Gaussian Mixture Model Based Distributionally Robust Optimal Power Flow With CVaR Constraints

Lei You, Student Member, IEEE, Hui Ma, Senior Member, IEEE, Tapan Kumar Saha, Fellow, IEEE, Gang Liu, Senior Member, IEEE

Abstract—This paper proposes a distributionally robust optimal power flow (OPF) model for transmission grids with wind power generation. The model uses the conditional value-at-risk (CVaR) constraints to control the reserve and branch flow limit violations caused by wind power forecast errors. Meanwhile, the Gaussian mixture model (GMM) is integrated into the CVaR constraints to guard against the non-Gaussian forecast error distributions. Unlike the previous studies considering the GMM constraints to guard against the non-Gaussian forecast error distributions, the proposed OPF model utilizes the Gaussian mixture model (GMM) to account for both the probability and magnitude of wind power forecast error at farm i.

Index Terms—Optimal power flow, Conditional value-at-risk, Gaussian mixture model, distributionally robust optimization, wind power, forecast error, ambiguity set.

NOMENCLATURE

Sets and Indices

| Symbol | Description |
|--------|-------------|
| \( b, g, l, i, m, n, k \) | Index for load, traditional unit, transmission line, wind farm, Gaussian component, bootstrap resample and random constraint, respectively. |
| \( \mathcal{B}, \mathcal{G}, \mathcal{L}, \mathcal{W} \) | Sets of buses, traditional units, transmission lines and wind farms, respectively |

Parameters

| Symbol | Description |
|--------|-------------|
| \( \bar{w}_i \) | Wind power central forecast at wind farm i. |
| \( H_W^{1,i}, H_W^{2,i}, H_D^{1,b}, H_D^{2,b} \) | Shift distribution factor of wind farm i, unit g and load b to line l, respectively. |
| \( c_g^p(\cdot) \) | Quadratic fuel cost function of unit g. |
| \( \bar{c}_g^{up}, c_g^{down} \) | Up-/down-ward reserve price of unit g. |
| \( d_b \) | Power demand of load b. |
| \( \bar{\tilde{P}}_g, \tilde{P}_g, \bar{R}_g^{up}, \tilde{R}_g^{down} \) | Maximum/minimum capacity of unit g. Maximum up-/down-ward reserve of unit g. |
| \( \tilde{f}_l \) | Transmission limit of line l. |
| \( M, N, K, J, W \) | Number of Gaussian components, bootstrap resamples, random constraints, historical data and wind farms, respectively. |

Variables

| Symbol | Description |
|--------|-------------|
| \( \zeta_i \) | Random wind power forecast error at farm i. |
| \( \alpha_g \) | Participation factor of unit g. |
| \( \tilde{p}_g \) | Nominal output of unit g. |
| \( \bar{R}_g^{up}, \tilde{R}_g^{down} \) | Up-/down-ward reserve provided by unit g. |
| \( \tilde{f}_l \) | Nominal power flow on line l. |
| \( \pi_m, \mu_m, \Sigma_m \) | Mixing weight, mean and covariance of \( m_{th} \) gaussian component, respectively. |

I. INTRODUCTION

The forecast errors of wind power can pose considerable risk to power system security. To handle the wind power forecast uncertainties, different decision-making approaches have been proposed, such as the stochastic optimization (SO) using discrete wind power scenarios [1] and the robust optimization (RO) leveraging the uncertainty sets of wind power [2].

The distributionally robust optimization (DRO) has been emerging as an intermediate between SO and RO. DRO assumes the true distribution of the forecast error lies in an ambiguity set and makes decisions to hedge against the worst-case distribution. The moment-based ambiguity set has been used for power system dispatch [3], [4]. However, such ambiguity set utilizes only the moment information of the forecast error and may be very conservative. Another type of ambiguity set leverages the statistical distance between distributions, such as the Kullback-Leibler (KL) distance and Wasserstein distance. In [5], a KL-based ambiguity set is used for the unit commitment. In [6], a Wasserstein-based ambiguity set is applied to the optimal power flow (OPF) problem. However, KL-DRO cannot be applied if the uncertainties have heavy tails [7], while the model size of Wasserstein-DRO grows heavily with the data size employed. In [8], [9], a new type of ambiguity set is also constructed based on the cumulative probability distributions of random variables, but the correlations between random variables cannot be captured.

The decision-making under uncertainty also often needs some risk management tools. Value-at-risk (VaR) and conditional value-at-risk (CVaR) are two popular tools and have been successfully applied to power system problems. Based on VaR, the chance constrained OPF and UC models have been widely used to ensure the power system security with specified probability levels [10], [11]. The use of VaR is intuitive since it limits the violation probability of a constraint. However, VaR neglects the aftermath if the violation occurs. Also, the VaR-based chance constraints are often non-linear and non-convex, and thus are computationally intractable.

Unlike VaR, CVaR is a coherent risk measure [12]. It is defined on the mean of the distributional tail exceeding VaR and can thus account for both the probability and magnitude of
constraint violation. Besides, the direct optimization of CVaR is a convex problem and is thus more tractable. In power systems, CVaR has been used for demand-side management [13] and unit commitment optimization [14].

Overall, both VaR and CVaR have their rationality and the decision maker may prefer one over the other in different situations. This paper focuses on CVaR. More specific, this paper aims to address two issues in the application of CVaR.

