Distributionally Robust Joint Chance-Constrained Dispatch for Integrated Transmission-Distribution Systems via Distributed Optimization

Junyi Zhai, *Member, IEEE*, Yuning Jiang, *Member, IEEE*, Yuanming Shi, *Senior Member, IEEE*
Colin N. Jones, *Member, IEEE*, Xiao-Ping Zhang, *Fellow, IEEE*

***Abstract*—This paper focuses on the distributionally robust dispatch for integrated transmission-distribution (ITD) systems via distributed optimization. Existing distributed algorithms usually require synchronization of all subproblems, which could be hard to scale, resulting in the under-utilization of computation resources due to the subsystem heterogeneity in ITD systems. Moreover, the most commonly used distributionally robust individual chance-constrained dispatch models cannot systematically and robustly ensure simultaneous security constraint satisfaction. To address these limitations, this paper presents a novel distributionally robust joint chance-constrained (DRJCC) dispatch model for ITD systems via asynchronous decentralized optimization. Using the Wasserstein-metric based ambiguity set, we propose data-driven DRJCC models for transmission and distribution systems, respectively. Furthermore, a combined Bonferroni and conditional value-at-risk approximation for the joint chance constraints is adopted to transform the DRJCC model into a tractable conic formulation. Meanwhile, considering the different grid scales and complexity of subsystems, a tailored asynchronous alternating direction method of multipliers (ADMM) algorithm that better adapts to the star topological ITD systems is proposed. This asynchronous scheme only requires local communications and allows each subsystem operator to perform local updates with information from a subset of, but not all, neighbors. Numerical results illustrate the effectiveness and scalability of the proposed model.**

***Index Terms*—Integrated transmission-distribution (ITD) systems, asynchronous alternating direction method of multipliers (ADMM), distributionally robust joint chance-constrained (DRJCC) optimization, Wasserstein metric**

NOMENCLATURE

Function

$\mathbb{I}_{\mathcal{X}}(\cdot)$	Indicator function of set $\mathcal{X}$, i.e., if $x \in \mathcal{X}$, $\mathbb{I}_{\mathcal{X}}(x) = 0$; Otherwise, $\mathbb{I}_{\mathcal{X}}(x) = 1$.
$\mathbb{E}_{\mathbb{P}}\{\cdot\}$	Expectation over distribution $\mathbb{P}$
$\mathbb{P}$	A probability measure/distribution

Sets

$\mathcal{A}_i$	Set of parent nodes of node i in DS
$\mathcal{C}_i$	Set of children nodes of node i in DS
$\mathcal{DG}_k$	Set of controllable units in DS k
$\mathcal{DL}_k$	Set of distribution lines in DS k indexed by $i \in \mathcal{DN}_k$
$\mathcal{DN}_k$	Set of non-root distribution nodes in DS k
$\mathcal{DW}_k$	Set of renewable energies in DS k
$\mathcal{K}$	Set of DSs
$\mathcal{TD}$	Set of boundary nodes connecting to DSs
$\mathcal{TG}$	Set of controllable units in TS
$\mathcal{TL}$	Set of transmission lines in TS
$\mathcal{TN}$	Set of transmission nodes in TS
$\mathcal{TW}$	Set of renewable energies in TS

Parameters

$\epsilon^D_{1-3}/\epsilon^T_{1-2}$	Risk parameters of DS/TS
$\hat{P}^{W,D}_i/\hat{Q}^{W,D}_i$	Forecasted active/reactive power of renewable energy i in DS
$\hat{P}^{W,T}_i$	Forecasted active power of renewable energy i in TS
$\overline{d}^D_i/\overline{u}^D_i$	Maximum downard/upward reserve capacity adjustment of controllable unit i in DS
$\overline{d}^T_i/\overline{u}^T_i$	Maximum downard/upward reserve capacity adjustment of controllable unit i in TS
$\overline{P}^{Tie}_i$	Maximum active power of tie-line in boundary node i of TS
$\overline{L}_l$	Line flow limit of transmission line l
θ	Wasserstein radius
$\underline{P}^{G,D}_i/\overline{P}^{G,D}_i$	Minimum/maximum active power of controllable unit i in DS
$\underline{P}^{G,T}_i/\overline{P}^{G,T}_i$	Minimum/maximum active power of controllable unit i in TS
$\underline{Q}^{G,D}_i/\overline{Q}^{G,D}_i$	Minimum/maximum reactive power of controllable unit i in DS
$\underline{V}_i/\overline{V}_i$	Minimum/maximum voltage magnitude of distribution node i
φ	Constant power factor
$c^{r,D}_i/c^{r,T}_i$	Cost coefficients of reserve adjustment of controllable unit i in DS/TS
c^D_{i1-3}/c^T_{i1-3}	Cost coefficients of controllable unit i in DS/TS
$M_{gl}/M_{jl}/M_{il}/M_{kl}$	Generation shift distribution factors of controllable units/renewable energies/loads/tie-line power injection in TS
$P^{L,D}_i/Q^{L,D}_i$	Active/reactive load of node i in DS
$P^{L,T}_i$	Active load of node i in TS
r_i/x_i	Resistance/reactance of distribution line i
V_0	Reference voltage of substation in the root node

This work was supported in part from the Youth Program of Natural Science Foundation of Jiangsu Province (BK20210103) and the Swiss National Science Foundation under the RISK project (Risk Aware Data Driven Demand Response, grant number 200021 175627) (Corresponding author: Yuning Jiang)

J. Zhai is with the College of New Energy, China University of Petroleum (East China) and with the State Grid (suzhou) City & Energy Research Institute, China (e-mail: zhaijunyi@163.com).

Y. Jiang and C. N. Jones are with the Automatic Control Laboratory, EPFL, Lausanne, Switzerland. (e-mail: yuning.jiang, colin.jones@epfl.ch).

Y. Shi is with the School of Information Science and Technology, ShanghaiTech University, China (e-mail:shiym@shanghaitech.edu.cn)

X. Zhang is with the Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham, United Kingdom (e-mail: X.P.Zhang@bham.ac.uk).

Random Variables

ξ_i^T Active power forecast errors of renewable energy i in TS

$\xi_i^{P,D}/\xi_i^{Q,D}$ Active/reactive power forecast errors of renewable energy i in DS

Decision Variables

α_i^T/α_i^D Participation factors of controllable unit i in TS/DS

$\hat{P}_i/\hat{Q}_i$ Active/reactive power flow of distribution line i in the nominal scenario

$\hat{P}_i^{G,D}/\hat{Q}_i^{G,D}$ Reference base-points of active/reactive power of controllable unit i in DS in the nominal scenario

$\hat{P}_i^{G,T}$ Reference base-points of active power of controllable unit i in TS in the nominal scenario

$\hat{V}_i$ Voltage magnitude of distribution node i in the nominal scenario

d_i^D/u_i^D Downard/upward reserve capacity adjustment of controllable unit i in DS

d_i^T/u_i^T Downard/upward reserve capacity adjustment of controllable unit i in TS

$P^{Tie,D}/Q^{Tie,D}$ Exchanged active/reactive power with TS

P_i/Q_i Active/reactive power flow of distribution line i under the realization of renewable energy

$P_i^{G,D}/Q_i^{G,D}$ Active/reactive power of controllable unit i in DS under the realization of renewable energy

$P_i^{G,T}$ Active power of controllable unit i in TS under the realization of renewable energy

$P_i^{Tie,T}$ Exchanged active power with DS in boundary node i of TS

V_i Voltage magnitude of distribution node i under the realization of renewable energy

Main symbols and notation are defined here for quick reference. Others are defined by their first appearance as required. Boldface lower case and upper case letters represent vectors and matrices, respectively. The superscript T and D represents variables of transmission system and distribution system, respectively. And the notation $\mathbb{Z}_{z_1}^{z_2} = \{z \in \mathbb{Z} | z_1 \leq z \leq z_2\}$ is used to denote integer ranges.

I. Introduction

A. Motivation and Background

The rapid uptake of distributed generations is leading to the rise of active distribution grids that can actively manage themselves in modern power systems. Power systems incorporating active distribution grids can be more reliable, efficient, and cost-effective [1]. As a result, research interests have been focused on the coordination of integrated transmission-distribution (ITD) systems.

Without the inclusion of active distribution grids, a centralized dispatch scheme can be utilized. However, distribution system (DS) and transmission systems (TS) in ITD systems are operated separately by DS operators (DSOs) and a TS operator (TSO). These independent entities are not fully aware of each other's networks and generating decisions. The transmission and distribution grids are parts of an interconnected system, any decisions made by TSO (DSOs) affect the DSOs' (TSO's) operation and decisions. In the United Kingdom, for instance, the centralized operation between TSO and DSOs becomes almost impossible under the deregulated electricity market environment. Thus, distributed optimization methods [2]–[17] have gradually became an efficient alternative for the cooperative operation of ITD systems. Given a distributed framework, the TSO and each DSO can operate independently and collaborate by sharing limited information. Accordingly, each subsystem keeps proprietary data, including operation states and topological information, confidential without compromising data privacy and decision-making independence. However, each subsystem of ITD systems may have different scales and different computing capacities in reality. Most of these distributed optimization methods are in a synchronous setting, where all workers need to wait for the slowest worker to finish its computation or communication. It may lead to the under-utilization of both computation and communication resources as some workers remain idle for most of the time [18], [19]. For the heterogeneous ITD systems, how to efficiently coordinate the operation of the entire system in an asynchronous way has been one of the motivations behind the rise of distributed optimization.

Another key issue that is required to consider in the coordinated operation of ITD systems discreetly is the uncertainties caused by integrated renewable energies. Large-scale renewable energies, such as wind generation, have been integrated into modern power systems with ever-increasing penetration. Efficient integration of renewable energies to meet the electricity demand while respecting the operational constraints is a fundamental challenge for ITD systems. The dispatch problem under uncertainties results in a robust optimization (RO) [10], [20]–[22] or a stochastic optimization (SO) [23], [24] problem. The decision-maker requires the security constraints to hold either for all realizations of the uncertainty in RO or a high probability in SO. The latter yields less conservative solutions but requires the decision-maker to know the distribution information of the uncertainty. This lack of a suitable distribution that describes uncertainties on the one hand, and availability of historical forecasting data on the other, have promoted the application of distributionally robust optimization, although the distributionally robust individual chance-constrained dispatch models [25]–[31] are well-known, only a few of researches recently investigate the distributionally robust joint chance-constrained (DRJCC) dispatch model [32]–[34] to systematically and robustly ensure simultaneous security constraint satisfaction.

B. Related works

In terms of the collaborative management of ITD systems, the distributed optimal power flow (OPF), unit commitment (UC), or economic dispatch (ED) models for ITD systems are studied in [2]–[8]. However, the above models do not consider the significant uncertainty of renewable energy, which may cause the solution costly and unreliable. In [9], a traditional chance-constrained ED model for integrated transmission-district energy systems via distributed optimization is proposed. The distributed two stage RO models for ITD systems are proposed in [10], [22] and the distributed distributionally robust optimization model for ITD systems is proposed in

[11]. The general distributed optimization algorithms can be classified into three types: 1) the augmented Lagrangian based approaches such as the analytical target cascading (ATC) [2], [4], [11], [12], [20], [35], [36], alternating direction method of multipliers (ADMM) [10], [14]–[16], [22], [37] and auxiliary problem principle (APP) [13]; 2) the Karush–Kuhn–Tucker conditions-based approaches such as the heterogeneous decomposition (HD) algorithm [5], [6] and the optimality condition decomposition (OCD) algorithm [17]; and 3) the benders decomposition (BD) algorithm [7]. However, the HD, OCD, and BD algorithms are only suitable to solve the deterministic dispatch problem. When the number of DSs and uncertain renewable energy units is large, their communication burden is still heavy [10]. In contrast, for better scalability, the augmented Lagrangian based methods can be utilized. Among these methods, ADMM has shown its superiority in convergence property. To enhance the cost-effectiveness of interconnected power systems, ADMM has been adopted in ITD systems operation [10], [22], residential distributed generation coordination [37], multi-area ED [14], multi-area UC [15], and multi-area power flow [16] problems.

Most of these distributed optimization methods are developed based on the premise that workers can solve regional subproblems synchronously. However, the scale and complexity of subproblems are usually dependent on the system's physical configuration, and therefore are heterogeneous while requiring different amounts of computation time. To overcome these drawbacks of synchronous optimization, the recent works [38], [39] have generalized the synchronous alternating direction method of multipliers (ADMM) to an asynchronous version. In asynchronous ADMM, all workers can perform their local updates based on the latest available information from a subset of neighbors, which prevents the speedy workers from spending most of the time idling. In [38], [39], the asynchronous ADMM-based dispatch models for meshed power networks are proposed. In [40], the large-scale UC problem is solved using asynchronous ADMM. In addition to the asynchronous optimization model, the asynchronous power flow model for ITD systems and multi-area power systems is also presented in [18] and [19].

