Aggregated Feasible Active Power Region for Distributed Energy Resources with a Distributionally Robust Joint Probabilistic Guarantee

Yihong Zhou, *Graduate Student Member, IEEE*, Chaimaa Essayeh, and Thomas Morstyn, *Senior Member, IEEE*

Abstract—Distributed Energy Resources (DERs) have valuable flexibility to provide grid services. The Aggregated Feasible Active Power Region (AFAPR) is useful for aggregating DERs and reducing the computational burden in system-wide DER scheduling. However, the uncertainty of DERs calls for a reliable AFAPR. This paper proposes a novel surrogate polytope method for deriving the inner approximation of the AFAPR that is jointly reliable for all DER constraints and linear network constraints across the scheduling period. Instead of directly applying the chance constraints to the low-level DER constraints and network constraints, the proposed method applies the Wasserstein Distributionally Robust Joint Chance Constraint (WDRJCC) to the surrogate polytope approximation of the AFAPR, which is reformulated into a tractable set of Mixed Integer Linear Programming (MILP) constraints. Our derived inner approximation to the reliable AFAPR is less conservative while still being reliable, as demonstrated by comparisons with four benchmarks in extensive case studies, and with the nonlinear Z-Bus power flow simulation applied to validate the satisfaction of network constraints. The historical data size required is small, making the proposed method easier to deploy. The scale of MILP constraints is small and does not increase with the network size nor with the number of DERs.

Index Terms—Distributed Energy Resources (DERs), flexibility, Aggregated Feasible Active Power Region (AFAPR), Wasserstein Distributionally Robust Joint Chance Constraint (WDRJCC).

I. INTRODUCTION

THE penetration of Distributed Energy Resources (DERs), including renewable generation, storage, electric heating and electric vehicles, has been a growing trend in recent years, and is anticipated to continue accelerating [1]. DERs' flexibility can be used for network constraints management, reserve provision, and frequency regulation [2]. However, directly incorporating a large population of DERs in system-wide scheduling introduces high computing efforts [3]. Distributed optimisation is a scalable approach for dispatching DERs [4], which can be implemented by sharing limited information while maintaining data privacy and decision-making independence of each subsystem. However, the main issue is its slow convergence time, especially when subsystems have different characteristics and computing capabilities [5].

Yihong Zhou is with the School of Engineering, University of Edinburgh, U.K. (Email: yihong.zhou@ed.ac.uk). Chaimaa Essayeh is with the Department of Engineering, Nottingham Trent University, U.K. (Email: chaimaa.essayeh@ntu.ac.uk). Thomas Morstyn is with the Department of Engineering Science, University of Oxford, U.K. (Email: thomas.morstyn@eng.ox.ac.uk).

Another promising solution is DER aggregation [5], which requires the identification of the Aggregated Feasible Power Region (AFPR). An AFPR describes all the aggregated active and reactive power that DERs can achieve while respecting the network constraints [2], [3]. The AFPR can be written as a set of optimisation constraints and can be integrated into upper-level scheduling problems that are more computationally tractable than direct scheduling [5]. This paper specifically focuses on the aggregation of active power (a sub-track of the AFPR), which is denoted the Aggregated Feasible Active Power Region (AFAPR) to distinguish it from the AFPR that aggregates both active and reactive power (also referred to the P-Q chart). Some existing work focused primarily on the AFAPR as here, while others considered the AFPR. Our literature review will include both tracks due to their close relationship, and the methods may be migrated to each other.

AFPR identification is challenging due to the presence of numerous network constraints and DER operational constraints, including time-coupling devices such as batteries [6], contributing to the large problem scale. Therefore, approximations to the exact AFPR are necessary. Authors in [7] made an approximation based on a geometric concept called homothet. A random sampling method was applied in [8], where the DER power trajectories were randomly drawn and only the feasible ones were kept. Ref. [3] applied an optimisation-based framework to maximise the volume of the AFAPR approximation, with all the DERs' operational constraints and network constraints included in the formulation. There are other topics similar to the AFPR where solution methods can be adaptive. One is the feasible power region for the tie-line of inter-connected systems [9]; the other is the Do-Not-Exceed limit (DNE) that sets power limits for each node in the network rather than the aggregated power [10].