First, the calculation of CVaR may be still not easy since it involves high-dimensional integral. Currently, the sample-based approximation of CVaR is often used [15], but it may cause a huge computational burden under large sample sizes. An analytical reformulation of CVaR is also provided in [15], but it is valid only when the uncertainties follow Gaussian distributions. In reality, the forecast error of wind power often follows non-Gaussian distributions [16]. Recently, Gaussian mixture model (GMM) has been used to fit non-Gaussian forecast errors [17], [18]. However, the CVaR under GMM still lacks an analytical form. It is also noteworthy that though GMM has been used for VaR-based power system scheduling [19], [20], there is no existing work applying GMM to CVaR-constrained power system dispatch problems.

Second, in practice, the precise probability distribution of the forecast error is often unavailable. It could be biased to use the distribution estimated directly from limited historical data or based on expert knowledge. To address this issue, the moment-based ambiguity set with the mean and covariance information has been used to consider the worst-case CVaR [21]. Since the utilization of only moment information may result in very conservative solutions, the unimodality knowledge is further included in [22] to reduce the conservatism. Besides, the Wasserstein-based ambiguity set has also been used for CVaR-constrained OPF [23], but the resulting model is not scalable since the model size is proportional to the number of historical data. Also for the CVaR-constrained OPF, [24] provides an alternative method by combining the sample-based CVaR approximation with an ambiguous discrete probability distribution. However, similar to [23], this method is still not scalable since its size increases heavily with the number of forecast error samples.

This paper proposes a new distributionally robust OPF model with CVaR constraints. A data-driven GMM-based ambiguity set is used to hedge against the ambiguous true distribution of wind power forecast error. A scalable solution approach without any sampling procedure is also developed for the proposed OPF model. The contributions of this paper include:

1) This paper extends the application of GMM in power system operational planning. First, previous research often uses GMM in VaR-constrained dispatch problems, while it is the first time for GMM to be applied to CVaR-constrained OPF. Second, the traditional GMM has fixed parameters (e.g., see [19], [20]), while the GMM in this paper is allowed to have uncertain parameters by adopting a data-driven GMM-based ambiguity set. The use of the proposed GMM is more robust to the distributional ambiguity of wind power forecast error.

2) A scalable cutting-plane-based methodology is used to solve the proposed OPF model. Unlike the traditional methods to handle CVaR, the proposed methodology does not rely on any samplings when evaluating the worst-case CVaR and enforcing the distributionally robust (DR)-CVaR constraints and thus avoids the associated high computational burden (see, e.g., [24]). By exploiting the special properties and structure of the proposed GMM-based ambiguity set, a method is designed to quickly evaluate the worst-case CVaR. Then, based on the convexity and sensitivities of CVaR, cutting-planes are used to enforce the DR-CVaR constraints. Finally, an iterative cutting-plane-based solution algorithm is developed for the proposed OPF model. The whole methodology demonstrates high computational efficiency even for very large-scale networks.

II. CVAR CONSTRAINED OPTIMAL POWER FLOW

A. Optimal Power Flow (OPF) Formulation

As the response to wind power forecast errors, we assume the outputs of traditional units are adjusted based on the policy:

\[
\Delta \hat{p}_g = -\alpha_g \sum_{i \in W} \xi_i, \forall g \in G
\]  

(1)

where the total power imbalance is allocated among the traditional units according to the participation factors \(\alpha_g\). If \(\sum_g \alpha_g = 1\), the system-wide power balance is guaranteed.

With the above policy, the OPF dispatch with uncertain wind power can be modelled as (P1):

\[
(P1): \min_{\hat{p}, R^{ap}, R^{dn}} \sum_{g \in G} \left( c_{f}^{G} (\hat{p}_g) + c_{g}^{up} R_{g}^{up} + c_{g}^{dn} R_{g}^{dn} \right)
\]

s.t. \(0 \leq R_{g}^{up} \leq \bar{R}_{g}^{up}, \forall g \in G\) \hspace{1cm} (2a)
\(0 \leq R_{g}^{dn} \leq \bar{R}_{g}^{dn}, \forall g \in G\) \hspace{1cm} (2b)
\(p_{g}^{up} + R_{g}^{up} \leq \hat{p}_{g} \leq p_{g}^{up} - R_{g}^{up}, \forall g \in G\) \hspace{1cm} (2c)
\(\sum_{g \in G} \hat{p}_{g} + \sum_{i \in W} \tilde{w}_{i} = \sum_{g \in G} d_{g}\) \hspace{1cm} (2d)
\(\sum_{g \in G} \alpha_g = 1; \alpha_g \geq 0, \forall g \in G\) \hspace{1cm} (2e)
\(-\alpha_g \sum_{i \in W} \xi_i \leq R_{g}^{up}, \forall g \in G\) \hspace{1cm} (2f)
\(\alpha_g \sum_{i \in W} \xi_i \leq R_{g}^{dn}, \forall g \in G\) \hspace{1cm} (2g)
\(f_{i} = \sum_{g \in G} H_{i,g}^{G} \hat{p}_{g} + \sum_{i \in W} H_{i,i}^{W} \tilde{w}_{i} - \sum_{g \in G} H_{i,g}^{D} d_{g}, \forall l \in L\) \hspace{1cm} (2h)
\(\hat{f}_{l} + \sum_{i \in W} (H_{i,i}^{W} - \sum_{g \in G} H_{i,g}^{G} \alpha_g) \xi_i \leq \tilde{f}_{l}, \forall l \in L\) \hspace{1cm} (2i)
\(\hat{f}_{l} + \sum_{i \in W} (H_{i,i}^{W} + \sum_{g \in G} H_{i,g}^{G} \alpha_g) \xi_i \geq \tilde{f}_{l}, \forall l \in L\) \hspace{1cm} (2j)

where (2a) minimizes the sum of the fuel cost and reserve-procuring cost; (2b)-(2d) limit the available up- and down-ward reserves; (2e)-(2f) enforce system-wide power balance; (2g)-(2h) bound the deployment of up- and down-ward reserves; (2i) indicates the nominal branch flow while (2j)-(2k) represent the random branch flow under the wind power fluctuations.