In terms of the distributionally robust optimization models, the main idea is to incorporate the available probability distribution information into an ambiguity set to characterize the true probability distribution of uncertain parameters. Accordingly, it eliminates the inherent dependence of SO on exact probability distributions and reduces the solution's conservatism of RO. The existing distributionally robust optimization models usually adopt two main types of ambiguity sets, moment-based and metric-based ambiguity sets. The first accommodates moment information of probability distributions, while the second specifies the closeness of probability distributions to an empirical distribution through a given statistical distance metric. The moment-based ambiguity sets comprise all distributions with an identical mean and covariance as the uncertain parameters [25], [29]–[32], [41]. The moment-based ambiguity set is relatively easier to manage than the metric-based ambiguity set, and can provide better tractability features [42]. However, this requires inferring the mean and covariance of uncertain parameters from the empirical data in order to construct the ambiguity set. The moment-based method only characterizes the ambiguity set by the first two moments of uncertain parameters, which cannot guarantee any convergence properties for the unknown distribution to the true distribution [43]. The probability distribution function contains more information than the moments and is not fully utilized in the moment-based method, which may lead to over-conservative decisions. In contrast, the Wasserstein-metric-based distributionally robust optimization method has recently gained much attention in power systems applications, including the OPF [26], [27], the UC [28], and the ED [33], [34] problems. The definition of ambiguity sets via the Wasserstein distance, directly utilizing the observed samples, brings several attractive properties in terms of finite sample guarantees, tractable reformulations, and asymptotic consistency. In power systems, the Wasserstein-based ambiguity sets also enable power system operators to control the conservativeness of the solution, thus ensuring flexibility in the power system operation. However, the aforementioned distributionally robust optimization models primarily focus on individual chance constraints.

Compared to individual chance constraints, the joint feature of distributionally robust chance constraints can provide a stronger guarantee on overall power systems security by enforcing the simultaneous satisfaction of multiple safety constraints with high probability. Needless to say, the distributionally robust joint chance-constrained problems are more challenging than individual chance-constrained problems. For instance, with only right-hand side disturbances, an individual chance-constrained problem can be transformed to an equivalent linearly constrained problem. In contrast, with only right-hand side disturbances, a joint chance-constrained problem is known to be convex only when the distributions are log-concave [44]. Thus, the distributionally robust joint chance constraints are less computationally tractable [34]. Fortunately, tractable conservative approximations can be obtained by using Bonferroni's inequality to decompose the joint chance constraint into several individual chance constraints [45], [46]. Recently, a few up to date literature focusing on the DC power flow based DRJCC dispatch model for transmission system are presented [32]–[34]. In [33], a Wasserstein metric-based DRJCC look-ahead dispatch model based on the Bonferroni approximation is discussed. And a further inner approximate robust bound was proposed, which was illustrated that it can achieve a better numerical scalability. In [34], two tractable approximations for Wasserstein metric-based DRJCC dispatch are proposed. A moment-based DRJCC OPF model in [32] is recently proposed based on the optimized Bonferroni approximation. Even so, the distributionally robust joint chance-constrained model applied in power systems is still limited due to the huge challenge in addressing the joint chance constraints.

C. Contributions and Organization

This paper concerns the distributionally robust dispatch problem for ITD systems via distributed optimization. A fully

decentralized asynchronous optimization scheme is proposed, which is particularly attractive for the co-optimization between the TSO and DSOs. In addition, the data-driven Wasserstein-distance based distributionally robust joint chance-constrained dispatch models for the transmission system and the distribution systems are investigated to handle renewable energy uncertainties. The contributions are summarized as follows:

1) Although [38], [39] have investigated the asynchronous ADMM-based decentralized dispatch model for power systems, they focused on the meshed networks. Considering the star topology of ITD systems and the subsystem heterogeneity, we propose a tailored asynchronous ADMM-based fully decentralized scheme to better adapt to the ITD systems. Compared to the classical synchronous ADMM [10], [14]–[16], [22], [37], the proposed asynchronous scheme only requires local communication and allows each subsystem operator to perform local updates with information from a subset of, but not all, neighbors. In particular, the central TSO's update only requires a subset of DSOs' information. Each DSO and TSO can thus, operate their local systems independently and asynchronously. Numerical results illustrate that the proposed asynchronous scheme for ITD systems has good scalability. Compared with the synchronous scheme, the tailored asynchronous scheme can reduce idle time and improve computational efficiency.
2) A DRJCC extension of asynchronous ADMM for ITD systems is proposed. Different from most of the distributionally robust individual chance-constrained dispatch [25]–[31], this paper focuses on the Wasserstein-distance-based distributionally robust joint chance-constrained dispatch. In order to facilitate the further reformulation of joint chance constraints, the DRJCC model based on the linearized DistFlow is proposed to describe the AC power flows and ensure the nodal voltage security in the distribution system under uncertainties, and the DRJCC model based on the DC power flow is proposed for the transmission system. Inspired by [34], a combined Bonferroni and CVaR approximation is presented to transfer the DRJCC models into the tractable conic formulations. A detailed empirical study on the out-of-sample performance reveals that the out-of-sample cost can attain a distinct minimum at a critical Wasserstein radius. Moreover, the critical radius gradually decreases with the increase of sample size, which shows that a sophisticated system operator who acknowledges the presence of ambiguity can reduce the out-of-sample cost.

The rest of this paper is organized as follows: Section II presents the separable formulation of ITD systems. Section III addresses the tractable approximation. Section IV proposes the asynchronous decentralized solution procedure. Case study and conclusion are presented in Section V and VI.

D. Preliminaries

Throughout this paper a couple of existing results from the field of distributionally robust optimization (DRO) [41], [47], in particular the data-driven approach [42], are used. The DRO problem can be written into a generic form

$$\min_{x\in\mathbb{X}} \max_{\mathbb{P}\in\mathcal{D}} \ \mathbb{E}_{\mathbb{P}}\left\{f(x,\xi)\right\} = \int_{\Omega} f(x,\xi)\mathbb{P}(d\xi) \tag{1}$$

with feasible set $\mathbb{X}\subseteq\mathbb{R}^n$, uncertainty set $\Omega\subseteq\mathbb{R}^m$, ambiguity set $\mathcal{D}$ and objective function $f:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$.

In general, as the distribution $\mathbb{P}$ is not precisely known, Problem (1) restricts it in the ambiguity set $\mathcal{D}$, which defines a family of distributions supported on Ω. Fortunately, in practice, we always have access to a finite number of samples $\widehat{\Xi} = \{\widehat{\xi}_1,..,\widehat{\xi}_N\} \subseteq \Omega$, which are assumed to be drawn independently from $\mathbb{P}$. Then, we can generate the discrete empirical distribution $\hat{\mathbb{P}}$ based on the sampling set $\widehat{\Xi}$. In order to construct an ambiguity set, we need the Wasserstein metric defined below.

Definition 1 (Wasserstein metric) *For two given distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ on $\mathbb{R}^n$, the type-1 Wasserstein distance is defined by*

$$\mathbb{W}(\mathbb{P}_1,\mathbb{P}_2) = \min_{\Pi}\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n} \|\boldsymbol{\xi}_1-\boldsymbol{\xi}_2\|\,\Pi(d\boldsymbol{\xi}_1,d\boldsymbol{\xi}_2)\right\}, \tag{2}$$

where Π is a joint distribution of $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ on $\mathbb{R}^n\times\mathbb{R}^n$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively.

According to Definition 1, one can denote the ambiguity set

$$\mathcal{D} := \left\{\mathbb{P}\in\mathcal{P}(\Omega) : \mathbb{W}(\mathbb{P},\hat{\mathbb{P}})\le\theta\right\} \tag{3}$$

where $\mathcal{P}(\Omega)$ is the set of all distributions on Ω and $\theta>0$. Here, the main idea of $\mathcal{D}$ is that, for a judiciously chosen radius θ, the ambiguity set $\mathcal{D}$ contains the unknown true distribution with high confidence.

II. Separable Formulation of ITD Systems

This section describes the typical structure of ITD systems and then presents the distributionally robust chance-constrained dispatch model for DS and TS, respectively.

A. Structure of ITD Systems

As shown in Fig. 1, a typical ITD systems that incorporates three distribution grids is connected to one transmission grid through their individual root nodes. In practice, the ITD systems allow the design of a decentralized decision-making scheme, in which each grid is operated locally while cooperating with its neighbors by sharing a limited set of data. As a result, proprietary data such as operation states and topological information can be kept confidential in each regional grid.

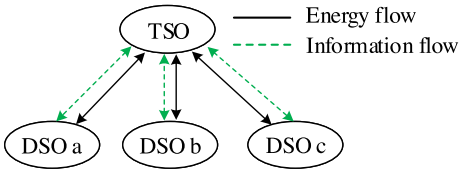

Fig. 1. Structure of decentralized operation for ITD systems

In particular, we consider a transmission system denoted by the tuple $(\mathcal{TN}, \mathcal{TL}, \mathcal{TG}, \mathcal{TW}, \mathcal{TD})$. Here, $\mathcal{TN}$ denotes the set of transmission nodes in TS, $\mathcal{TL} \subseteq \mathcal{TN} \times \mathcal{TN}$ the set of transmission lines in TS, $\mathcal{TG} \subseteq \mathcal{TN}$ the set of controllable units in TS, $\mathcal{TW} \subseteq \mathcal{TN}$ the set of renewable energies in TS, and $\mathcal{TD} \subseteq \mathcal{TN}$ the set of boundary nodes connecting to the root node of DS. For each DS $k \in \mathcal{K}$ with index set $\mathcal{K}$ of all DS, we consider a radial network denoted by the tuple $(\mathcal{DN}_k, \mathcal{DL}_k, \mathcal{DG}_k, \mathcal{DW}_k)$. Here, $\mathcal{DG}_k \subseteq \mathcal{DN}_k$ denotes the set of controllable units in DS k, $\mathcal{DW}_k \subseteq \mathcal{DN}_k$ the set of renewable energies in DS k, $\mathcal{DN}_k$ the set of non-root distribution nodes in DS k. The root node is indexed as 0 which connects the DS and TS. Each node is associated with a parent node $\mathcal{A}_i$ and a set of children nodes $\mathcal{C}_i$. Since the distribution network is radial, it is $|\mathcal{A}_i| = 1, i \in \mathcal{DN}_k$ and all distribution lines $i \in \mathcal{DL}_k$ are indexed by $\mathcal{DN}_k$, where i is the index of the downstream node of distribution line i.

The following assumptions are made in modeling the DR-JCC dispatch problem of ITD systems.

1) The subsystem operator has access to a finite number of i.i.d. (independend and identically distributed) samples, which are locally drawn from an underlying distribution.
2) In accordance with the distributionally robust studies of [26]–[28], [32]–[34], [48], the wind power curtailment is not considered.
3) The DRJCC model for distribution system does not impose distribution line flow limits, because real-life distribution systems are typically voltage-constrained and line flow limits can be disregarded [49].
4) The power loss of transmission grid and distribution grids can be ignored.

B. Uncertainty Modeling for the Distribution System

In the presence of renewable energy forecast error, the controllable units in each DS are assumed to adjust their power based on the following linear decision rules (LDRs) to guarantee that the forecasting errors of renewable energy in each DS are fully mitigated.

$$P_i^{G,D} = \hat{P}_i^{G,D} - \alpha_i^D \sum_{j \in \mathcal{DW}_k} \xi_j^{P,D} \ , \ i \in \mathcal{DG}_k, k \in \mathcal{K} \quad (4a)$$

$$Q_i^{G,D} = \hat{Q}_i^{G,D} - \alpha_i^D \sum_{j \in \mathcal{DW}_k} \xi_j^{Q,D} \ , \ i \in \mathcal{DG}_k, k \in \mathcal{K} \quad (4b)$$

$$\sum_{i \in \mathcal{DG}_k} \alpha_i^D = 1 \quad \text{with} \quad \alpha_i^D \in [0,1] \ , \ k \in \mathcal{K} \quad (4c)$$

where the active power forecast errors can also cause fluctuations of reactive power, the reactive power forecast errors can be assumed to be proportional [11], i.e., $\xi_i^{Q,D} = \sqrt{(1-\varphi^2)/\varphi^2}\xi_i^{P,D}$.

