrificing statistical performance.

A. Parameter Estimation using Expectation-Maximization

Accurately estimating the model parameters (w,μ,Σ) from observed data is crucial. Maximum Likelihood Estimation (MLE) is a well-known method used to estimate the parameters of any probability distribution. It aims to obtain the best possible parameters that maximize the likelihood function. In practice, the Expectation-Maximization (EM) algorithm is an iterative method to solve the MLE problem [13], [14]. The EM Algorithm alternates between an expectation (E-step) and maximization (M-step). At each iteration, the E-step first computes the expected value of the log-likelihood function using the current estimated parameters. The M-step then updates the parameter estimates by maximizing the function computed in the E-step.

We propose a modified version of the EM algorithm to fit a DGMM to a given dataset. The algorithm, which we call FixedEM, attempts to find the locations of the J discrete masses $(\mu_{K+1},\mu_{K+2},\ldots,\mu_{K+J})$ by first fixing J covariance matrices of a $(K+J)$-component GMM to $\epsilon^2 I$, where I denotes the $d\times d$ identity matrix and $\epsilon \in (0,1)$ is a small positive constant. The FixedEM algorithm is otherwise identical to the EM algorithm except that the J fixed covariances are not updated during the M-step. The intuition is that a normal distribution with a very small (but non-zero) covariance approximates a discrete mass at the mean of the distribution.

The FixedEM procedure is shown in Algorithm 1. Here, $X = \{x_1,x_2,\ldots,x_n\}$ denotes the dataset consisting of n observations, $L^{\max}$ denotes the maximum number of iterations, and tol denotes the convergence tolerance. In the algorithm, $Z \in \mathbb{R}^{n\times(K+J)}$ is the matrix of latent data indicating the membership of each observation to each of the components. It can be shown [13], [14] that at every iteration, the estimated parameters provide an increase in the value of the likelihood function until a local maximum is attained. In practice, we run the algorithm several times, each time starting from a different initialization and choose those final parameter values that attain the largest likelihood.

B. Optimal Model Selection

A key challenge when fitting any mixture model is to determine the appropriate number of mixture components. Introducing a large number of components can better capture the shape of the data distribution, but this can also increase model complexity and result in overfitting. On the other hand, a very small number of components can fail to capture all characteristics of the data. In this work, we use the well-known Bayesian Information Criterion (BIC) [15], [16] to quantify this tradeoff. The BIC score is computed as

$$\mathrm{BIC} = P\ln(n) - 2\ln(\hat{Q}), \tag{3}$$

where n is the number of observations used to estimated the model, $\hat{Q}$ is the likelihood computed using the parameters estimated from the FixedEM algorithm, and P is the number of estimated model parameters. Observe that increasing the number of mixture components (and hence, the number of parameters) to achieve a better likelihood is appropriately

Algorithm 1 Fixed Expectation-Maximization for DGMM

procedure FIXEDEM(X, K, J, $L^{\max}$, ϵ, tol)
  Initialize $\mu^{(0)}, w^{(0)}, \Sigma_1^{(0)}, \Sigma_2^{(0)}, \ldots, \Sigma_K^{(0)}$ randomly.
  $\Sigma_j^{(0)} \leftarrow \epsilon^2 I$ for $j = K+1, K+2, \ldots, K+J$.
  for $l = 1, 2, \ldots, L^{\max}$ **do**
    $Z^{(l)} \leftarrow \text{EXPECTATION}(X, w^{(l-1)}, \mu^{(l-1)}, \Sigma^{(l-1)})$
    $w^{(l)}, \mu^{(l)}, \Sigma^{(l)} \leftarrow \text{MAXIMIZATION}(X, Z^{(l)})$
    $\Delta \leftarrow \|(w^{(l)}, \mu^{(l)}, \Sigma^{(l)}) - (w^{(l-1)}, \mu^{(l-1)}, \Sigma^{(l-1)})\|$
    if $\Delta <$ tol **then break end if**
  end for
  return $w^{(i)}, \mu^{(i)}, \Sigma^{(i)}$
end procedure
function EXPECTATION(X, w, μ, Σ)
  for all $x_i \in X$ **do**
    for $k = 1, 2, \ldots, K+J$ **do**
      $$Z_{ik} \leftarrow \frac{w_k \cdot \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K+J} w_j \cdot \mathcal{N}(x_i|\mu_j, \Sigma_j)}$$
    end for
  end for
  return Z
end function
function MAXIMIZATION(X, Z)
  for $k = 1, 2, \ldots, K+J$ **do**
    $N_k \leftarrow \sum_{i=1}^{n} Z_{ik}$
    $w_k \leftarrow \frac{N_k}{\sum_{j=1}^{n} N_j}$
    $\mu_k \leftarrow \frac{1}{N_k} \sum_{i=1}^{n} Z_{ik} \cdot x_i$
    if $k \le K$ **then**
      $\Sigma_k \leftarrow \frac{1}{N_k} \sum_{i=1}^{n} Z_{ik} \cdot (x_i - \mu_k)(x_i - \mu_k)^\top$
    else
      $\Sigma_k \leftarrow \epsilon^2 I$
    end if
  end for
  return w, μ, Σ
end function