In the following analysis, the random constraints (2g)-(2h) and (2j)-(2k) are all represented by the general form:
\[ y_k(x)^T \xi \leq h_k(x), \forall k = 1, \ldots, K \]  
where \( x \) is the decision vector in (P1), \( y_k(x) \) and \( h_k(x) \) are two affine transformations of \( x \), \( \xi \) is defined as the following GMM:

\[ CVaR_{1-\beta} (y_k^T \xi) = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{\beta} E_x [y_k^T \xi - t]^+ \right\} \quad (4) \]
where \( E_x [.]^+ = E_x [\max(.,0)]. \)

In (4), CVaR is the conditional expectation of \( y_k^T \xi \) exceeding the optimal solution of \( t \) (i.e., the value-at-risk of \( y_k^T \xi \) at the probability level \( 1 - \beta \)). Then, we use the CVaR constraints in (5) to control the violation risk of (3).

\[ CVaR_{1-\beta} (y_k^T \xi) \leq h_k(x), \forall k = 1, \ldots, K \quad (5) \]

However, the use of (5) has the following two issues:

i) Usually, we only have limited historical data of \( \xi \) and it is hard to estimate the precise distribution of \( \xi \). It can be biased to estimate CVaR by assuming any parametric distribution of \( \xi \).

ii) CVaR lacks a general computational efficient form. Though (5) can be handled by a sample-based approximation, the approximation quality could be sensitive to the number of samples and the sampling technique, and the computational burden could be huge when the number of samples is large.

### III. DISTRIBUTIONALLY ROBUST FORMULATION

To address the aforementioned issues in Section II-B, we propose a distributionally robust approach combining the CVaR constraints (5) with a GMM-based ambiguity set. This is based on the following considerations:

i) GMM is adopted due to its flexibility in modelling various types of multivariate non-Gaussian probability distributions.

ii) The GMM fitted on limited training data may deviate from the unknown true distribution of \( \xi \), so a GMM-based ambiguity set is used to account for a GMM with uncertain parameters and hedge against the distributional ambiguity of \( \xi \).

iii) Though the GMM-based CVaR still lacks an analytical form, the use of GMM makes it possible to apply a cutting-plane method to enforce the CVaR constraints (5) (see Section IV). Thus, the sample-based approximation of CVaR is avoided.

#### A. Gaussian Mixture Model (GMM)

The probability density function (PDF) of the wind power forecast error \( \xi \) is defined as the following GMM:

\[ \text{PDF} (\xi | \Psi) = \sum_{m=1}^{M} \pi_m \phi(\mu_m, \Sigma_m) \]

where \( \phi(\mu_m, \Sigma_m) \) is the multivariate Gaussian distribution with the mean \( \mu_m \) and covariance matrix \( \Sigma_m \), and it is the \( m \)-th Gaussian component of the GMM; \( \{\pi_1, \ldots, \pi_M\} \) are the mixing weights of the \( M \) Gaussian components; \( \Psi \) is the vector form of \( \{\pi_1, \ldots, \pi_M, \mu_1, \ldots, \mu_M, \Sigma_1, \ldots, \Sigma_M\} \).

Given the historical data of \( \xi \), the parameters \( M \) and \( \Psi \) can be obtained by the maximum likelihood estimation (MLE) [25].

#### B. Constructing Credible Regions for GMM Parameters

Different to the traditional GMM with fixed parameters, we assign some GMM parameters with certain uncertainty and allow them to be variable within some credible regions. This yields an ambiguous GMM. The credible regions are as follows.

i) Weights \( \pi_1, \ldots, \pi_M \) for all GMM components:

\[ C^\pi = \left\{ \pi \mid \pi_m \geq 0, \sum_m \pi_m \leq 1, \forall m = 1, \ldots, M \right\} \]

where \( \delta \) is a confidence level (e.g., 0.95) and \( [\pi^\delta_m, \pi^\delta_m] \) is the \( \delta \)-confidence interval of \( \pi_m \).

ii) Mean \( \mu_m \) for each GMM component \( m = 1, \ldots, M \):

\[ C^\mu_m = \left\{ \mu_m \mid (\mu_m - \hat{\mu}_m)^T \Lambda_m^{-1} (\mu_m - \hat{\mu}_m) \leq \gamma^\mu_m \right\} \]

where \( \mu_m \) lies in a \( \hat{\mu}_m \)-centered ellipsoid shaped by \( \Lambda_m \) and bounded by \( \gamma^\mu_m \).

iii) Covariance \( \Sigma_m \) for each component \( m = 1, \ldots, M \):

\[ C^\Sigma_m = \left\{ \Sigma_m \mid \|\Sigma_m - \hat{\Sigma}_m\|_F \leq \gamma^{\Sigma_m} \right\} \]

where \( \| \cdot \|_F \) denotes the Frobenius norm, and \( \Sigma_m \) lies inside a \( \hat{\Sigma}_m \)-centered Frobenius-norm ball bounded by \( \gamma^{\Sigma_m} \).

#### C. Quantifying the Uncertainty of GMM Parameters

Next, we provide two optional methods to quantify the uncertainty of \( \pi_m, \mu_m \) and \( \Sigma_m \) and determine the coefficients in the credible regions \( C^\pi, C^\mu \) and \( C^\Sigma \).