The aforementioned studies [3], [6]–[8] assumed that future DER power capabilities are known. However, DERs like solar and wind generation are uncertain in the scheduling phase, so a pressing need is to find an AFPR with a desired reliability (probabilistic) guarantee. Ref. [11] proposed a sampling method to find the reliable PQ chart but requiring excessive computing time. Ref. [12] proposed a data-driven method to learn the price response behaviour of a Distribution Network Operator's (DSO's) aggregated power. Ref. [13] used a sample-based method to tackle the uncertainty in AFAPR considering non-convex DERs. However, their derived AFPRs/AFAPRs [12], [13] do not have an explicit probabilistic guarantee. To reach this probabilistic guarantee, the optimisation framework can be applied. Authors in [14] leveraged Robust Optimisation (RO) to find the reliable AFAPR, but the

use of RO may lead to over-conservativeness [15]. Works in [6] and [16] applied the Chance Constraint (CC) to derive the AFAPR, which can explicitly model the constraint satisfaction probability and reduces the conservativeness compared to RO [17]. However, the CC method assumes that the distribution is known, which is not usually available in practice and leads to bias [15]. The Distributionally Robust Chance Constraint (DRCC) tackles this issue by considering the ambiguity of the distribution. Refs. [18] and [19] applied DRCC to guarantee the individual reliability of stochastic constraints in their AFPR or AFAPR derivation.

In power system applications, it is desired to have a joint reliability guarantee in which all constraints are met simultaneously with high probability [20], calling for a Joint CC (JCC) approach. However, in the context of the AFPR identification, there is a lack of work using JCC due to its intractability, especially when a large number of DER constraints and network constraints are involved. Therefore, aforementioned studies in reliable AFPR or AFAPR [6], [16], [18], [19] applied individual CC (ICC) or Distributionally Robust ICC (DRICC) that controls the constraint violation probability individually. ICC under the same violation rate as JCC leads to a less reliable AFPR as the joint violation rate is not guaranteed. To tackle the JCC for the energy uncertainty of individual Electric Vehicles (EV) in the context of EV aggregation, [21] applied a safe Bonferroni ICC approximation to the JCC by setting a much lower violation probability, but the Bonferroni method can lead to over-conservativeness [17].

When not considering the uncertainty, we notice that the AFAPR admits a surrogate polytope approximation [6], [22], which is essentially a small set of linear constraints for aggregated power only while implicitly considering the DER constraints and network constraints. Therefore, a JCC formulation starting from this surrogate polytope becomes more tractable. By leveraging the Wasserstein Distributionally Robust Joint Chance Constraint (WDRJCC) [23], the JCC based on the surrogate polytope is exactly reformulated into a set of MILP constraints with low complexity, which does not scale with the number of DERs nor with the size of the network. Therefore, our work differs from others using the exact JCC reformulation technique like the virtual power plant scheduling in [24] and DNE in [10] in that their reformulations are directly applied to individual DERs, which may have scalability issues.

In summary, the contributions of this paper are as follows:

- We propose a novel surrogate polytope method to inner-approximate the jointly reliable AFAPR over the entire scheduling horizon, taking into account network constraints and DER constraints including time-coupling effects. The inner approximation has a lower conservativeness due to the exact reformulation of JCC. The linear distribution network model is applied for its tractability in line with existing literature in DER aggregation considering network constraints [3], [6], [14], [16], [18]. The nonlinear Z-Bus power flow simulation is applied in case studies to verify the satisfaction of the exact network constraints.
- We apply WDRJCC to the surrogate polytope, leading to good performance achieved even with limited historical
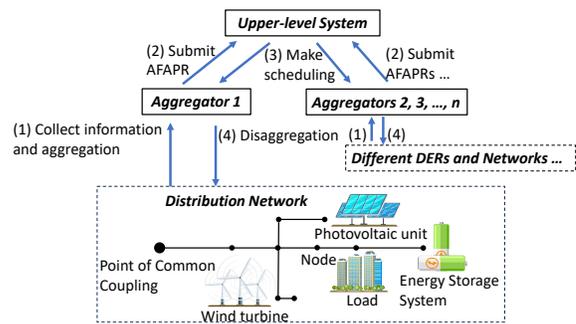
Figure 1. The scheduling process using the AFAPR.

data, which makes the proposed method easier to deploy.
- Our applied surrogate polytope tackles the intractability issue of the large-size JCC. The derived jointly reliable AFAPR is represented as a tractable set of MILP constraints, leading to computationally efficient upper-level scheduling problems. The MILP size does not scale with the number of DERs nor with the network size.
- Comparisons with benchmarks and comprehensive analysis demonstrate the superior performance of the proposed method. The benchmarks include two safe Bonfferoni approximations to the JCC and two unsafe ICC models.