penalized by the BIC score. Therefore, models with lower BIC scores are preferred when deciding between several candidate models. In particular, for the DGMM (1), it can be verified that $P = O(Kd^2 + Jd)$. This means that whereas adding one Gaussian component introduces $O(d^2)$ new parameters (since each element of Σ_k counts as one additional parameter), adding one discrete component only introduces $O(d)$ new parameters, resulting in a much simpler model.

III. Results and Performance Evaluations

A. Dataset

We use the ACTIVSg2000 dataset [17], a synthetic test case that simulates a 2000-bus system in Texas. This case uses historical weather measurements that are then mapped to renewable generation via wind power curves and solar PV generation models [18], and provides a realistic depiction of wind power generation statistics. The dataset contains time series data of bus-level power generation for 87 wind generating units. The hourly wind power generation for each of the 87 wind generating units is provided over a span of 366 days in 2016. We normalized the power generation at each bus by its maximum observed value, and thus all our data points are within the bounded interval $[0, 1]$.

B. Experimental Setup

This section discusses the performance assessment and metrics of our proposed FixedEM algorithm. The hyperparameters used in the experiments are:

- The number of Gaussian components, $K \in \{2, 4, 6, 8, 12, 16, 20\}$.
- The number of discrete components, $J \in \{0, 1, 2, 3, 4\}$. In our implementation of Algorithm 1, we set $\epsilon = 0.01$. Recall that ϵ is used to approximate the J discrete masses using Gaussian components with covariance $\epsilon^2 I$.

Two sets of experiments are conducted:

- Experiment 1 (univariate analysis): We implemented Algorithm 1 on the first 10 wind generating units from the ACTIVSg2000 dataset. For each unit, 100 different initializations are performed for each possible (K, J) pair. For each K, when $J = 0$ (i.e., fitting a classical GMM), we recorded the lowest BIC score of 100 runs using FixedEM and denoted it as BIC*(EM-GMM); we also implemented the Density-Preserving Hierarchical EM (DPHEM) algorithm from [7] (the bandwidth of the initial Kernel Density Estimation model was set to 0.01), recorded its lowest BIC score from 100 runs and denoted it as BIC*(DPHEM-GMM). We define BIC*(GMM) to be the smaller of the two BIC scores. When $J \in \{1, 2, 3, 4\}$ (i.e., fitting a DGMM), we recorded the lowest BIC score of 100 runs for each J, took the minimum over all four possible J values, and denoted it as BIC*(DGMM).
- Experiment 2 (multivariate analysis): This experiment extends the analysis to a multivariate context, aiming to model the spatial correlation in wind power generation across four units. Using the correlation matrix of wind power generation of all 87 generating units, we select four correlated and four uncorrelated units. In each of these two cases, we treat the wind power generation at a specific hour of the year as a single 4-dimensional data point. We then follow the same procedure as in Experiment 1 to obtain the corresponding (EM-based) BIC*(GMM) and BIC*(DGMM).

To decide whether GMM or DGMM is better, we compute the relative improvement in BIC scores, Δ_{BIC}, defined as follows:

$$\Delta_{\text{BIC}} = \frac{\text{BIC}^*(\text{GMM}) - \text{BIC}^*(\text{DGMM})}{|\text{BIC}^*(\text{GMM})|} \times 100\%, \quad (4)$$

A positive value of Δ_{BIC} indicates that the proposed DGMM has a lower BIC score and would be preferred over the GMM, while a negative value of Δ_{BIC} suggests that the GMM is the preferred model.