The DC power flow based distributionally robust models have been proven to effectively trade-off the likelihood of constraint violations and the security cost to avoid constraints violations [32]–[34]. However, the distributed energy resources in distribution system mainly complicate the voltage regulation, and therefore the DC power flow model is not technically suitable since it parametrizes voltage magnitudes at rated values. As the distribution grid is a radial topology, each distribution line has to carry the complete net load of all its downstream nodes. The linearized DistFlow [50] approximation for AC power flows is used here, and the distribution line flow affected by the uncertain renewable energy injections is derived as

$$P_i = \hat{P}_i - \boldsymbol{B}_{i*} \left(\boldsymbol{\xi}^{P,D} - \boldsymbol{\alpha}^D \sum_{j \in \mathcal{DN}_k} \xi_j^{P,D} \right) \quad (5a)$$

$$Q_i = \hat{Q}_i - \boldsymbol{B}_{i*} \left(\boldsymbol{\xi}^{Q,D} - \boldsymbol{\alpha}^D \sum_{j \in \mathcal{DN}_k} \xi_j^{Q,D} \right) \quad (5b)$$

for all $i \in \mathcal{DL}_k$ and $k \in \mathcal{K}$. Here $\boldsymbol{\xi}^{P,D}$ and $\boldsymbol{\xi}^{Q,D}$ denote the vector of active and reactive power forecast errors of all non-root nodes in DS, $\boldsymbol{\alpha}^D$ denotes the vector of participation factors of all non-root nodes in DS. If node i has no renewable energy injection, then $\xi_i^{P,D} = \xi_i^{Q,D} = 0$. If node i has no controllable units, then $\alpha_i^D = 0$. For each distribution system k, $\boldsymbol{B}_{i*}$ denotes the i-th row of $\boldsymbol{B} \in \mathbb{R}^{|\mathcal{DL}_k| \times |\mathcal{DN}_k|}$ with elements $b_{(ij)} = 1$ if line i is part of the path from root to bus j and $b_{(ij)} = 0$, otherwise.

Accordingly, the nodal voltage magnitude affected by the uncertain renewable energy injections can be derived as

$$\begin{aligned} V_i &= V_{\mathcal{A}_i} - (r_i P_i + x_i Q_i)/V_0 \\ &= \hat{V}_i + \boldsymbol{B}_{*i}^\top \Bigg[\boldsymbol{RB}(\boldsymbol{\xi}^{P,D} - \boldsymbol{\alpha}^D \sum_{j \in \mathcal{DN}_k} \xi_j^{P,D}) \\ &\qquad + \boldsymbol{XB}(\boldsymbol{\xi}^{Q,D} - \boldsymbol{\alpha}^D \sum_{j \in \mathcal{DN}_k} \xi_j^{Q,D}) \Bigg] / V_0 \end{aligned} \quad (6)$$

for all $i \in \mathcal{DN}_k$, $k \in \mathcal{K}$. Here $\boldsymbol{R}$ and $\boldsymbol{X}$ are $|\mathcal{DL}_k| \times |\mathcal{DL}_k|$ matrices with diagonal entries consisting of the line resistances and reactances respectively: $R_{(ii)} = r_i$, $R_{(ij,i \neq j)} = 0$, $\boldsymbol{X}$ in analogy.

C. DRJCC Dispatch Model for the Distribution System

The DRJCC model for each distribution system k (omit the subscript k) is formulated as

$$\begin{aligned} \min \max_{\mathbb{P} \in \mathcal{D}^D} \ \mathbb{E}_{\mathbb{P}} \Bigg\{ & \sum_{i \in \mathcal{DG}} \left[c_{i1}^D (P_i^{G,D})^2 + c_{i2}^D P_i^{G,D} + c_{i3}^D \right] \\ & + \sum_{i \in \mathcal{DG}} c_i^{r,D} \left(u_i^D + d_i^D \right) \Bigg\} \end{aligned} \quad (7)$$

subject to

$$\hat{P}_i = \sum_{j \in \mathcal{C}_i} \hat{P}_j - \hat{P}_i^{G,D} - \hat{P}_i^{W,D} + P_i^{L,D}, \ i \in \mathcal{DN} \quad (8a)$$

$$\hat{Q}_i = \sum_{j \in \mathcal{C}_i} \hat{Q}_j - \hat{Q}_i^{G,D} - \hat{Q}_i^{W,D} + Q_i^{L,D}, \ i \in \mathcal{DN} \quad (8b)$$

$$\hat{V}_i = \hat{V}_{\mathcal{A}_i} - \left(r_i \hat{P}_i + x_i \hat{Q}_i \right) / V_0, \ i \in \mathcal{DN} \quad (8c)$$

$$\hat{P}_1 = P^{Tie,D}, \ \hat{Q}_1 = Q^{Tie,D} \quad (8d)$$

$$\underline{P}_i^{G,D} \leq \hat{P}_i^{G,D} - d_i^D, \ 0 \leq d_i^D \leq \overline{d}_i^D \ , \ i \in \mathcal{DG} \quad (8e)$$

$$\hat{P}_i^{G,D} + u_i^D \leq \overline{P}_i^{G,D}, \ \ 0 \leq u_i^D \leq \overline{u}_i^D, \ i \in \mathcal{DG} \quad (8f)$$

$$\min_{\mathbb{P}\in\mathcal{D}^D} \mathbb{P}\left[\underline{V}_i \le V_i \le \overline{V}_i, \forall i \in \mathcal{DN}\right] \ge 1-\epsilon_1^D \tag{8g}$$

$$\min_{\mathbb{P}\in\mathcal{D}^D} \mathbb{P}\left[-d_i^D \le -\alpha_i^D \sum_{j\in\mathcal{DW}} \xi_j^{P,D} \le u_i^D, \forall i \in \mathcal{DG}\right] \ge 1-\epsilon_2^D \tag{8h}$$

$$\min_{\mathbb{P}\in\mathcal{D}^D} \mathbb{P}\left[\underline{Q}_i^{G,D} \le Q_i^{G,D} \le \overline{Q}_i^{G,D}, \forall i \in \mathcal{DG}\right] \ge 1-\epsilon_3^D \tag{8i}$$

$$\text{(4b), (6)} \tag{8j}$$

The objective function (7) aims to find the optimal decisions for DS that minimizes the worst-case expected production costs. (8a)-(8d) denote the linearized DistFlow equation for the forecast renewable energy. (8e) and (8f) guarantee the reserve availability considering the output limits of controllable units in each DS. Chance constraints (8g)-(8i) ensure that even under the worst case distribution, the constraints of nodal voltage, adjustment reserve capacity of controllable units, and reactive power of controllable units are still satisfied with the prescribed reliability.

D. DRJCC Dispatch Model for the Transmission System

Similar to the DS, the controllable units in TS are also assumed to adjust their power based on LDRs to guarantee that the forecasting errors of renewable energy in TS are fully mitigated.

$$P_i^{G,T} = \hat{P}_i^{G,T} - \alpha_i^T \sum_{j\in\mathcal{TW}} \xi_j^T,\ i \in \mathcal{TG} \tag{9a}$$

$$\sum_{i\in\mathcal{TG}} \alpha_i^T = 1 \quad \text{with} \quad \alpha_i^T \in [0,1] \tag{9b}$$

Then, based on the standard DC power flow model, the DRJCC model for the transmission system is summarized as

$$\min \max_{\mathbb{P}\in\mathcal{D}^T} \ \mathbb{E}_{\mathbb{P}}\left\{\sum_{i\in\mathcal{TG}}\left[c_{i1}^T (P_i^{G,T})^2 + c_{i2}^T P_i^{G,T} + c_{i3}^T\right] + \sum_{i\in\mathcal{TG}} c_i^{r,T}\left(u_i^T + d_i^T\right)\right\} \tag{10}$$

subject to

$$\sum_{i\in\mathcal{TG}} \hat{P}_i^{G,T} + \sum_{i\in\mathcal{TW}} \hat{P}_i^{W,T} = \sum_{i\in\mathcal{TN}} P_i^{L,T} + \sum_{i\in\mathcal{TD}} P_i^{Tie,T} \tag{11a}$$

$$0 \le P_i^{Tie,T} \le \overline{P}^{Tie},\ i \in \mathcal{TD} \tag{11b}$$

$$\underline{P}_i^{G,T} \le \hat{P}_i^{G,T} - d_i^T,\ 0 \le d_i^T \le \overline{d}_i^T,\ i \in \mathcal{TG} \tag{11c}$$

$$\hat{P}_i^{G,T} + u_i^T \le \overline{P}_i^{G,T},\ \ 0 \le u_i^T \le \overline{u}_i^T,\ i \in \mathcal{TG} \tag{11d}$$

$$\min_{\mathbb{P}\in\mathcal{D}^T} \mathbb{P}\left[-d_i^T \le -\alpha_i^T \sum_{j\in\mathcal{TW}} \xi_j^T \le u_i^T, \forall i \in \mathcal{TG}\right] \ge 1-\epsilon_1^T \tag{11e}$$

$$\min_{\mathbb{P}\in\mathcal{D}^T} \mathbb{P}\left[-\overline{L}_l \le \sum_{j\in\mathcal{TW}} M_{jl}(\hat{P}_j^{W,T} + \xi_j^T) - \sum_{i\in\mathcal{TN}} M_{il} P_i^{L,T} + \sum_{g\in\mathcal{TG}} M_{gl} P_g^{G,T} - \sum_{k\in\mathcal{TD}} M_{kl} P_k^{Tie,T} \le \overline{L}_l,\ \forall l \in \mathcal{TL}\right] \ge 1-\epsilon_2^T \tag{11f}$$

$$\text{(9)} \tag{11g}$$

The objective function (10) aims to find the optimal decisions for the TS that minimizes the worst-case expected production costs. (11a) denotes the power balance. (11b) denotes the active power limit of tie-line. (11c) and (11d) guarantee the reserve availability considering the output limits of controllable units in TS. Chance constraints (11e)-(11f) ensure that even under the worst case distribution, the constraints of adjustment reserve capacity of controllable units, the transmission line flow are still satisfied with the prescribed reliability. Note that, as the DC power flow model is considered here, the reactive power of substations in the boundary nodes is assumed to be enough for the DSs.

Remark 1 *If the wind power curtailment is considered, the extra LDRs form of wind output limit similar with the controllable units is needed. Then a new chance constraint about wind power curtailment is added to the DRJCC model. How to consider the wind power curtailment is addressed in Appendix.*

E. Coupling of Transmission and Distribution System

Regional coupling constraints tend to guarantee the agreement on tie-line power exchange between TS and each DS. While implementing the distributed optimization, it is necessary that the output active power from TS should be equal to the input active power to DS. Hence, the regional coupling is given by equalities

$$P_i^{Tie,T} = P_i^{Tie,D},\ i \in \mathcal{TD} \tag{12}$$

III. Tractable Approximation

This section aims to transfer the proposed DRJCC model into a tractable form under some mild approximations. As both DSOs and the TSO are ignorant of $\mathbb{P}$, the Wasserstein metric based data-driven ambiguity set $\mathcal{D}$ defined by (3) is used. Here, we assume that a finite number of samples $\widehat{\xi}_i$, $i \in \mathbb{Z}_1^N$ drawn independently from $\mathbb{P}$ are accessible. Under this ambiguity set, tractable conic reformulations of the worst-case objective function and chance constraints in the DRJCC model are presented. Accordingly, we summarize the tractable approximation into a standard form of distributed optimization problem.

A. Reformulation of Objective Function

Both objective functions of the distribution system model (7) and transmission system model (10) can be regarded as a sum of a quadratic function about the fuel cost at the reference output base-points of controllable units, a linear function about the reserve adjustment cost, and a linear function to represent the incremental cost under uncertainties [27]. Throughout the rest of this section, we omit the superscript T

for notational simplification. Let us take the TS model as an example. First, we define an approximation of (10) by

$$\min \sum_{i\in\mathcal{TG}}\left[c_{i1}(\hat{P}_i^G)^2+c_{i2}\hat{P}_i^G+c_{i3}\right]+\sum_{i\in\mathcal{TG}}c_i^r\left(u_i+d_i\right) + \max_{\mathbb{P}\in\mathcal{D}}\mathbb{E}_{\mathbb{P}}\left\{\sum_{i\in\mathcal{TG}}\left(-c_{i0}\alpha_i\sum_{j\in\mathcal{TW}}\xi_j\right)\right\} \quad (13)$$

with $c_{i0}=2c_{i1}P_i^{G*}+c_{i2}$, where P_i^{G*} is the optimal output obtained through a deterministic model under forecasting renewable energy. In this way, the worst-case expectation term is separated from the deterministic term, facilitating further reformulation. Besides, the quadratic function term is approximated by a piecewise linear function to improve computational efficiency.