**Method 1: A non-parametric bootstrap method** [26]

This method estimates the empirical distribution of a statistic by calculating the statistic on many bootstrap-resamples. This empirical distribution is used to construct the confidence region of the statistic. The bootstrap procedure is as follows [26]:

i) With all historical data of \( \xi \), use MLE to obtain the optimal parameters \( M^* \) and \( \Psi^* \), where \( M^* \) is chosen according to the Bayesian information criterion; then extract the maximum likelihood posterior group membership probability matrix \( Z^* \).

ii) Construct \( N \) bootstrap resamples \( \{Y_1, Y_2, \ldots, Y_N\} \) by drawing random observations with replacement from the original historical observations of \( \xi \).

iii) For each \( Y_n \), initialize the MLE with \( M^* \) and \( Z^* \); then conduct the MLE to obtain the corresponding optimal \( \Psi_m^* = \{\pi_1^*, \ldots, \pi_M^*, \mu_1^*, \ldots, \mu_M^*, \Sigma_1^*, \ldots, \Sigma_M^*\} \).

With the obtained parameters \( \Psi_m^* \), we can determine the coefficient of the credible regions in Section III-B.

For the credible region \( C^\pi \), we estimate

\[ \pi_m^* = \frac{1}{N} \sum_n \delta_m \]

\[ (10a) \]
\[ \Delta^\delta = \Gamma_{1/2}[\pi^1_m, \ldots, \pi^N_m] \]  
(10b)

where \( \Gamma_{1/2}[\cdot] \) and \( \Gamma_{1/2}[\cdot] \) are the upper \( \frac{1}{2} \) and \( \frac{1}{2} \) quantiles of a data series, respectively.

For the credible region \( C^m_\mu \), we estimate
\[
\hat{\mu}_m = \frac{1}{N} \sum_{n=1}^N \mu^*_m
\]
(11a)

\[ \Lambda_m = \frac{1}{N-1} \sum_{n=1}^N (\mu^*_m - \mu_m)(\mu^*_m - \mu_m)^T \]  
(11b)

\[ \gamma^*_m = \Gamma_\delta \left[ (\mu^*_m - \mu_m)^T \Sigma^{-1}_m (\mu^*_m - \mu_m), \ldots, (\mu^*_m - \mu_m)^T \Sigma^{-1}_m (\mu^*_m - \mu_m) \right] \]  
(11c)

For the credible region \( C^\Sigma_m \), we estimate
\[
\hat{\Sigma}_m = \frac{1}{N} (\Sigma^1_m + \ldots + \Sigma^N_m)
\]
(12a)

\[ \gamma^\Sigma_m = \Gamma_\delta \left[ \| \Sigma^1_m - \hat{\Sigma}_m \|_F, \ldots, \| \Sigma^N_m - \hat{\Sigma}_m \|_F \right] \]  
(12b)

Method 2: A method from a Bayesian viewpoint

This method performs asymptotic inference on the estimator of \( \Psi \). More specific, under suitable regularity conditions and for large data sizes of \( \hat{\xi} \), the posterior distribution of \( \Psi \) can be approximated by the following normal distribution and it does not depend on the prior distribution of \( \Psi \) (see Section 4 in [27]):

\[
\phi \left( \Psi^*, \left( \hat{f}(\Psi^*) \right)^{-1} \right)
\]
where \( \Psi^* \) is the MLE of \( \Psi \), \( \hat{f}(\Psi^*) \) is the estimated Fisher information matrix evaluated at \( \Psi^* \) and it can be calculated based on the methods in [28], \( J \) is the data size of \( \hat{\xi} \). With this normal distribution, it is easy to determine the coefficients in the above credible regions and the details are omitted here.

The Bayesian method does not need the bootstrap resamples and is computationally very cheap. As the data size gets larger, the Bayesian method can be more accurate and is expected to offer a lower level of conservativeness associated with the credible regions. This is desirable since it helps make less-conservative decisions when more data information is at hand.

D. Distributionally Robust CVaR-constrained OPF

After determining the credible regions of the GMM parameters, we can construct the following GMM-based ambiguity set to account for an ambiguous GMM distribution:

\[
\mathcal{D} = \left\{ \sum_{m=1}^{M^*} \pi_m \phi(\mu_m, \Sigma_m) : \mu_m \in C^\mu_m, \Sigma_m \in C^\Sigma_m, \forall m = 1, \ldots, M^* \right\} 
\]
(13)

With the ambiguity set in (13), we can consider the following distributionally robust CVaR (DR-CVaR) constraints:

\[
\max_{\gamma \in \mathcal{D}} \text{CVaR}_{1-\beta}(\gamma^T \xi) \leq h_k(x), \forall k = 1, \ldots, K 
\]  
(14)

where the worst-case CVaR of \( \gamma^T \xi \) is no higher than \( h_k(x) \).

Now, we formulate our DR-CVaR-constrained OPF as (P2):

\[
(P2): \min_{\gamma} (2a), \text{ s.t.: } (2b)-(2f), (2i), (14). 
\]
(15)

IV. Solution Methodology

Next, we design an iterative cutting-plane algorithm to enforce the DR-CVaR constraints (14) and solve problem (P2).