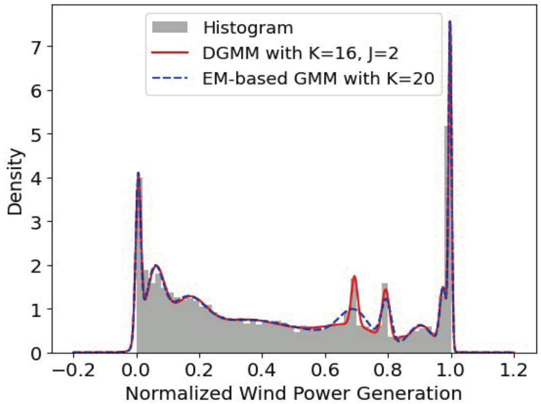

Fig. 1: Histogram and Estimated Densities of Unit 1

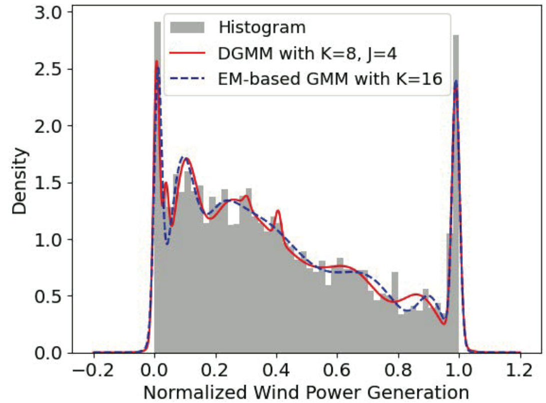

Fig. 2: Histogram and Estimated Densities of Unit 3

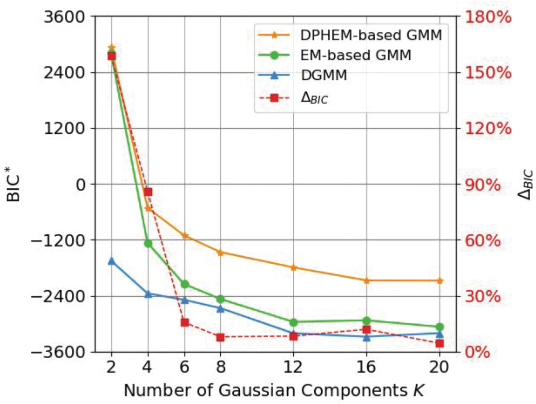

Fig. 3: BIC Scores of Estimated Models of Unit 1

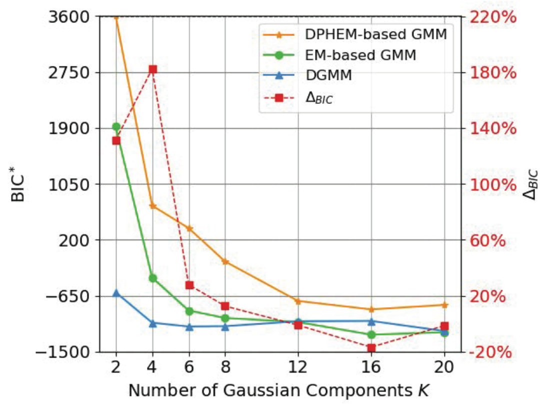

Fig. 4: BIC Scores of Estimated Models of Unit 3

C. Analysis of Results

For Experiment 1, the histogram of the wind power generation and densities of the EM-based GMM and DGMM that have the lowest BIC scores for units 1 and 3 are shown in Fig. 1 and Fig. 2, respectively. In addition to brevity, we choose units 1 and 3 for two primary reasons. Indeed, not only are they representative of the entire dataset (in the sense that the other 85 generating units either have a strong correlation with unit 1 or with unit 3), but also their histograms have different shapes, where the frequencies at the boundaries of unit 1 are more dominant than those of unit 3. We do not show the DPHEM-based GMM density curve, as it is outperformed by the EM-based GMM for both units. This is consistent with the observation made in [7] that the EM-based GMM has better performance for small component numbers ($K \leq 20$).