In order to reformulate the second worse-case expectation term in (13), we use notations $\boldsymbol{c}_0,\boldsymbol{\alpha}\in\mathbb{R}^{|\mathcal{TG}|}$ and $\boldsymbol{\xi}\in\mathbb{R}^{|\mathcal{TW}|}$ to stack c_{i0}, α_i over $i\in\mathcal{TG}$ and ξ_j over $j\in\mathcal{TW}$, respectively. Moreover, the support of $\boldsymbol{\xi}$ is given by a polytope $\Xi=\{\boldsymbol{\xi}\in\mathbb{R}^{|\mathcal{TW}|}:\boldsymbol{H}\boldsymbol{\xi}\le\boldsymbol{h}\}$. In this paper, we consider $\boldsymbol{H}$ and $\boldsymbol{h}$ denoted by $\boldsymbol{H}=[\boldsymbol{I}\ \ -\boldsymbol{I}]^\top$ and $\boldsymbol{h}=[(\overline{\boldsymbol{P}}^W-\hat{\boldsymbol{P}}^W)^\top\ (\hat{\boldsymbol{P}}^W-\underline{\boldsymbol{P}}^W)^\top]^\top$ with $\boldsymbol{I}$ identity matrix and $\underline{\boldsymbol{P}}^W/\overline{\boldsymbol{P}}^W$ the vector of lower/upper limit of wind outputs. Then, we define its compact form

$$\max_{\mathbb{P}\in\mathcal{D}}\ \mathbb{E}_{\mathbb{P}}\left\{\sum_{i\in\mathcal{TG}}\left(-c_{i0}\alpha_i\sum_{j\in\mathcal{TW}}\xi_j\right)\right\} = \max_{\mathbb{P}\in\mathcal{D}}\ \mathbb{E}_{\mathbb{P}}\left\{-\boldsymbol{c}_0^\top\boldsymbol{\alpha}\boldsymbol{1}^\top\boldsymbol{\xi}\right\} \quad (14)$$

with $\boldsymbol{1}=[1,1,...,1]^\top\in\mathbb{R}^{|\mathcal{TW}|}$ and according to [51, Corollary 5.1], the worst-case expectation term (14) is equivalent to the conic program

$$\min_{\eta^o\ge0,\boldsymbol{\beta}^o,\boldsymbol{\tau}^o}\quad \eta^o\theta+\frac{1}{N}\sum_{i=1}^N\beta_i^o \quad (15a)$$

$$\text{s.t.}\quad -\boldsymbol{c}_0^\top\boldsymbol{\alpha}\boldsymbol{1}^\top\widehat{\boldsymbol{\xi}}_i+{\boldsymbol{\tau}_i^o}^\top(\boldsymbol{h}-\boldsymbol{H}\widehat{\boldsymbol{\xi}}_i)\le\beta_i^o, i\in\mathbb{Z}_1^N \quad (15b)$$

$$\|\boldsymbol{H}^\top\boldsymbol{\tau}_i^o+\boldsymbol{1}\boldsymbol{\alpha}^\top\boldsymbol{c}_0\|_\infty\le\eta^o, i\in\mathbb{Z}_1^N \quad (15c)$$

for a given radius $\theta>0$. As a result, the worst-case expectation objective (10) can be approximately reformulated into a tractable conic form by substituting (15) into (13).

B. Reformulation of Joint Chance Constraints

Joint chance constraints (8g)-(8i) and (11e)-(11f) are non-convex and very challenging to derive tractable equivalent reformulations due to their implicit form. To this end, the combined Bonferroni and CVaR approximation is proposed to reformulate the joint chance constraints for transmission and distribution grids. Let us take the TS model as an example, and use notation $\boldsymbol{z}$ to stack variables $(\hat{P}_i^G,\hat{P}_i^{Tie},u_i,d_i)$ over i. The generic formulation of joint chance constraints for TS can be given as follows

$$\min_{\mathbb{P}\in\mathcal{D}}\ \mathbb{P}\left[\boldsymbol{A}(\boldsymbol{\alpha})\boldsymbol{\xi}\le\boldsymbol{b}(\boldsymbol{z})\right]\ge1-\epsilon \quad (16)$$

where each element of $\boldsymbol{A}(\boldsymbol{\alpha})$ and $\boldsymbol{b}(\boldsymbol{z})$ is given by an affine map of $\boldsymbol{\alpha}$ and $\boldsymbol{z}$, respectively. Let us assume that there are M affine inequalities involved in the joint chance constraints, we use notation $\boldsymbol{A}(\boldsymbol{\alpha})=\left[\boldsymbol{a}_1(\boldsymbol{\alpha})\cdots\boldsymbol{a}_M(\boldsymbol{\alpha})\right]^\top$ and $\boldsymbol{b}(\boldsymbol{z})=\left[\boldsymbol{b}_1(\boldsymbol{z})\cdots\boldsymbol{b}_M(\boldsymbol{z})\right]^\top$. The joint chance constraint is thus, equivalent to

$$\min_{\mathbb{P}\in\mathcal{D}}\ \mathbb{P}\left[\boldsymbol{a}_m(\boldsymbol{\alpha})^\top\boldsymbol{\xi}\le\boldsymbol{b}_m(\boldsymbol{z}),\forall\, m\in\mathbb{Z}_1^M\right]\ge1-\epsilon. \quad (17)$$

Given a set of risk parameters $\epsilon_m\ge0$, $m\le M$, with $\sum_{m=1}^M\epsilon_m=\epsilon$, we can exploit Bonferroni's inequality to split the original joint chance constraints up into a family of M simpler but more conservative individual chance constraints. As discussed in [34], this amounts to approximating the feasible set by

$$\Omega:=\left\{(\boldsymbol{\alpha},\boldsymbol{z})\ \middle|\ \begin{array}{l}\forall\, m\in\mathbb{Z}_1^M,\\ \min\limits_{\mathbb{P}\in\mathcal{D}}\mathbb{P}\left[\boldsymbol{a}_m(\boldsymbol{\alpha})^\top\boldsymbol{\xi}\le\boldsymbol{b}_m(\boldsymbol{z})\right]\ge1-\epsilon_m,\end{array}\right\}. \quad (18)$$

However, optimizing over Ω is still hard. We then define the CVaR at level $\epsilon\in(0,1)$ for a given measurable loss function $\ell(\boldsymbol{\xi})$ by

$$\mathbb{P}\text{-}\,\mathrm{CVaR}_\epsilon(\ell(\boldsymbol{\xi})):=\inf_\delta\left\{\delta+\frac{1}{\epsilon}\mathbb{E}\left[\ell(\boldsymbol{\xi})-\delta\right]_+\right\}$$

with $[\cdot]_+=\max\{0,\cdot\}$. Thus, the CVaR based approximation of Ω is given by

$$\Phi:=\left\{(\boldsymbol{\alpha},\boldsymbol{z})\ \middle|\ \begin{array}{l}\forall\, m\in\mathbb{Z}_1^M,\\ \max\limits_{\mathbb{P}\in\mathcal{D}}\mathbb{P}\text{-}\,\mathrm{CVaR}_{\epsilon_m}\left[\boldsymbol{a}_m(\boldsymbol{\alpha})^\top\boldsymbol{\xi}-\boldsymbol{b}_m(\boldsymbol{z})\right]\le0\end{array}\right\},$$

Remark 2 *In [52], Φ has been shown that it is the best convex inner approximation, i.e., $\Phi\subseteq\Omega$, in a sense made precise. Moreover, if $\epsilon_m\le N^{-1}$ holds for all $m\in\mathbb{Z}_1^M$, we have $\Phi=\Omega$. A detailed proof refers to [53, Corollary 2].*

As [34, Proposition 1] discussed, Φ can be further derived into the explicit conic form

$$\Phi:=\left\{(\boldsymbol{\alpha},\boldsymbol{z})\ \middle|\ \begin{array}{l}\exists\,\boldsymbol{\gamma}\in\mathbb{R}^M,\ \boldsymbol{\eta}\in\mathbb{R}^M,\ \boldsymbol{\beta}\in\mathbb{R}^{N\times M},\\ \eta_m\theta+\frac{1}{N}\sum_{i=1}^N\beta_{im}\le0,\ m\in\mathbb{Z}_1^M\\ \forall\, i\in\mathbb{Z}_1^N,\ \ m\in\mathbb{Z}_1^M:\\ \left\{\begin{array}{l}\gamma_m\le\beta_{im},\ \boldsymbol{\tau}_{i,m}\ge0\\ \|\epsilon_m\boldsymbol{H}^\top\boldsymbol{\tau}_{im}-\boldsymbol{a}_m(\boldsymbol{\alpha})\|_\infty\le\epsilon_m\eta_m\\ \boldsymbol{a}_m(\boldsymbol{\alpha})^\top\widehat{\boldsymbol{\xi}}_i-\boldsymbol{b}_m(\boldsymbol{z})+(\epsilon_m-1)\gamma_m\\ \qquad+\epsilon_m\boldsymbol{\tau}_{im}^\top(\boldsymbol{h}-\boldsymbol{H}\widehat{\boldsymbol{\xi}}_i)\le\epsilon_m\beta_{im}\end{array}\right.\end{array}\right\}. \quad (19)$$

Here, $\beta_{i,m}$ is the element of matrix $\boldsymbol{\beta}$ at i-th row and m-th column.

Under the combined Bonferroni and CVaR approximation, all the joint chance constraints in the DRJCC model of

the transmission system have been replaced by the conic representation. Then, the resulting DRJCC model has been transformed into a single tractable conic program that can be solved efficiently by off-the-shelf solvers. The tractable approximation of DRJCC model for DS can be derived similarly.

Remark 3 *If the Wasserstein metric is defined in terms of the ℓ_1-norm or ℓ_∞-norm, the tractable conic program is reduced to a quadratic program with linear constraints [42].*

Remark 4 *The choice of the risk parameters ϵ_m affects the performance of the combined Bonferroni and CVaR approximation. As recommended in [52], we set $\epsilon_m = \epsilon/M$ for all $m \leq M$.*

C. Problem Formulation

Based on the Wasserstein metric 1, Section III-A and III-B have elaborated how to reformulate the worse-case objectives and chance constraints in the proposed DRJCC model discussed in Section II into a tractable form. Accordingly, the overall optimization problem is summarized as below

$$\min_{\bar{\boldsymbol{x}},\{\boldsymbol{x}_k\}_{k\in\mathcal{K}}} \quad \bar{F}(\bar{\boldsymbol{x}}) + \sum_{k\in\mathcal{K}} F_k(\boldsymbol{x}_k) \tag{20a}$$

$$\text{s.t.} \quad \boldsymbol{A}_k\bar{\boldsymbol{x}} = \boldsymbol{B}_k\boldsymbol{x}_k \quad | \ \boldsymbol{\lambda}_k \ , \ \ k \in \mathcal{K} \ , \tag{20b}$$

$$\bar{\boldsymbol{x}} \in \overline{\mathcal{X}} \ , \ \ \boldsymbol{x}_k \in \mathcal{X}_k \ , \ \ k \in \mathcal{K} \ . \tag{20c}$$

Here, $\bar{\boldsymbol{x}} \in \mathbb{R}^{d_T}$ and $\boldsymbol{x}_k \in \mathbb{R}^{d_{D_k}}$, $k \in \mathcal{K}$ stack the local decision variables of TS and each DS, respectively. Selection matrices $\boldsymbol{A}_k$ and $\boldsymbol{B}_k$, $k \in \mathcal{K}$ with elements either 0 or 1 are used to denote the coupling constraints (12) between the TS and DS while $\boldsymbol{\lambda}_k$ denotes the corresponding Lagrangian multipliers of constraint (20b). Moreover, we use compact notations $\overline{\mathcal{X}}$ and $\mathcal{X}_k$, $k \in \mathcal{K}$ to collect all deterministic constraints and the tractable reformulation of the chance constraints in (11) and (8), respectively.

The operation problem for the TSO and each DSO is an independent decision-making process, which is particularly suitable for distributed optimization. Only tie-line information is shared to minimize the data exchange between TSO and each DSO. Therefore, the overall problem (20) can be solved in a fully decentralized way to preserve regional information privacy and dispatch independence of respective system operators. Besides, due to the different network topology, scale and complexity of each subsystem, the computation of each regional grid may require different amounts of time. In the synchronous decentralized computation, all system operators need to wait for the slowest one to finish its computation or communication. This may lead to the under-utilization of both computation and communication resources as some regional grids remain idle for most of the time. Thus, an asynchronous decentralized algorithm is preferable where all system operators can perform their local updates based on the latest available information from a subset of but not all neighbors. Therefore, it can prevent the speedy system operators from spending most of the time idling.