A. Cutting-plane Method

To handle the DR-CVaR constraints (14), we will utilize two important and very desirable properties of CVaR:

i) since \( \gamma^*_k \xi \) is convex (affine) in \( y_k \), \( \text{CVaR}_{1-\beta}(\gamma^*_k \xi) \) is convex in \( y_k \) (see the Corollary 11 in [12]).

ii) it is easy to verify that \( \gamma^*_k \xi \) satisfies the Assumptions 1-3 in [29], so the gradients of \( \text{CVaR}_{1-\beta}(\gamma^*_k \xi) \) with respect to \( y_k \) can be calculated as the following conditional expectations (see Theorem 3.1 in [29]):

\[
\nabla_{y_k,i} \text{CVaR}_{1-\beta}(\gamma^* k \xi) = E_{\gamma} \left[ \frac{\xi_i y_k^T \xi - \text{VaR}_{1-\beta}(\gamma^*_k \xi)}{y_k^T \xi} \right], \forall i 
\]
(16)

where \( y_{k,i} \) denotes the \( i \)th element of \( y_k \), and \( \text{VaR}_{1-\beta}(\gamma^*_k \xi) \) is the value-at-risk of \( y^T_k \xi \) at the probability level \( 1 - \beta \).

In [29], a Monte-Carlo sampling method is used to estimate the VaR, CVaR and its gradients in (16). However, this method is computationally inefficient, and it is difficult to account for the distributional ambiguity of \( \xi \) within such method.

Next, we will show that, thanks to the use of GMM, a precise and scalable method can be applied to calculate the VaR, CVaR and gradients in (16) while considering the distributional ambiguity of \( \xi \) as modelled by (13).

Step 1: identify the worst-case \( \mu_m \) and \( \Sigma_m \).

For a given \( y_k \), the corresponding worst-case CVaR of \( y^T_k \xi \) can be reformulated as (17) by combining (4), (13) and (14):

\[
\max_{\gamma \in \mathcal{D}} \text{CVaR}_{1-\beta}(\gamma^T \xi) = \max_{\gamma \in \mathcal{D}} \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{\beta} E_{\gamma} [y^T_k \xi \xi - t] \right\} 
\]
(17)

In (17), the auxiliary variable \( Q_m \) is an integral under a multivariate normal distribution

\[
Q_m = \int_{\xi \in \mathcal{W}} [y^T_k \xi - t]^+ \phi(\xi | \mu_m, \Sigma_m) d\xi
\]
and it has an analytical form

\[
Q_m = \sigma_m^2 + \phi_m(\sigma_m) + (\mu_m - t)[1 - \Phi_m(\sigma_m)] 
\]
(18)

where \( \phi_m \) and \( \Phi_m \) are respectively the PDF and cumulative distribution function (CDF) of the normal distribution with mean \( \mu_m = y^T_k \mu_m \) and standard deviation \( \sigma_m = \sqrt{y^T_k \Sigma_m y_k} \).

It can be verified that \( Q_m \) is monotonically increasing with \( \mu_m \) and \( \sigma_m \) for any \( t \). Thus, the optimal solution of (17) corresponds to the maximum values of \( \mu_m \) and \( \sigma_m \).

The maximum \( \mu_m \) and \( \sigma_m \) can be obtained by identifying the worst-case \( \mu_m \in C^\mu_m \) and worst-case \( \Sigma_m \in C^\Sigma_m \). Due to
the special structures of the $\mathcal{C}_m$ and $\mathcal{E}_m^\infty$ in (8)-(9), the worst-case $\mu_m$ and $\Sigma_m$ have closed-form solutions:

$$\hat{\mu}_m^\max = y_k^T \mu_{m\infty}$$  \hspace{1cm} (19a)$$

$$\mu_{m\infty} = \hat{\mu}_m + \frac{\Sigma_m}{y_k^T \Lambda_m y_k} \Lambda_m y_k$$  \hspace{1cm} (19b)$$

$$\tilde{\sigma}_m^\max = \sqrt{y_k^T \Sigma_{m\infty} y_k}$$  \hspace{1cm} (19c)$$

$$\Sigma_{m\infty} = \overline{\Sigma} + \gamma_m y_k y_k^T \frac{\eta}{\|\eta\|^2}$$  \hspace{1cm} (19d)$$

where the superscript $\omega\infty$ means ‘the worst-case’.

Step 2: identify the worst-case $\pi$ and the corresponding $\text{VaR}_{1-\beta}(y_k^T \tilde{E})$.

It is easy to verify that $Q_m$ is convex in $t$, so (17) can be rewritten as (20) according to the max-min theorem [30]:

$$\min_{t \in \mathbb{R}} \left[ t + \frac{1}{\beta} \max_{\pi \in \mathcal{P}} \sum_{m=1}^{M} \pi_m Q_m |\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max \right]$$  \hspace{1cm} (20)$$

To simplify expressions, let $T(t)$ be the function inside the square bracket in (20), and so (20) is equal to

$$\min_{t \in \mathbb{R}} T(t)$$  \hspace{1cm} (21)$$

Due to the convexity of $Q_m$ in $t$, it is easy to proof that $T(t)$ is also convex in $t$. Also, problem (21) involves only one decision variable $t$, and $T(t)$ could be non-smooth. These observations motivate us to design a bisection search method to solve (21):

Algorithm 1: A bisection search method to solve (21)

1. Receive $\hat{\mu}_m^\max$ and $\tilde{\sigma}_m^\max$, $\forall m$.
2. Let $\bar{t}_m = \hat{\mu}_m^\max + \Phi^{-1}(1-\beta)\tilde{\sigma}_m^\max$, $\forall m$, where $\Phi^{-1}(1-\beta)$ is the inverse of the CDF of standard normal distribution evaluated at $1-\beta$.
3. Let $\xi = \min_{m} \bar{t}_m$, $\bar{T} = \max_{m} \bar{t}_m$.
4. while $|t - \xi| > 10^{-5}$
   