Fig. 1 shows that with $K = 16, J = 2$, the estimated DGMM of unit 1 can more accurately captures the nuances of the original data distribution compaed to the GMM with $K = 20$. Notably, the estimated density of the DGMM successfully reproduces the peaks observed at 0.7 and 0.8, which the classical GMM fails to delineate precisely. This comparison highlights the DGMM's efficiency in achieving a better fit with a reduced total component count of $K+J = 18$, in contrast to the 20 components used by the classical GMM.

Fig. 2 reveals similar estimated densities for the GMM and DGMM of unit 3. Compared to unit 1, the frequencies at the boundaries for unit 3 are not as pronounced as those observed in the central range of the distribution. Nevertheless, the heightened frequencies at 0 and 1 allows DGMM to capitalize on its strength: the ability of capturing the peaks at these boundary values and thus provide a more refined fit with fewer Gaussian components, as also evidenced by the BIC scores.

Fig. 3 contains the plots of $\mathrm{BIC}^*(\text{EM-GMM})$, $\mathrm{BIC}^*(\text{DPHEM-GMM})$, and $\mathrm{BIC}^*(\text{DGMM})$ for each K, as well as the relative improvement, Δ_{BIC}, for unit 1. The $\mathrm{BIC}^*(\text{DGMM})$ curve resides entirely below both the $\mathrm{BIC}^*(\text{EM-GMM})$ and the $\mathrm{BIC}^*(\text{DPHEM-GMM})$ curves. For a fixed number of Gaussian components (say $K = 4$), incorporating additional components with fixed variances leads to models with lower BIC scores. On the other hand, to attain the same BIC score, the DGMM requires a leaner Gaussian component structure, augmented with up to only 4 discrete masses, whereas a GMM requires more Gaussian components. Fig. 4 exhibits the same four curves as Fig. 3 for unit 3. We observe a modest divergence in BIC scores between the EM-based GMM and DGMM at $K = 16$. However, the relative improvement of DGMM compared to GMM is significant (by more than 100%) for smaller K (say $K \in \{2, 4, 6\}$), which is also the case for unit 1.

Figs. 5 and 6 present the BIC scores for Experiment 2.

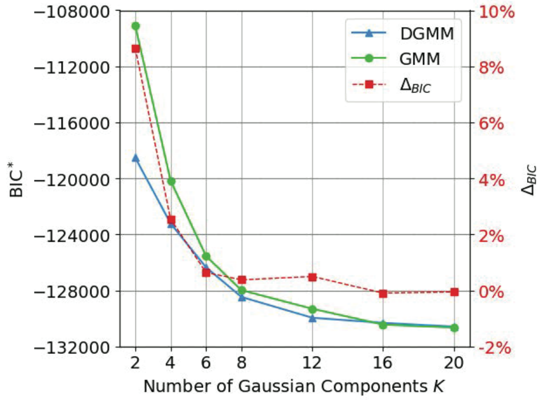

Fig. 5: BIC Score of Estimated Models of 4 Correlated Units

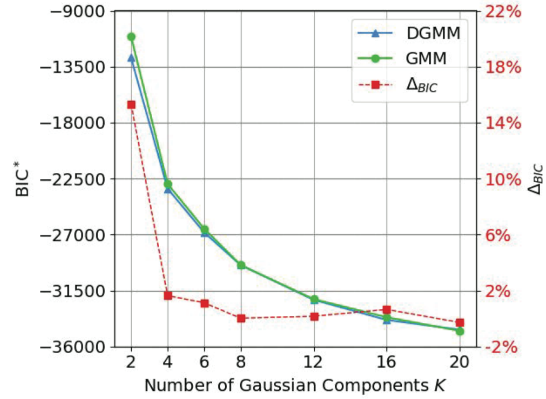

Fig. 6: BIC Score of Estimated Models of 4 Uncorrelated Units

In both scenarios, the only significant improvements are observed at smaller K, since $\mathrm{BIC}^*(\mathrm{GMM})$ and $\mathrm{BIC}^*(\mathrm{DGMM})$ curves tend to overlap as the number of Gaussian components increases. One explanation could be the higher dimensionality of the data points: the gain from reducing the number of free parameters might be too small to significantly penalize the likelihood and improve the BIC score in such cases.