IV. Asynchronous Decentralized Solution Procedure

In this section, we present a tailored asynchronous ADMM based decentralized algorithm for solving the resulting optimization problem (20), which can better adapt to the star topology of ITD systems. To this end, we first denote the augmented Lagrangian function of (20) by

$$\bar{F}(\bar{\boldsymbol{x}}) + \sum_{k\in\mathcal{K}} \Big(F_k(\boldsymbol{x}_k) + L_k(\boldsymbol{A}_k\bar{\boldsymbol{x}}, \boldsymbol{B}_k\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \Big)$$

with

$$L_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{z}^\top(\boldsymbol{x} - \boldsymbol{y}) + \frac{\rho}{2}\left\|\boldsymbol{x} - \boldsymbol{y}\right\|_2^2$$

such that the classical ADMM method for solving (20) is given by the iterations

$$\boldsymbol{x}_k^+ = \operatorname*{argmin}_{\boldsymbol{x}_k\in\mathcal{X}_k} F_k(\boldsymbol{x}_k) + L_k(\boldsymbol{A}_k\bar{\boldsymbol{x}}, \boldsymbol{B}_k\boldsymbol{x}_k, \boldsymbol{\lambda}_k), \ k \in \mathcal{K} \tag{21a}$$

$$\boldsymbol{\lambda}_k^+ = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}_k\bar{\boldsymbol{x}} - \boldsymbol{B}_k\boldsymbol{x}_k) \ , \ \ k \in \mathcal{K} \tag{21b}$$

$$\bar{\boldsymbol{x}}^+ = \operatorname*{argmin}_{\bar{\boldsymbol{x}}\in\overline{\mathcal{X}}} \bar{F}(\bar{\boldsymbol{x}}) + \sum_{k\in\mathcal{K}} L_k(\boldsymbol{A}_k\bar{\boldsymbol{x}}, \boldsymbol{B}_k\boldsymbol{x}_k^+, \boldsymbol{\lambda}_k^+) \ . \tag{21c}$$

Although the update of $(\boldsymbol{x}_k, \boldsymbol{\lambda}_k)$ can be parallelized over $k \in \mathcal{K}$, Step (21c) has to wait Step (21a) and Step (21b) finished for all $k \in \mathcal{K}$. In order to overcome this problem, we introduce a consensus variable for parallelizing Step (21c) by

$$\boldsymbol{A}_k\bar{\boldsymbol{x}} = \boldsymbol{y}_k \ , \ \ \boldsymbol{B}_k\boldsymbol{x}_k = \boldsymbol{y}_k$$

and their associated pair of Lagrangian multipliers $(\boldsymbol{\lambda}_k^T, \boldsymbol{\lambda}_k^D)$ for all $k \in \mathcal{K}$. As a result, a variant of classical ADMM is yielded as follows:

$$\boldsymbol{x}_k^+ = \operatorname*{argmin}_{\boldsymbol{x}_k\in\mathcal{X}_k} F_k(\boldsymbol{x}_k) + L_k(\boldsymbol{B}_k\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k^D), \ k \in \mathcal{K} \tag{22a}$$

$$\bar{\boldsymbol{x}}^+ = \operatorname*{argmin}_{\bar{\boldsymbol{x}}\in\overline{\mathcal{X}}} \bar{F}(\bar{\boldsymbol{x}}) + \sum_{k\in\mathcal{K}} L_k(\boldsymbol{A}_k\bar{\boldsymbol{x}}, \boldsymbol{y}_k, \boldsymbol{\lambda}_k^T) \ , \tag{22b}$$

$$\boldsymbol{\lambda}_k^{D,+} = \boldsymbol{\lambda}_k^D + \rho(\boldsymbol{B}_k\boldsymbol{x}_k^+ - \boldsymbol{y}_k) \ , \ \ k \in \mathcal{K} \tag{22c}$$

$$\boldsymbol{\lambda}_k^{T,+} = \boldsymbol{\lambda}_k^T + \rho(\boldsymbol{A}_k\bar{\boldsymbol{x}}^+ - \boldsymbol{y}_k) \ , \ \ k \in \mathcal{K} \tag{22d}$$

$$\boldsymbol{y}_k^+ = \arg\min_{\boldsymbol{y}_k} \ L_k(\boldsymbol{B}_k\boldsymbol{x}_k^+, \boldsymbol{y}_k, \boldsymbol{\lambda}_k^{D,+}) \tag{22e}$$

$$+ L_k(\boldsymbol{A}_k\bar{\boldsymbol{x}}^+, \boldsymbol{y}_k, \boldsymbol{\lambda}_k^{T,+}), \ k \in \mathcal{K} \ .$$

It is clear that the update of $\boldsymbol{x}_k^+$ can be executed in parallel with the update $\bar{\boldsymbol{x}}^+$ in the tailored ADMM iterations (22). Moreover, as functions L_k are quadratic, we can work out the update $\boldsymbol{y}_k$ explicitly as

$$\boldsymbol{y}_k^+ = \boldsymbol{A}_k\bar{\boldsymbol{x}}^+ + \boldsymbol{B}_k\boldsymbol{x}_k^+ - \frac{\boldsymbol{\lambda}_k^{T,+} + \boldsymbol{\lambda}_k^{D,+}}{\rho} \ , \ \ k \in \mathcal{K} \ .$$

However, in practice, the computational time of solving (22a) and (22b) might be enormously different such that (22) still leads to an inefficiency on communication. In order to mitigate this disadvantage, we adopt the framework proposed in [38] to design an asynchronous ADMM approach.

Algorithm 1 and 2 outline an asynchronous ADMM variant from TS's perspective and each DS's perspective. Here, we use notation ℓ and ℓ_k, $k \in \mathcal{K}$ to denote the local iteration counter in TS and DS. Moreover, we assume that the communication

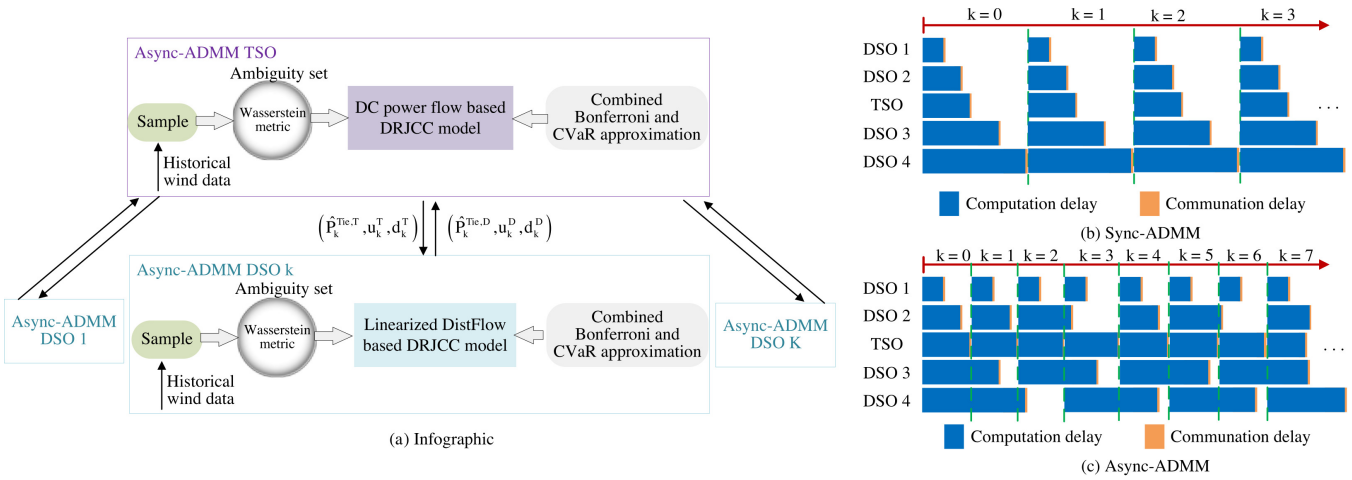

Fig. 2. The proposed distributed DRJCC optimization framework

Algorithm 1 Asynchronous ADMM in TS

Initialization:

- choose $\boldsymbol{\lambda}_k^{T,0}$, $k \in \mathcal{K}$ and set $\ell = 0$;
- solve TS problem $\bar{\boldsymbol{x}}^0 = \arg\min_{\bar{\boldsymbol{x}} \in \overline{\mathcal{X}}} \bar{F}(\bar{\boldsymbol{x}})$;
- send $(\boldsymbol{A}_k \bar{\boldsymbol{x}}^0, \boldsymbol{\lambda}_k^{T,0})$ to all DS $k \in \mathcal{K}$.

Repeat:

1) Wait until $\bar{k} \in \overline{\mathcal{K}}$ DS's $(\boldsymbol{B}_k \boldsymbol{x}_k^\ell, \boldsymbol{\lambda}_k^{D,\ell})$;
2) Evaluate

$$\bar{\boldsymbol{y}}_k^{\ell+1} = \begin{cases} \boldsymbol{A}_k \bar{\boldsymbol{x}}^\ell + \boldsymbol{B}_k \boldsymbol{x}_k^\ell - \frac{\boldsymbol{\lambda}_k^{T,\ell} + \boldsymbol{\lambda}_k^{D,\ell}}{\rho} & k \in \overline{\mathcal{K}}, \\ \boldsymbol{y}_k^\ell & k \in \mathcal{K} \backslash \overline{\mathcal{K}}. \end{cases}$$

3) Update primal and dual by

$$\bar{\boldsymbol{x}}^{\ell+1} = \underset{\bar{\boldsymbol{x}} \in \overline{\mathcal{X}}}{\operatorname{argmin}} \ \bar{F}(\bar{\boldsymbol{x}}) + \sum_{k \in \mathcal{K}} L_k(\boldsymbol{A}_k \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}_k^{\ell+1}, \boldsymbol{\lambda}_k^{T,\ell}) \ ,$$
$$\boldsymbol{\lambda}_k^{T,\ell+1} = \boldsymbol{\lambda}_k^T + \rho(\boldsymbol{A}_k \bar{\boldsymbol{x}}^{\ell+1} - \bar{\boldsymbol{y}}_k^{\ell+1}) \ , \quad k \in \mathcal{K} \ .$$

4) Set $\ell \leftarrow \ell + 1$ and send $(\boldsymbol{A}_k \bar{\boldsymbol{x}}^\ell, \boldsymbol{\lambda}_k^{T,\ell})$ to all DS $k \in \mathcal{K}$.

Algorithm 2 Asynchronous ADMM in DS k

Initialization:

- choose $\boldsymbol{\lambda}_k^{D,0}$, set $\ell_k = 0$;
- solve DS problem $\boldsymbol{x}_k^0 = \arg\min_{\boldsymbol{x}_k \in \mathcal{X}_k} F_k(\boldsymbol{x}_k)$;
- send $(\boldsymbol{B}_k \boldsymbol{x}_k^0, \boldsymbol{\lambda}_k^{D,0})$ to TS.

Repeat:

1) Wait $(\boldsymbol{A}_k \bar{\boldsymbol{x}}_k^{\ell_k}, \boldsymbol{\lambda}_k^{T,\ell_k})$ from TS;
2) Update $\boldsymbol{y}_k^{\ell_k+1}$ by evaluating

$$\boldsymbol{y}_k^{\ell_k+1} = \boldsymbol{A}_k \bar{\boldsymbol{x}}^{\ell_k} + \boldsymbol{B}_k \boldsymbol{x}_k^{\ell_k} - \frac{\boldsymbol{\lambda}_k^{T,\ell_k} + \boldsymbol{\lambda}_k^{D,\ell_k}}{\rho} \ . \tag{23}$$

3) Update primal and dual by

$$\boldsymbol{x}_k^{\ell_k+1} = \underset{\boldsymbol{x}_k \in \mathcal{X}_k}{\operatorname{argmin}} \ F_k(\boldsymbol{x}_k) + L_k(\boldsymbol{B}_k \boldsymbol{x}_k, \boldsymbol{y}_k^{\ell_k+1}, \boldsymbol{\lambda}_k^{D,\ell_k})$$
$$\boldsymbol{\lambda}_k^{D,\ell_k+1} = \boldsymbol{\lambda}_k^{D,\ell_k} + \rho(\boldsymbol{B}_k \boldsymbol{x}_k^{\ell_k+1} - \boldsymbol{y}_k^{\ell_k+1})$$

4) Set $\ell_k \leftarrow \ell_k + 1$ and send $(\boldsymbol{A}_k \boldsymbol{x}_k^{\ell_k}, \boldsymbol{\lambda}_k^{D,\ell_k})$ TS.

delay is bounded. This implies that all transfer information would eventually arrive at its destination.

The main idea of the proposed algorithm is to let TS update its associated primal and dual iterates with obtaining only $|\overline{\mathcal{K}}| \geq 1$ DS's information but not all of them, i.e., $\overline{\mathcal{K}} \subseteq \mathcal{K}$. In a result, TS updates its consensus variable $\bar{\boldsymbol{y}}_k$, $k \in \mathcal{K}$ by using limited DS's information. Notice that $\bar{\boldsymbol{y}}_k$ in Algorithm 1 might not be equal to $\boldsymbol{y}_k$ in Algorithm 2 as the update (23) only depends on the part of neighbors' information. Notice that only $|\overline{\mathcal{K}}|$ is fixed but the elements might be different from each iteration. In addition, how to choose $\overline{\mathcal{K}}$ does not affect the convergence guarantee that has been established in [39, Theorem 1].