The manuscript is organised as follows. Section II introduces the problem definition, and Section III describes our proposed method. Case studies are carried out in Section IV while Section V concludes the paper.

II. PROBLEM DEFINITION

The process of DER scheduling via aggregation is depicted in Fig. 1. An aggregator manages a set of DERs including PhotoVoltaic units (PV), Wind Turbines (WT), Energy Storage (ES), and Controllable Loads (CL) in a distribution network, which connects to the upper-level system through the Point of Common Coupling (PCC). The aggregator collects information about these DERs, and we assume that the aggregator has access to distribution network information, authorised by the DSO under some contracted conditions [25]. This access is important in the aggregation process to ensure reliable distribution network operation, and this importance can facilitate the collaboration between aggregators and the DSO. Also, the aggregator role may be done by the DSO itself to facilitate the TSO-DSO coordination [26]. When the aggregator and the DSO are two separate entities, there could be conflicting objectives. For example, the aggregator seeks maximum profit while the DSO aims at minimising operational cost. In this case, additional coordination strategies must be in place to ensure fair consideration of different objectives [27], [28]. This paper focuses primarily on the feasibility characterisation at the PCC (the AFAPR) aggregating downstream DERs and the network, which is a flexible and scalable concept that can also be used for aggregation at different nodes or for different parts of a distribution network [2], improving the computing efficiency of the coordination strategies.

There may be more than one aggregator located in the same downstream network. For each aggregator, the impact on network constraints by the power of invisible DERs owned by other aggregators can be modelled as uncertain non-controllable loads, as will be described in Section II-A3. It is

also possible that the DSO as mediator collects all the information from the downstream aggregators and derives the reliable AFAPR based on our proposed method to facilitate upper-level scheduling and avoid the conservativeness to counter the additional uncertainty.

Given the DER and the network information, the AFAPR is derived and submitted to the upper-level system, which at the same time receives the AFAPRs from other aggregators for different DERs and networks, and performs a system-wide scheduling on these AFAPRs. The use of AFAPRs reduces computing efforts compared to directly scheduling all DERs of all the aggregators. The private DER information and network details are also preserved by each aggregator. As the next step, each aggregator is informed of the scheduled aggregated power, which is disaggregated to individual DERs in the real-time implementation process. Our key focus is the derivation of the reliable AFAPR for the aggregated active power; while the reactive power aggregation, upper-level scheduling, and disaggregation are not our main scope. Our proposed method may be applied to the AFPR case as well, as will be discussed in Remark 4 in Section II-C.

We will next present the mathematical details of the DERs, network, and the jointly reliable AFAPR problem. The lower letter a represents a scalar, the bold lower letter $\boldsymbol{a}$ represents a vector, and the bold upper letter $\boldsymbol{A}$ represents a matrix.

A. DER Model

1) PV and WT: The PV operational power is modelled as
$$\forall t \in [T], \forall i_{\text{PV}} \in [N^{\text{PV}}],$$
$$0 \le p_{t,i_{\text{PV}}}^{\text{PV},\psi} \le \overline{p}_{t,i_{\text{PV}}}^{\text{PV},\psi} \tag{1}$$
$$(p_{t,i_{\text{PV}}}^{\text{PV},\psi}, q_{t,i_{\text{PV}}}^{\text{PV},\psi}) \in \mathcal{PQ}_{t,i_{\text{PV}}}^{\text{PV},\psi} \tag{2}$$
WT has the same form of constraints which are not displayed. $\overline{p}_{t,i_{\text{PV}}}^{\text{PV},\psi}$ is the maximum active power generation of the PV connected to bus i_{PV} at time step t, which depends on the weather and the capacity limit and is uncertain. Here, $\psi \in \Psi$ represents the phase index, where Ψ comprises a, b, c for Wye connections or ab, bc, ca for Delta connections. $[N^{\text{PV}}]$ collects the buses connected with PVs. The actual PV generation $p_{t,i_{\text{PV}}}^{\text{PV},\psi}$ can be scheduled to zero as Eq. (1). PV and WT can also provide reactive power support $q_{t,i_{\text{PV}}}^{\text{PV},\psi}$ and $q_{t,i_{\text{PV}}}^{\text{WT},\psi}$. Their reactive power constraints are described by their PQ-charts [29] in Eq. (2), which is illustrated as the area surrounded by the capacity limit (green curves), active power limits (red curves), and reactive power limits (blue curves) in Fig. 2. To address nonconvexity in PQ charts, we convert them into sets of linear constraints (grey areas) following [16].