IV. Conclusions

This paper proposed a novel Discrete-Gaussian Mixture Model (DGMM) for wind power generation. The DGMM generalizes the GMM since some of the mixture components can be discrete masses. The optimal locations of these discrete components can be found by the FixedEM algorithm. Our experimental results reveal that the DGMM can effectively capture the multimodal and skewed characteristics present in wind power data. Compared to the classical GMM fitted using either the EM or the recently proposed Density Preserving Hierarchical EM algorithm, the DGMM uses fewer Gaussian components to achieve a better or equal fit at the expense of introducing only a few additional discrete components. This leads to a reduction in model complexity without compromising the information represented in the data. This advantage is particularly pronounced when computation power is limited and only a small number of Gaussian components are allowed. These results are also in harmony with our motivation. Indeed, the effective use of a small number of Gaussian components to model uncertainty offers a pathway to the computationally efficient solution of the optimization problems for long-term planning.

References

[1] L. A. Roald, D. Pozo, A. Papavasiliou, D. K. Molzahn, J. Kazempour, and A. Conejo, "Power systems optimization under uncertainty: A review of methods and applications," *Electric Power Systems Research*, vol. 214, p. 108725, 2023.

[2] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson, "Analytical reformulation of security constrained optimal power flow with probabilistic constraints," in *2013 IEEE Grenoble Conference*, pp. 1–6, 2013.

[3] L. Roald and G. Andersson, "Chance-constrained AC optimal power flow: Reformulations and efficient algorithms," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2906–2918, 2017.

[4] D. Ke, C. Y. Chung, and Y. Sun, "A Novel Probabilistic Optimal Power Flow Model With Uncertain Wind Power Generation Described by Customized Gaussian Mixture Model," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 200–212, 2016.

[5] Z. Wang, C. Shen, F. Liu, X. Wu, C.-C. Liu, and F. Gao, "Chance-Constrained Economic Dispatch With Non-Gaussian Correlated Wind Power Uncertainty," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4880–4893, 2017.

[6] Y. Yang, W. Wu, B. Wang, and M. Li, "Analytical Reformulation for Stochastic Unit Commitment Considering Wind Power Uncertainty With Gaussian Mixture Model," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2769–2782, 2020.

[7] Y. Gao, X. Xu, Z. Yan, and M. Shahidehpour, "Gaussian Mixture Model for Multivariate Wind Power Based on Kernel Density Estimation and Component Number Reduction," *IEEE Transactions on Sustainable Energy*, vol. 13, pp. 1853–1856, July 2022.

[8] A. J. Conejo, L. Baringo, S. J. Kazempour, and A. S. Siddiqui, "Investment in electricity generation and transmission," *Cham Zug, Switzerland: Springer International Publishing*, vol. 106, 2016.

[9] F. Pourahmadi, J. Kazempour, C. Ordoudis, P. Pinson, and S. H. Hosseini, "Distributionally robust chance-constrained generation expansion planning," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2888–2903, 2019.

[10] Z. Hu, W. Sun, and S. Zhu, "Chance constrained programs with Gaussian mixture models," *IISE Transactions*, vol. 54, no. 12, pp. 1117–1130, 2022.

[11] J. Luedtke, S. Ahmed, and G. L. Nemhauser, "An integer programming approach for linear programs with probabilistic constraints," *Mathematical programming*, vol. 122, no. 2, pp. 247–272, 2010.

[12] S. Küçükyavuz and R. Jiang, "Chance-constrained optimization: A review of mixed-integer conic formulations and applications," *arXiv preprint arXiv:2101.08746*, p. 2, 2021.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[14] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. Hoboken, NJ: John Wiley & Sons, 2007.

[15] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[16] P. Stoica and P. Babu, "On the Proper Forms of BIC for Model Order Selection," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4956–4961, 2012.

[17] H. Li, J. H. Yeo, A. L. Bornsheuer, and T. J. Overbye, "The Creation and Validation of Load Time Series for Synthetic Electric Power Systems," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 961–969, 2021.

[18] T. J. Overbye, F. Safdarian, W. Trinh, Z. Mao, J. Snodgrass, and J. Yeo, "An approach for the direct inclusion of weather information in the power flow," in *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS)*, HICSS, January 2023.