The schematic of the holistic model framework is displayed in Fig. 2. Under the synchronous protocol, the speedy agents which might be TS or DS, would spend most of the time idling, and thus the parallel computational resources cannot be fully utilized. Instead, with the lock removed, the speedy agents can update their variables more frequently in the asynchronous protocol. On the flip side, the asynchronous one introduces delayed variable information and thereby requires a larger number of iterations to reach the same solution accuracy than its synchronous counterpart. This implies that none of the workers have to be synchronized with each other and does not need to wait for the slowest worker either. In the considering heterogeneous network, the local DSs and TS might have different computational powers, or the data sets can be non-uniformly distributed across the network. Thus, the agents can require different computational times in solving the local subproblems. Besides, the communication delays can also be different, e.g., due to probabilistic communication failures and message retransmission.

V. Numerical Results

In this section, we present numerical studies on two benchmarks to illustrate the effectiveness of the proposed scheme. In both cases, the risk parameters of chance constraints in TS and

each DS are all set to 5%. The constant power factor for all DSs is set to 0.75. The construction procedure of data-driven ambiguity set refers to [34] using the wind outputs data from southeastern Australia from 2012 to 2013 [54]. As discussed in Remark 3, we use ℓ_1-norm for the Wasserstein metric on $\mathbb{R}^W$ such that the DRJCC model is reduced to a quadratic program.

For the implementation of the asynchronous ADMM algorithm, the initial values of the coupling variables are set to zero while the initial guess of the Lagrangian multiplier is set to 0.02. Moreover, we set $|\overline{\mathcal{K}}| = 2$ in Case 1 and $|\overline{\mathcal{K}}| = 3$ in Case 2. The simulation is employed in `MATLAB R2021a` on an Intel Core i7-9700K, 3.6 GHz, 16 GB RAM PC with 8 cores. `Gurobi` 9.0 is used to solve each subproblem locally on each core. All regional subproblems in synchronous and asynchronous setting are solved by using the Matlab Parallel Computing Toolbox.

A. Test System T39D3

Case 1 is referred to as T39D3, where an modified IEEE 39-bus transmission system is connected with one modified IEEE 33-bus distribution system, one modified IEEE 69-bus distribution system, and one modified IEEE 85-bus distribution system located at transmission nodes 3, 10, and 17, respectively. The root node of each DS connects the TS. Accordingly, the three DSs are respectively named as DS 1, 2, 3 and the three tie-lines are respectively named as Tie-line 1, 2, 3. The TS has five wind farms located at transmission nodes 3, 6, 17, 20, and 25, respectively. DS 1 has three distributed wind generations located at distribution nodes 2, 3, and 6, respectively. DS 2 has three distributed wind generations located at distribution nodes 3, 4, and 8, respectively. DS 3 has four distributed wind generations located at distribution nodes 2, 5, 13, and 60, respectively. The network topology, controllable units parameters, wind outputs, load of TS and each DS, and other detailed data in Case 1 has been made available online [55].

The sample set is set to $N = 50$ for TS and each DS. Then, another $S = 1000$ test samples from the data source are carried out to test the out-of-sample performance, which is assessed by the following out-of-sample operation cost

$$\widehat{\mathcal{O}}^O = \boldsymbol{c}(\widehat{\boldsymbol{z}}) + \frac{1}{S}\sum_{i=1}^{S} \boldsymbol{c}_0^\top \widehat{\boldsymbol{\alpha}} \mathbf{1}^\top \widehat{\xi}_{N+i}$$

and the out-of-sample constraint violation probability

$$\widehat{\mathcal{O}}^C = \sum_{j\in\mathcal{J}} \left\{ \frac{1}{S}\sum_{i=1}^{S} \mathbb{I}_{\boldsymbol{A}^j(\widehat{\boldsymbol{\alpha}})\widehat{\xi}_{N+i} > \boldsymbol{b}^j(\widehat{\boldsymbol{z}})} \right\}$$

where $c(\widehat{\boldsymbol{z}})$ represents the deterministic term in (13), $\widehat{\boldsymbol{z}} = (\widehat{\boldsymbol{P}}^G, \boldsymbol{P}^{Tie}, \boldsymbol{u}, \boldsymbol{d})$ and $\widehat{\boldsymbol{\alpha}}$ denote the optimal value obtained from the sample set, set $\mathcal{J}$ denotes the index of the joint chance constraints.

1) Impact of Wasserstein Radius: The out-of-sample constraint violation probability for TS and each DS with different Wasserstein radii are shown in Fig.3. Different Wasserstein radii result in different out-of-sample performance. With the increasing of Wasserstein radius, the solution leads to a lower risk of constraint violation for TS and each DS. This is because the larger Wasserstein balls lead to higher robustness to sampling errors. This is expected as larger Wasserstein radii result in more conservative solutions. For sufficiently large values of radius, the proposed DRJCC model can guarantee that the out-of-sample constraint violation probabilities of all chance constraints are smaller than the selected risk parameters.

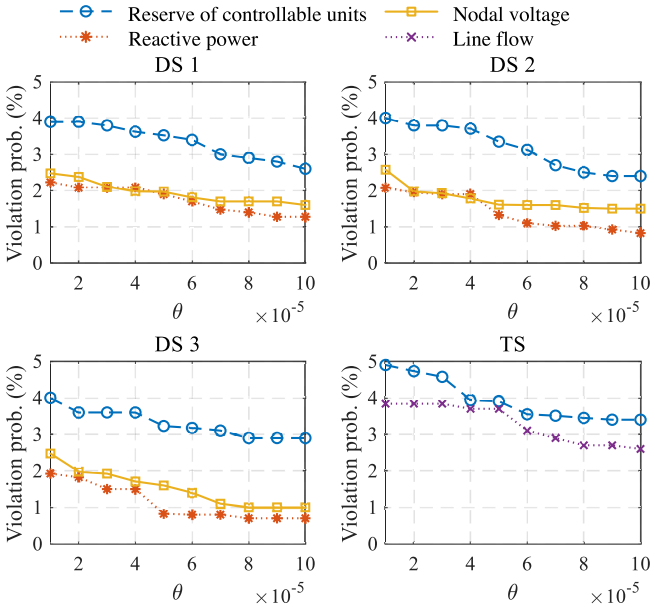

Fig. 3. Out-of-sample constraint violation probability

2) Comparison with stochastic & robust optimization: This section compares the proposed DRJCC model with radius $\theta = 0.0001$ to the SO-based and RO-based models. The RO-based model is constructed as a LDRs-based adjustable robust optimization model, which requires all chance constraints to be satisfied for all realizations in the support of uncertain wind farm outputs. The SO-based model is constructed as a sample average approximation model. The S test samples is carried out to compare the out-of-sample operation cost and constraint reliability of TS and each DS, shown in Fig. 4.

The total operation cost of the DRJCC model is larger than SO while smaller than RO model. The DRJCC and RO models can ensure lower constraint violation probability than the required risk level, while the SO model cannot guarantee the required system reliability level. This is because the SO assumes uncertain variables follow the empirical probability distribution generated from N samples, so the solution is sensitive to the perturbation in the true distribution of uncertainties. Since RO completely ignores the probabilistic information, the RO model is the most conservative solution while SO yields the most aggressive solution. We can conclude that the proposed DRJCC model can flexibly balance the system economic and constraint reliability in comparison with RO and SO models.

3) Impact of Risk Parameters: The risk parameter ϵ reflects the system operator's risk attitude. Table I shows the out-of-sample total operation cost with the same radius $\theta = 0.0001$ under different risk parameters. It can be observed that a larger

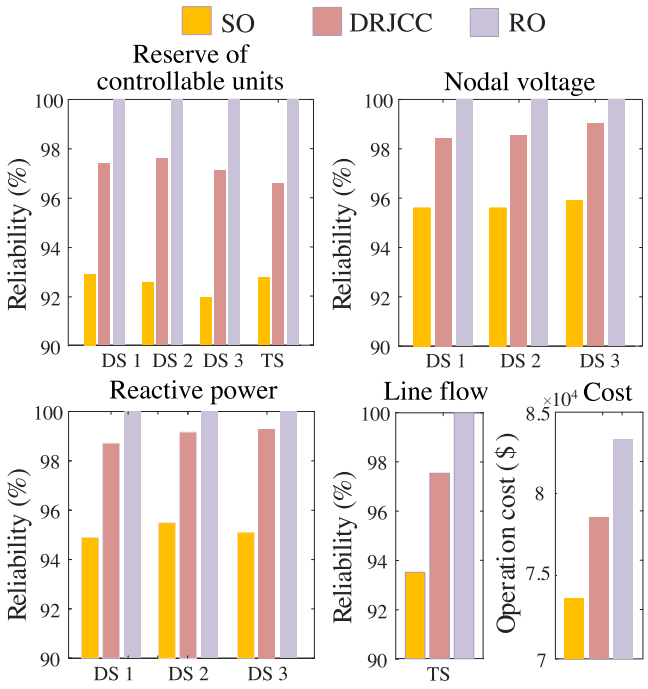

Fig. 4. Out-of-sample constraint reliability and operation cost

TABLE I
COMPARISON OF DIFFERENT RISK PARAMETERS

ϵ	2%	3%	5%	10%	15%	20%
Operation cost ($)	79528	79069	78628	78364	78051	77824

ϵ allows to tolerate some constraint violation probability, in a result, the total operation cost will decrease. We can conclude that the choice of risk parameter also has an important impact on the conservatism of solution. In fact, the system operators can reduce the total operation costs by increasing the risk parameters as long as they can tolerate a relatively high operational risk.

4) Convergence Performance and Solution Quality: The tailored asynchronous ADMM scheme is compared with the synchronous ADMM and the traditional centralized method for the DRJCC model of ITD systems with radius $\theta = 0.0001$ and risk parameter $\epsilon = 5\%$. The centralized model is also solved using the Gurobi solver, which leads to operation by a single central super entity with complete knowledge and control of the entire ITD systems. The solution quality is summarized in Table II, where the results are nearly the same as the centralized method.

TABLE II
COMPARISON OF SYNC-ADMM, ASYNC-ADMM, CENTRALIZED METHOD

Scheme	Tie-line power (MW)			Generation
	Tie-line 1	Tie-line 2	Tie-line 2	cost ($)
Centralized	239.7	440.2	702.5	78628.1
Sync. ADMM	240.0	440.0	702.3	78628.3
Async. ADMM	239.9	440.3	702.2	78628.3

The distributed optimization for ITD systems is particularly attractive for the co-optimization between TSO and DSOs where they will not need to share their system data due to the legal framework and data confidentiality under deregulated electricity market environments. In the United Kingdom, for instance, the integrated operation between transmission system and distribution systems becomes almost impossible under the deregulated electricity market. While the proposed distributed optimization is to minimize the data exchange between the transmission and distribution systems, then the ITD systems can carry out joint co-optimization together. Meanwhile, the proposed distributed co-optimization scheme is suitable for the implementation by distributed cloud computing.

5) Computation, Idle, and Execution Time: Table III demonstrates the computation time, idle time, and execution time of asynchronous and synchronous ADMM. The Primal and dual residue of tie-line power during the iteration process is compared in Fig. 5. Note that the execution time represents the total time that the algorithm takes to converge, including the parallel computation time and idle time. Here, the communication time is not considered as passing message from one worker to the other usually takes a couple of milliseconds, which is very small compared to local computation time. Thus, there is no idle time for the slowest TS in synchronous scheme. It is clearly shown asynchronous scheme spends relatively less time idling, while the synchronous scheme suffers from greater amount of idle times. In asynchronous scheme, 18% of the time is wasted in idling, while it is 49.2% for synchronous scheme. The asynchronous and synchronous ADMM converges after 32s and 46s with all the primal and dual residues smaller than the predefined convergence criterion. This is due to the fact that in asynchronous ADMM a worker uses the most updated information of its neighbors in a more timely manner than in the synchronous case and updates its local variables more frequently. Global computational progress in the synchronous scheme is held up by the slowest subproblem DS 3 which simultaneously incurs very high idle time on the fastest DS 1. Due to the idle time for the slowest worker at each iteration, a lot of time is wasted on waiting for all workers using a synchronous scheme. The higher computation time for DS 3 forms the main bottlenecks for global progress of synchronous scheme which are readily circumvented by asynchronous scheme.