   $t^n = \frac{t + \xi}{2}$
   
   Calculate $Q_m(t^n)|\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max$, $\forall m$, according to (18); $\pi^n = \text{Sort}(\mathcal{E}_m^\xi, Q_m(t^n))|\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max$, $\forall m$; Fix the $\pi$ in $T(t)$ as $\pi^n$, calculate the gradient of $T(t)$ at $t^n$:
   
   $\nabla_i T(t^n)|\pi^n, \tilde{\mu}_m^\max, \tilde{\sigma}_m^\max$ = \frac{1}{2} \sum_{m=1}^{M} \frac{\sum_{m=1}^{M} \pi_m Q_m(t^n)|\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max}}{\pi_m Q_m(t^n)|\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max}}$
   
   if $\nabla_i T(t^n)|\pi^n, \tilde{\mu}_m^\max, \tilde{\sigma}_m^\max$ < 0, then $t = t^n$; else $\xi = t^n$; end if end while
5. Output $\pi^n, \pi^n$ and the corresponding $\text{VaR}_{1-\beta}(y_k^T \tilde{E}) = t^n$.

where the structure of $\mathcal{C}_m$ allows us to use the sorting method in [31] (i.e., SortMethod) to identify the $\pi \in \mathcal{C}_m$ that corresponds to the maximum $\sum_{m=1}^{M} \pi_m Q_m(t^n)|\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max$ (i.e., $\pi^n$).

Step 3: develop the cutting-plane of $\max_{\pi \in \mathcal{P}} \text{VaR}_{1-\beta}(y_k^T \tilde{E})$

With the identified $\mu_{m\infty}, \Sigma_{m\infty}$ and $\pi_{m\infty}$, the worst-case $P \in \mathcal{P}$ is determined and let $\text{CVaR}_{1-\beta}(y_k^T \tilde{E})$ be the CVaR corresponding to such $P$. For the given $y_k^T$, $\text{CVaR}_{1-\beta}(y_k^T \tilde{E})$ can be calculated by substituting $\hat{\mu}_m, \tilde{\sigma}_m^\max, \pi_{m\infty}$ and $t = \text{VaR}_{1-\beta}(y_k^T \tilde{E})$ into (17)-(18). The gradients of $\text{CVaR}_{1-\beta}(y_k^T \tilde{E})$ at $y_k^T$ can be obtained according to (16) with the analytical forms:

$$\nabla_{y_k} \text{CVaR}_{1-\beta}(y_k^T \tilde{E}) = \frac{1}{\beta} \sum_{m=1}^{M} \pi_m \left[ 1 - \Phi_{\beta} \left( \frac{\text{VaR}_{1-\beta}(y_k^T \tilde{E})|\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max} {\| \eta \|^2} \right) \right] \nabla_{y_k} \Phi_{\beta}(t_m y_k^T + \Phi_{\beta}(t_m y_k^T)|\tilde{\mu}_m^\max, \tilde{\sigma}_m^\max)} \forall i$$

It is easy to verify that $\max_{\pi \in \mathcal{P}} \text{CVaR}_{1-\beta}(y_k^T \tilde{E})$ is also convex in $y_k$, and the value and gradients of $\max_{\pi \in \mathcal{P}} \text{CVaR}_{1-\beta}(y_k^T \tilde{E})$ at $y_k^T$ are the same as those of $\text{CVaR}_{1-\beta}(y_k^T \tilde{E})$. Thus, we can use the cutting-plane supporting $\text{CVaR}_{1-\beta}(y_k^T \tilde{E})$ at $y_k^T$ to linearize the $k_{th}$ DR-CVaR constraint in (14). That is

$$\text{CVaR}_{1-\beta}(y_k^T \tilde{E}) + \sum_{i} \nabla_{y_k,i} \text{CVaR}_{1-\beta}(y_k^T \tilde{E})(y_k,i) (y_k,i) \leq h_k(x)$$

B. Iterative Solution Algorithm for Problem (P2)

With the cutting-plane-based constraint (23), we can develop an iterative solution algorithm to gradually enforce the DR-CVaR constraints in (14) and solve the problem (P2) in (15).

Algorithm 2: Iterative solution algorithm for problem (P2)

1. Solve the following problem (P3), which is the relaxation of (P2):

$$(P3): \left( \min_{\pi \in \mathcal{P}} \text{(2b)} \right) \text{and the cutting-plane-based constraints which have been generated }$$

2. Obtain the optimal solution $\pi$, and form $y_k^*, \forall k = 1, ..., K$.
3. for $k = 1, ..., K$, do
   
   Use the Steps 1-3 in Section IV-A to calculate $\text{CVaR}_{1-\beta}(y_k^T \tilde{E})$.
   
   if $\text{CVaR}_{1-\beta}(y_k^T \tilde{E}) > h_k(x)$ then
      
      Generate the cutting-plane-based constraint (23).
   end if
   
   end for
4. if $\text{CVaR}_{1-\beta}(y_k^T \tilde{E}) \leq h_k(x)$, $\forall k$, then terminate; Otherwise, go to Step 1.

We also have one enhancement to accelerate Algorithm 2. The upward-reserve DR-CVaR constraints are rewritten as:

$$\alpha_{g} \max_{\pi \in \mathcal{P}} \text{CVaR}_{1-\beta}(-T \tilde{E}) \leq R_{g_{up}}, \forall g$$

We can first calculate the worst-case CVaR of $-T \tilde{E}$ through the Steps 1-3 in Section IV-A (the result is denoted as $\Theta_{UP}$) and use (25) to replace the original constraints in (24).

$$\alpha_{g} \Theta_{UP} \leq R_{g_{up}}, \forall g$$

Similarly, after calculating the worst-case CVaR of $T \tilde{E}$ evaluated at $1-\beta$ (the result is $\Theta_{DN}$), we can use (26) to replace the downward-reserve DR-CVaR constraints in (P2).

$$\alpha_{g} \Theta_{DN} \leq R_{g_{dn}}, \forall g$$

Now, the reserve-related DR-CVaR constraints in (P2) are replaced by linear constraints (25)-(26), and we only need to apply the cutting-planes to branch-flow DR-CVaR constraints.
C. Remarks on the Solution Methodology

1) The whole methodology does not rely on any sampling procedure. First, for the evaluation of the worst-case CVaR, the Steps 1-3 in Section IV-A can be quickly implemented since they do not require any sample-based simulations but only involve the bisection search and some analytical solutions. Second, for the solving of the proposed OPF model, cutting-planes are used to enforce the DR-CVaR constraints. Compared to the traditional methods with sample-based approximation of CVaR (e.g., [24]), the cutting-plane method is computationally more desirable since it makes the scale of the proposed OPF model irrelevant to the size of historical samples.

2) The structures of $C^\pi$, $C^m$, $C^\Sigma$ in (7)-(9) are crucial for the scalability of the Steps 1-3 in Section IV-A: a) the structures of $C^m$ and $C^\Sigma$ for the analytical solutions of $\hat{\mu}^{max}$ and $\hat{\sigma}^{max}$; b) the structure of $C^\pi$ for the fast implementation of Algorithm 1.

3) Our experiments show that usually a very limited number of cutting-planes are actually active in each iteration, so the size of problem (P3) increases slowly with the iteration.

V. CASE STUDIES

The proposed OPF model is tested on a modified IEEE 2736-bus system, with the original data from [32]. The system has 2736 buses, 289 controllable units, 3504 transmission lines and 10 wind farms. The total capacities of the controllable units and wind farms are 28.88 GW and 3.08 GW, respectively. The flow limits of all transmission lines are reduced by 20% to create transfer congestions. All numerical tests are carried out on a desktop with Intel Core i7-8700 CPU and 16GB RAM.

A. Construction of the Proposed Ambiguity Set

In this paper, the underlying marginal distributions of the forecast errors at all wind farms and the associated correlations are extracted from Wind Integration National Dataset Toolkit [33]. Based on the marginal distributions and correlations, the Nataf method in [34] is used to generate some forecast error samples, which forms the historical dataset available to system operators. With the samples as the input, the bootstrap procedure in Section III-C is used to construct the proposed ambiguity set. The number of the bootstrap resamples is 2000. The confidence level $\delta$ in (10)-(12) is 95%.

As an example, 4000 forecast error samples are generated as the input. The number of the Gaussian components given by the bootstrap method is six. For the $4^{th}$ Gaussian component, the associated bootstrap means of the forecast errors at wind farms 1, 2 and 5 are plotted in Fig. 1. It is seen that there is a correlation between the means for wind farm 1 and 2. This is speculated to be related to the high dependency between the forecast errors themselves (the correlation coefficient of the two forecast errors is around 0.75). As a comparison, the means for wind farms 1 and 5 have no obvious dependency since the corresponding forecast errors are weakly correlated. The potential correlation between the means is another reason why we use ellipsoids as the credible regions of the means (the first reason can be seen in Section IV-C), considering ellipsoids have a good capability of modelling correlated factors [35].

B. Comparison of Different CVaR-constrained OPF Models

The proposed OPF model, denoted by DG-OPF, is compared with three other OPF models with CVaR constraints:

i) NA-OPF: the GMM with fixed parameters is used and no distributional ambiguity is considered. Here, the GMM fitted at the first step of the bootstrap procedure is used.

ii) M-OPF: The traditional moment-based ambiguity set is used, which utilizes only the estimated mean $\hat{\mu}$ and covariance $\sum$ of the forecast error (see its structure in [3], [4]). In this case, the worst-case CVaR of $y_k^T \xi$ has a simple form [21]:

$$\quad y_k^T \hat{\mu} + \sqrt{\frac{1 - \beta}{\beta}} \sqrt{y_k^T \sum y_k}$$

iii) U-OPF: This model applies not only the mean and covariance information but also the unimodality knowledge to construct its ambiguity set. The model can be formulated as a semi-definite programming problem (see more details in [22]).

DG-OPF, NA-OPF and M-OPF are solved by GUROBI while U-OPF is solved by MOSEK.

We consider two sample sizes: 200 and 4000. The risk levels $\beta$ for reserve and branch flow-CVaR constraints are set to 2% and 4%, respectively. For each sample size, we do the following procedure ten times: a) generate samples; b) solve all the four models; c) re-generate $10^6$ forecast error scenarios using the Nataf method; d) for each model, keep the optimal operating strategy fixed, and do the out-of-sample test (to check if all CVaR constraints are satisfied under the $10^6$ scenarios).