B. Test System T118D7

Case 2 is a T118D7 system with a modified IEEE 118-bus transmission system connected with seven modified IEEE benchmark distribution systems. The transmission system and distribution systems data are from `MATPOWER 7.1` [56]. The IEEE 118-bus transmission system is connected with one IEEE 33-bus distribution system, one IEEE 69-bus distribution system, one IEEE 85-bus distribution system, one IEEE 136-bus distribution system, one IEEE 141-bus distribution system, one IEEE 85-bus distribution system, and one IEEE 33-bus distribution system at transmission nodes 8, 15, and 25, 34, 65, 80, 100, respectively. The root node of each DS connects the TS. There are five wind farms located at transmission nodes 8, 25, 32, 48, and 63 in TS. The number of distributed wind

TABLE III
COMPARISONS OF TIME FOR SYNC. AND ASYNC. ADMM

Regions		Sync. ADMM		Async. ADMM	
		Time(s)	% of total	Time(s)	% of total
TS	Computation	22	48.9%	27	84.4%
	Idle	23	51.1%	5	15.6%
DS1	Computation	6	13.0%	23	71.9%
	Idle	40	87.0%	9	28.1%
DS2	Computation	19	41.3%	26	81.3%
	Idle	27	58.7%	6	18.8%
DS3	Computation	46	100.0%	29	90.6%
	Idle	0	0.0%	3	9.4%
Total	Computation	93	50.8%	105	82.0%
	Idle	90	49.2%	23	18.0%
Execution		46	\	32	\

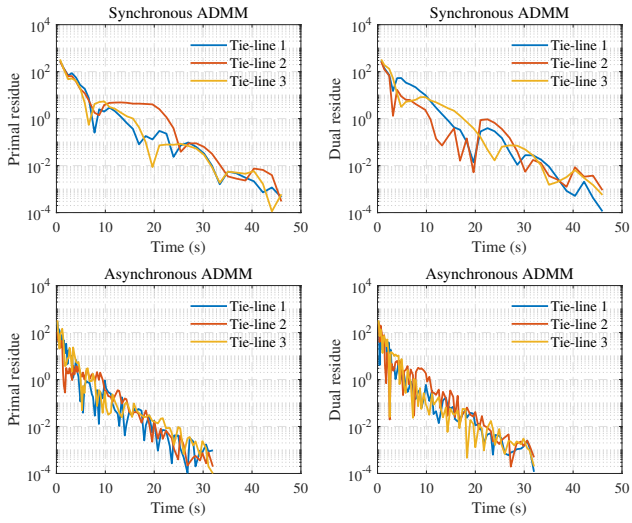

Fig. 5. Primal and dual residue about execution time for Case 1

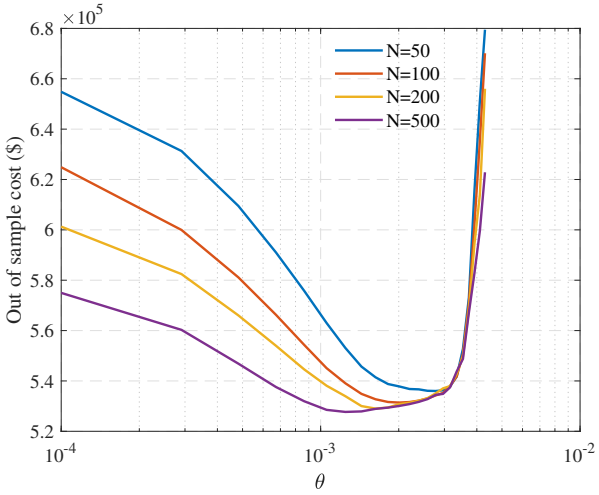

Fig. 6. Average out-of-sample cost under different radius and sample size

generations in the seven DSs are respectively 3, 3, 4, 6, 9, 7, and 4. The network topology, controllable units' parameters, wind location and wind outputs, and other parameters in Case 2 are available online [55]. In Case 2, the sample set is set to $N \in \{50, 100, 200, 500\}$ for TS and each DS. Another $S = 1000$ test samples from the data source are also carried out to test the out-of-sample performance. Different from Case 1, the empirical out-of-sample cost in Case 2 is assessed under the premise that only the decisions $\widehat{z}$ in the nominal scenario is implemented and that the power adjustments of controllable units in TS and each DS are determined by solving a traditional deterministic OPF problem. To obtain a feasible recourse action, drastic measures such as load shedding and wind power spilling must be involved in the deterministic OPF to ensure that the model is feasible.

1) Relationship between Wasserstein Radius and Sample set: Using the sample data, a detailed empirical study on the out-of-sample performance for varying Wasserstein radii is carried out on Case 2. Figure 6 illustrates the empirical out-of-sample cost as a function of Wasserstein radii for different sample sizes, averaged over 30 independent simulation runs. We can observe that the out-of-sample cost have a negative relationship with the sample size, say the larger available samples are, the smaller out-of-sample cost would be. That is, the conservativeness of the proposed DRJCC method is mitigated with the increase of sample size. The reason is that the more historical data is available, the more probabilistic information of the true probability distribution is revealed, and the less conservative the solution is. A closer inspection of the results from Fig. 6 reveals that the out-of-sample cost attains a distinct minimum at a critical Wasserstein radius $\theta^{opt} > 0$. This shows that there exists a best Wasserstein radius corresponding to a lowest out-of-sample cost for different sample size. Meanwhile, this best Wasserstein radius gradually decreases with the increase of sample size. We can conclude that by setting $\theta = \theta^{opt}$, a sophisticated system operator who acknowledges the presence of ambiguity can reduce the out-of-sample cost.

The radius θ is the scalar parameter that controls the solution robustness, which is similar to the uncertainty budget in RO model. For a larger Wasserstein radius, we require the chance constraints to hold for a larger set of distributions. When the number of available samples is small, the system operator can choose a larger radius to improve the robustness of the solution to yet-to-be-realized uncertain parameters. When a large number of historical data is available, the empirical distribution tends to be a good representation of the uncertain parameters and the system operator can reduce the radius to obtain a smaller operation cost. In the implementation, the operator firstly sets a probability for the chance constraint a priori (In practice, the chance constraint is usually required to be satisfied with a high probability (e.g., 90%, 95%) to guarantee a high security level for power systems [26], [57]). Then, the system operator may perform a statistical analysis to identify the desired level of robustness (i.e., the desired radius) based on the violation of specific operational constraints. Ideally, one should select the best possible Wasserstein radius $\theta = \theta^{opt}$ that minimizes the out-of-sample operation cost over

all Wasserstein radii; note that θ^{opt} inherits the dependence on the sample set. As the true distribution is unknown, however, it is impossible to evaluate and minimize the out-of-sample performance. In practice, the best we can hope for is to approximate $\theta = \theta^{opt}$ using the sample data. Statistics offers several methods to accomplish this goal, such as the holdout method [51] and the k-fold cross validation [27].

2) Comparison with Moment-based DRJCC Model: Based on the observations from Fig. 6, the out-of-sample performace of the proposed Wasserstein metric-based DRJCC model with the best radius $\theta = \theta^{opt}$ is compared with the moment-based DRJCC model, shown in 7. The ambiguity set for moment-based model is the set of all probability distributions with given sample mean and sample covariance. Then, the moment-based DRJCC models for TS and DSs can be respectively, formulated as a second-order cone program **P1** as discussed in [32]. It can be seen that the lowest reliability of all chance constraints and the out-of-sample cost in the moment-based model almost stay constant and keep at a relatively unnecessary conservative level under different sample sets. However, the performance of the proposed Wasserstein metric-based model can be usually improved with the increase of sample sets. This is because the proposed Wasserstein metric-based model incorporates a variety of historical data to enhance the characterization of the true probability distribution. Accordingly, the true distribution is revealed with increasing accuracy as the number of available data increases. However, the moment-based model utilizes the first two moments to construct the ambiguity set. Once the moment information is determined, the ambiguity set is fixed, and the solution conservativeness is also determined. Because the samples used to construct the ambiguity set is selected randomly, the mean and covariance barely changes with the sample set grows from 50 to 500, so the solution of the moment-based model almost keeps constant.

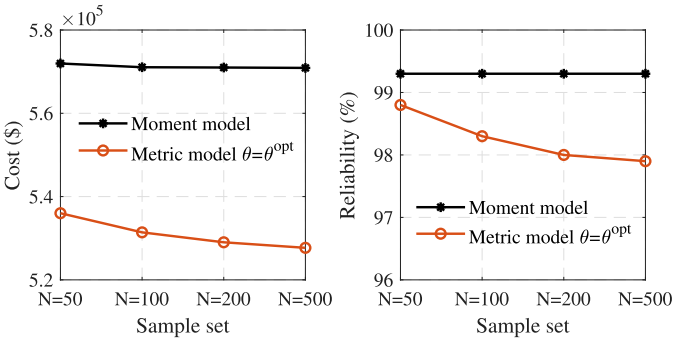

Fig. 7. Comparison of the out-of-sample cost and the lowest out-of-sample constraint reliability

3) Convergence Performance and Solution Quality: The tailored asynchronous ADMM scheme is also compared with its synchronous counterpart to demonstrate its convergence performance in Case 2 with radius $\theta = 0.001$ and risk parameter $\epsilon = 5\%$. The average primal and dual residue of power for the seven tie-lines during the iteration process in Case 2 is compared in Fig. 8. The asynchronous and synchronous ADMM converges after 76s and 103s. This indicates that the proposed asynchronous optimization scheme for ITD systems has good scalability. Of course, if the scale of transmission system is much larger than that of distribution systems, the computing of transmission system will take more time. This will degrade the computational efficiency of the asynchronous algorithm. Under this circumstance, some rules can be adopted to simplify the transmission system model, such as the fast inactive constraint filtration method [11].

In actual ITD systems, if there are tens or hundreds of DS connected with the TS, the computational efficiency will highly depend on how to choose the subset of TS's neighbors who exchange information with TS at each iteration. Some random sampling based subsystem selection strategy could be used [58].

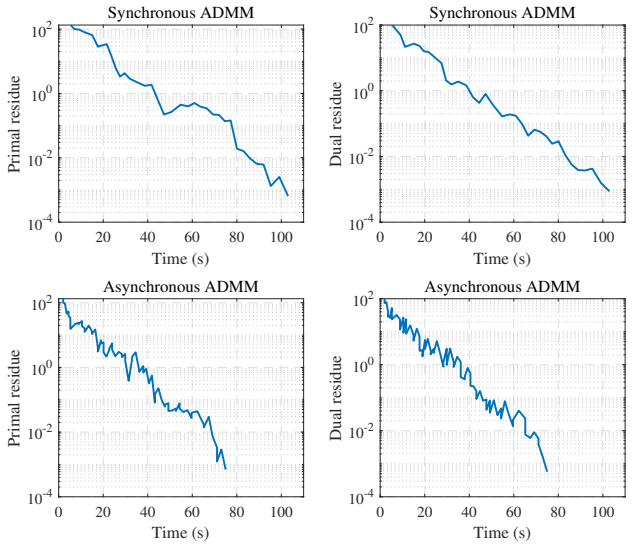

Fig. 8. Average primal and dual residue about execution time for Case 2

In summary, the tailored asynchronous ADMM could be more suitable and scalable for the practical deployment and more efficient than its synchronous counterpart. The impact of communication delay on the convergence performance due to information exchange between subsystems will be minimal. This asynchronous property is more fault-tolerable for delayed or missing information and is invulnerable to communication network condition as well as communication bottleneck. Despite strongly asynchronous systems, more frequent asynchronous updates are able to successfully drive the problem towards the global solution much faster. Besides, for many engineering applications, only a mild level of solution accuracy is needed. Thus the asynchronous updates have the potential to be computational efficiency.

VI. Conclusions

This paper proposes a DRJCC dispatch model for ITD systems via distributed optimization. Based on the data-driven Wasserstein ambiguity set, we proposed the DRJCC models for transmission and distribution systems, respectively. A combined Bonferroni and CVaR approximation is adopted to transform the distributionally robust model into a tractable

conic formulation. A detailed empirical study on the out-of-sample performance of the DRJCC model reveals that the out-of-sample cost can attain a distinct minimum at a critical Wasserstein radius. Meanwhile, a tailored asynchronous ADMM-based fully decentralized dispatch scheme is proposed to better adapt to the star topology of ITD systems, where each subsystem operator can perform local updates with information from a subset of, but not all, neighbors. The proposed asynchronous decentralized optimization scheme has been illustrated by applying two IEEE benchmarks that it has good scalability and can improve computational efficiency for ITD systems.

Future work will investigate several open and more challenging problems, including 1) developing the optimized Bonferroni approximation for metric-based distributionally robust joint chance-constrained program to reduce conservatism, 2) developing an asynchronous and inexact ADMM algorithm to better adapt to the different complexity of each subsystem in ITD systems, and 3) developing the AC-OPF based distributionally robust joint chance-constrained approach for ITD systems.

VII. Acknowledgment

The authors would like to thank eight anonymous reviewers. Their insightful comments and valuable suggestions greatly improved the quality of this paper.

References

[1] A. Baharvandi, J. Aghaei, A. Nikoobakht, T. Niknam, V. Vahidinasab, D. Giaouris, and P. Taylor, "Linearized hybrid stochastic/robust scheduling of active distribution networks encompassing pvs," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 357–367, 2020.

[2] A. Kargarian, Y. Fu, and H. Wu, "Chance-constrained system of systems based operation of power systems," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3404–3413, 2016.

[3] C. Lin, W. Wu, and M. Shahidehpour, "Decentralized ac optimal power flow for integrated transmission and distribution grids," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2531–2540, 2020.

[4] A. Mohammadi, M. Mehrtash, and A. Kargarian, "Diagonal quadratic approximation for decentralized collaborative TSO+DSO optimal power flow," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2358–2370, 2019.

[5] Z. Li, Q. Guo, H. Sun, and J. Wang, "A new lmp-sensitivity-based heterogeneous decomposition for transmission and distribution coordinated economic dispatch," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 931–941, 2018.

[6] ——, "Coordinated economic dispatch of coupled transmission and distribution systems using heterogeneous decomposition," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4817–4830, 2016.

[7] C. Lin, W. Wu, B. Zhang, B. Wang, W. Zheng, and Z. Li, "Decentralized reactive power optimization method for transmission and distribution networks accommodating large-scale dg integration," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 363–373, 2017.

[8] C. Lin, W. Wu, X. Chen, and W. Zheng, "Decentralized dynamic economic dispatch for integrated transmission and active distribution networks using multi-parametric programming," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4983–4993, 2018.

[9] J. Yu, Z. Li, Y. Guo, and H. Sun, "Decentralized chance-constrained economic dispatch for integrated transmission-district energy systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6724–6734, 2019.

[10] Z. Chen, C. Guo, S. Dong, Y. Ding, and H. Mao, "Distributed robust dynamic economic dispatch of integrated transmission and distribution systems," *IEEE Transactions on Industry Applications*, pp. 1–1, 2021.

[11] P. Li, Q. Wu, M. Yang, Z. Li, and N. D. Hatziargyriou, "Distributed distributionally robust dispatch for integrated transmission-distribution systems," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1193–1205, 2021.

[12] A. Kargarian, Y. Fu, and Z. Li, "Distributed security-constrained unit commitment for large-scale power systems," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1925–1936, 2015.

[13] C. Liu, J. Wang, Y. Fu, and V. Koritarov, "Multi-area optimal power flow with changeable transmission topology," *IET Generation, Transmission Distribution*, vol. 8, no. 6, pp. 1082–1089, 2014.

[14] G. Chen and Q. Yang, "An admm-based distributed algorithm for economic dispatch in islanded microgrids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 3892–3903, 2018.

[15] Y. Mou, A. Papavasiliou, and P. Chevalier, "A bi-level optimization formulation of priority service pricing," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2493–2505, 2020.

[16] S. D. Manshadi, G. Liu, M. E. Khodayar, J. Wang, and R. Dai, "A distributed convex relaxation approach to solve the power flow problem," *IEEE Systems Journal*, vol. 14, no. 1, pp. 803–812, 2020.

[17] A. G. Bakirtzis and P. N. Biskas, "A decentralized solution to the dc-opf of interconnected power systems," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1007–1013, 2003.

[18] K. Tang, S. Dong, Y. Liu, L. Wang, and Y. Song, "Asynchronous distributed global power flow method for transmission–distribution coordinated analysis considering communication conditions," *Electric Power Systems Research*, vol. 182, p. 106256, 2020.

[19] H. Zhang, B. Zhang, A. Bose, and H. Sun, "A distributed multi-control-center dynamic power flow algorithm based on asynchronous iteration scheme," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1716–1724, 2018.

[20] J. Zhai, M. Zhou, J. Li, X. Zhang, G. Li, C. Ni, and W. Zhang, "Hierarchical and robust scheduling approach for vsc-mtdc meshed ac/dc grid with high share of wind power," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 793–805, 2021.

[21] Z. Li, W. Wu, B. Zhang, and B. Wang, "Adjustable robust real-time power dispatch with large-scale wind power integration," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 357–368, 2015.

[22] Z. Chen, Z. Li, C. Guo, J. Wang, and Y. Ding, "Fully distributed robust reserve scheduling for coupled transmission and distribution systems," *IEEE Transactions on Power Systems*, vol. 36, no. 1, pp. 169–182, 2021.

[23] Q. Wang, Y. Guan, and J. Wang, "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 206–215, 2011.

[24] C. Zhao, Q. Wang, J. Wang, and Y. Guan, "Expected value and chance constrained stochastic unit commitment ensuring wind power utilization," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2696–2705, 2014.

[25] A. Zhou, M. Yang, Z. Wang, and P. Li, "A linear solution method of generalized robust chance constrained real-time dispatch," *IEEE Trans. Power Syst,*, vol. 33, no. 6, pp. 7313–7316, 2018.

[26] A. Zhou, M. Yang, M. Wang, and Y. Zhang, "A linear programming approximation of distributionally robust chance-constrained dispatch with wasserstein distance," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3366–3377, 2020.

[27] L. Yao, X. Wang, Y. Li, C. Duan, and X. Wu, "Distributionally robust chance-constrained ac-opf for integrating wind energy through multi-terminal vsc-hvdc," *IEEE Trans. Sustain. Energy*, vol. 11, no. 3, pp. 1414–1426, 2020.

[28] R. Zhu, H. Wei, and X. Bai, "Wasserstein metric based distributionally robust approximate framework for unit commitment," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2991–3001, 2019.

[29] Y. Zhang, S. Shen, and J. L. Mathieu, "Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1378–1388, 2017.

[30] K. Baker, E. Dall'Anese, and T. Summers, "Distribution-agnostic stochastic optimal power flow for distribution grids," in *2016 North American Power Symposium (NAPS)*, 2016, pp. 1–6.

[31] A. Hassan, R. Mieth, M. Chertkov, D. Deka, and Y. Dvorkin, "Optimal load ensemble control in chance-constrained optimal power flow," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5186–5195, 2019.

[32] L. Yang, Y. Xu, H. Sun, and W. Wu, "Tractable convex approximations for distributionally robust joint chance constrained optimal power flow under uncertainties," *IEEE Transactions on Power Systems*, pp. 1–1, 2021.

[33] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, and A. Cherukuri, "Wasserstein distributionally robust look-ahead economic dispatch," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2010–2022, 2021.

[34] C. Ordoudis, V. A. Nguyen, D. Kuhn, and P. Pinson, "Energy and reserve dispatch with distributionally robust joint chance constraints," *Oper. Res. Lett.*, vol. 49, no. 3, pp. 291–299, 2021.

[35] M. Zhou, J. Zhai, G. Li, and J. Ren, "Distributed dispatch approach for bulk ac/dc hybrid systems with high wind power penetration," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3325–3336, 2018.

[36] Y. Jiang, P. Sauerteig, B. Houska, and K. Worthmann, "Distributed optimization using aladin for mpc in smart grids," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 5, pp. 2142–2152, 2021.

[37] A. Attarha, P. Scott, and S. Thiébaux, "Affinely adjustable robust admm for residential der coordination in distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1620–1629, 2020.

[38] J. Guo, G. Hug, and O. Tonguz, "Impact of communication delay on asynchronous distributed optimal power flow using admm," in *2017 IEEE International Conference on Smart Grid Communications*, 2017, pp. 177–182.

[39] J. Guo, G. Hug, and O. uz, "Asynchronous admm for distributed non-convex optimization in power systems," *arXiv preprint arXiv:1710.08938*, 2017.

[40] P. Ramanan, M. Yildirim, E. Chow, and N. Gebraeel, "An asynchronous, decentralized solution framework for the large scale unit commitment problem," *IEEE Transactions on Power Systems*, 2019.

[41] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, 2010.

[42] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Math. Program.*, vol. 171, no. 1, pp. 115–166, 2018.

[43] G. Rui and A. J. Kleywegt, "Distributionally robust stochastic optimization with wasserstein distance," 2016.

[44] W. Chen, M. Sim, J. Sun, and C.-P. Teo, "From cvar to uncertainty set: Implications in joint chance-constrained optimization," *Operations research*, vol. 58, no. 2, pp. 470–485, 2010.

[45] W. Xie, S. Ahmed, and R. Jiang, "Optimized bonferroni approximations of distributionally robust joint chance constraints," *Math. Prog.*, pp. 1–34, 2019.

[46] A. R. Hota, A. Cherukuri, and J. Lygeros, "Data-driven chance constrained optimization under wasserstein ambiguity sets," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 1501–1506.

[47] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Oper. Res.*, vol. 58, no. 4-part-1, pp. 902–917, 2010.

[48] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally robust chance-constrained approximate ac-opf with wasserstein metric," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4924–4936, 2018.

[49] J. D. Taft, P. De Martini, and L. Kristov, "A reference model for distribution grid control in the 21st century," *Pacific Northwest National Lab.(PNNL), Richland, WA (United States)*, 2015.

[50] H. Yeh, D. F. Gayme, and S. H. Low, "Adaptive VAR control for distribution circuits with photovoltaic generators," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1656–1663, 2012.

[51] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Math. Prog.*, vol. 171, no. 1, pp. 115–166, 2018.

[52] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2007.

[53] Z. Chen, D. Kuhn, and W. Wiesemann, "Data-driven chance constrained programs over Wasserstein balls," *Optimization Online*, 2018.

[54] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 763–770, 2016.

[55] https://github.com/JunyiZhai1990/ITD-systems.

[56] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2010.

[57] R. Mieth and Y. Dvorkin, "Data-driven distributionally robust optimal power flow for distribution systems," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 363–368, 2018.

[58] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

APPENDIX

A. Considering the Wind Power Curtailment

If one would like to consider the wind power curtailment, the wind power will become decision variables. Taking the TS as an example, the uncertain wind power can be expressed as

$$\tilde{P}_j^{W,T} = \hat{P}_j^{W,T} + \xi_j^T,\ j \in \mathcal{TW} \tag{24}$$

where $\tilde{P}_j^{W,T}$ denotes the uncertain power of wind j.

Similar to the LDRs form of the controllable units' output, the wind output can be written in a LDRs form as follows:

$$P_j^{W,T} = \dot{P}_j^{W,T} - \sum_{k\in\mathcal{TW}} \beta_{jk}^T \xi_k^T,\ j \in \mathcal{TW} \tag{25}$$

where $P_j^{W,T}$ denotes the output of wind j under the realization of wind generation, $\dot{P}_j^{W,T}$ denotes the output of wind j in the nominal scenario, β_{jk}^T denotes the adjustable term under the realization of wind generation.

As a result, the power balance equation (11a) is changed as (26). The chance constraint for transmission line flow limit (11f) is changed as (27). Moreover, a new chance constraint about the wind power curtailment is added as (28), which ensure that the wind power outputs cannot exceed the random wind power.

$$\sum_{i\in\mathcal{TG}} \hat{P}_i^{G,T} + \sum_{i\in\mathcal{TW}} \dot{P}_i^{W,T} = \sum_{i\in\mathcal{TN}} P_i^{L,T} + \sum_{i\in\mathcal{TD}} P_i^{Tie,T} \tag{26}$$

$$\begin{aligned}\min_{\mathbb{P}\in\mathcal{D}^T} \mathbb{P}\Bigg[-\overline{L}_l^T \le \sum_{j\in\mathcal{TW}} M_{jl}\left(\dot{P}_j^{W,T} - \sum_{k\in\mathcal{TW}} \beta_{jk}^T \xi_k^T\right) \\ + \sum_{g\in\mathcal{TG}} M_{gl}\left(\hat{P}_g^{G,T} - \alpha_g^T \sum_{j\in\mathcal{TW}} \xi_j^T\right) - \sum_{k\in\mathcal{TD}} M_{kl} P_k^{Tie,T} \\ - \sum_{i\in\mathcal{TN}} M_{il} P_i^{L,T} \le \overline{L}_l^T,\ \forall l \in \mathcal{TL}\Bigg] \ge 1-\epsilon_2^T\end{aligned} \tag{27}$$

$$\min_{\mathbb{P}\in\mathcal{D}^T} \mathbb{P}\left[P_j^{W,T} \le \tilde{P}_j^{W,T},\ \forall j \in \mathcal{TW}\right] \ge 1-\epsilon_3^T,\ \text{with } 24, 25 \tag{28}$$