For each OPF model, the average, maximum and minimum cost results are listed in Table I. Meanwhile, for each model, the worst-case violations of various CVaR constraints observed in the out-of-sample tests are reported in Table II. It is seen from Table I and II that NA-OPF has the lowest cost but fails to meet the reliability requirements (the violated CVaR constraints are marked by boldface type). The reason is that NA-OPF makes decisions based on a deterministic GMM distribution, which is estimated from a small dataset with limited distributional information. This GMM could deviate from the underlying true distribution, and the NA-OPF dispatch may become unsafe. In contrast to NA-OPF, each of the other three models accounts for not a single but a family of possible true distributions. This needs additional cost but can hedge against the ambiguity in the true distribution and yield more robust operational strategies.

For DG, M and U-OPF models, they all satisfy the reliability requirements, but this is at the expense of different costs. As shown in Table I, M-OPF needs the highest total cost and the
inclusion of the unimodality information in U-OPF decreases the required cost. Compared to M- and U-OPF, DG-OPF needs much lower cost. To better present this point, the additional costs of the three models used to hedge against the distributional ambiguity (i.e., the additional costs in relative to the total cost of NA-OPF) are listed in Table III. Under the data size 200, M and U-OPF need additional 7.97% and 6.74% costs in average, respectively, while such percentage is reduced to 2.21% in DG-OPF. Similar results are observed for the data size 4000.

| Data size | Total Cost (10%) | NA-OPF | DG-OPF | M-OPF | U-OPF |
|-----------|------------------|--------|--------|-------|-------|
| 200       | avg              | 1.0792 | 1.1031 | 1.1653 | 1.1520 |
|           | max              | 1.0831 | 1.1189 | 1.1727 | 1.1574 |
|           | min              | 1.0747 | 1.0932 | 1.1511 | 1.1399 |
| 4000      | avg              | 1.0790 | 1.1021 | 1.1627 | 1.1495 |
|           | max              | 1.0799 | 1.1162 | 1.1645 | 1.1511 |
|           | min              | 1.0783 | 1.0941 | 1.1597 | 1.1469 |

We also compare the worst-case distributions in M, U and DG-OPF models. For the upward reserve CVaR constraints, the associated worst-case distributions of the total wind power forecast error are plotted in Fig. 2. The empirical histogram of the total forecast error formed by the $10^6$ scenarios for out-of-sample-tests and the upward reserves procured by the three models are also plotted. The worst-case distribution of M-OPF is a two-supported discrete distribution, and the left support has a total probability mass of 2% and is far from the zero point, which causes the procurement of the large amount of upward reserve; In the U-OPF with the unimodality information, the worst-case distribution consists of two uniform distributions, and the left one has a total probability mass of 2.89%, and the occurrence probability of very large downward forecast errors is reduced; For DG-OPF, its worst-case distribution is still a continuous GMM and it has a shorter stretch at the left-hand tail compared to the cases in M and U-OPF, and thus the occurrence probability of very large downward forecast errors is the lowest.

The difference in the conservativeness of DG, M and U-OPF can be also observed in Table IV, which lists the amounts of the upward reserves procured by these models. In average, the amount for M-OPF is more than 70% higher than that for DG-OPF. Though the amount for M-OPF guarantees no violations of the upward reserve-related CVaR constraints, it could be unnecessarily high.

| Data size | Upward Reserve (MW) | NA-OPF | DG-OPF | M-OPF | U-OPF |
|-----------|---------------------|--------|--------|-------|-------|
| 200       | avg                 | 592.83 | 834.98 | 1496.72 | 1393.93 |
|           | max                 | 688.28 | 1072.74 | 1602.50 | 1481.33 |
|           | min                 | 525.92 | 715.76 | 1349.84 | 1258.56 |
| 4000      | avg                 | 625.30 | 848.82 | 1463.64 | 1365.99 |
|           | max                 | 645.28 | 941.57 | 1482.60 | 1382.59 |
|           | min                 | 605.85 | 774.24 | 1437.54 | 1339.99 |

Data size Total CPU time (s) DG-OPF M-OPF U-OPF
200 avg 3.26 1.95 8.75
    max 3.56 2.74 12.10
    min 3.05 1.56 5.14
4000 avg 3.52 2.20 8.87
      max 3.87 2.64 9.04
      min 3.08 1.54 8.78

Table V lists the solving time of DG, M, and U-OPF. As can be seen, the solving time of DG-OPF is longer than that of M-OPF but shorter than that of U-OPF. The reasons include that DG-OPF has more complex reformulations of DR-CVaR.
constraints compared to M-OPF, and DG-OPF involves only linear programs while U-OPF is a semi-definite model. Still, DG-OPF can be solved within seconds on this 2736-bus system. This proves the usefulness of DG-OPF in short-term dispatch (e.g., 5 minutes ahead) on large networks. The solving time of DG-OPF also changes slightly as the data size increases. This is expected since the scale of DG-OPF is irrelevant to the data size. Moreover, DG-OPF uses averagely only 37 ms to evaluate the worst-case CVaRs for totally 7008 branch flow constraints. This shows the scalability of the Steps 1-3 in Section IV-A.

In summary, compared to M and U-OPF, DG-OPF uses more economical dispatch to satisfy the same reliability requirements while still maintaining a high computational efficiency.

VI. CONCLUSION

This paper proposes a GMM-based distributionally robust OPF model with CVaR constraints. A new ambiguity set based on uncertain GMM parameters is constructed, and a scalable cutting-plane-based solution methodology is designed. Case studies on a modified 2736-bus system show that the proposed GMM-based ambiguity set is capable of effectively hedging against the distributional ambiguity of wind power forecast error while offering less conservative dispatch than the moment and unimodality-based ambiguity sets. The case studies also reveal the scalability of the proposed OPF model, validating its potential application for short-term dispatch on large networks.

VII. REFERENCES