2) ES: The operational power of an ES is modelled as:
$$\forall t \in [T], \forall i_{\text{ES}} \in [N^{\text{ES}}]$$
$$p_{t,i_{\text{ES}}}^{\text{ES},\psi} = \widehat{p}_{t,i_{\text{ES}}}^{\text{ES},\psi} + \breve{p}_{t,i_{\text{ES}}}^{\text{ES},\psi} \tag{3}$$
$$\underline{p}_{i_{\text{ES}}}^{\text{ES},\psi} \le \breve{p}_{t,i_{\text{ES}}}^{\text{ES},\psi} \le 0 \tag{4}$$
$$0 \le \widehat{p}_{t,i_{\text{ES}}}^{\text{ES},\psi} \le \overline{p}_{i_{\text{ES}}}^{\text{ES},\psi} \tag{5}$$
$$e_{t,i_{\text{ES}}} = e_{t-1,i_{\text{ES}}} + \sum_{\psi}(\widehat{p}_{t,i_{\text{ES}}}^{\text{ES},\psi}\widehat{\eta} + \breve{p}_{t,i_{\text{ES}}}^{\text{ES},\psi}/\breve{\eta}) \cdot \Delta t \tag{6}$$
$$\underline{e}_{i_{\text{ES}}} \le e_{t,i_{\text{ES}}} \le \overline{e}_{i_{\text{ES}}} \tag{7}$$
$$(p_{t,i_{\text{ES}}}^{\text{ES},\psi}, q_{t,i_{\text{ES}}}^{\text{ES},\psi}) \in \mathcal{PQ}_{t,i_{\text{ES}}}^{\text{ES},\psi} \tag{8}$$
Here, $[N^{\text{ES}}]$ collects the buses connected with ESs and $p_{t,i_{\text{ES}}}^{\text{ES},\psi}$ is the power of the ES connected to bus i_{ES} at phase ψ and time
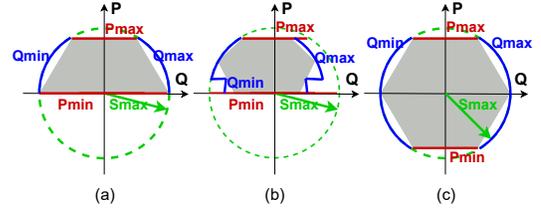
Figure 2. The PQ capability chart [29] for (a) PV, (b) WT, and (c) ES.

step t. The ES has separately modelled charging power $\widehat{p}_{t,i_{\text{ES}}}^{\text{ES},\psi}$ and discharging power $\breve{p}_{t,i_{\text{ES}}}^{\text{ES},\psi}$, subject to the charging power limit $\overline{p}_{i_{\text{ES}}}^{\text{ES},\psi}$ and discharging limit $\underline{p}_{i_{\text{ES}}}^{\text{ES},\psi}$ as Eqs. (4)-(5). The charging and discharging affect the energy level $e_{t,i_{\text{ES}}}$ as the time-coupling constraint (6), which is subject to the charging (discharging) efficiency $\widehat{\eta}$ ($\breve{\eta}$) due to the power loss. Δt is the length of one time step. The energy level of an ES is limited within the maximum energy $\overline{e}_{t,i_{\text{ES}}}$ and the minimum energy $\underline{e}_{t,i_{\text{ES}}}$ (7). The ES's PQ chart is also linearised in Fig. 2. Note that simultaneous charging and discharging can be avoided by adding a linear penalty term $\omega_{i_{\text{ES}}} \sum_{t,\psi,i^{\text{ES}}}(\widehat{p}_{t,i_{\text{ES}}}^{\text{ES},\psi} - \breve{p}_{t,i_{\text{ES}}}^{\text{ES},\psi})$ to the scheduling objective with $\omega_{i_{\text{ES}}}$ being a preset penalty coefficient [30]. This penalty also represents the linear cost of battery degradation w.r.t. the throughput. We model this cost in both our aggregated cost function and the disaggregation process in Section III-D. Ref. [31] proved this result for $\forall \omega_{i_{\text{ES}}} > 0$ in a DC OPF problem under some common power system statuses via KKT conditions.

3) Load: We consider CLs that can adjust their power within a continuous range modelled as
$$\forall t \in [T], \forall i_{\text{CL}} \in [N^{\text{CL}}],$$
$$\sigma_{\min}\ddot{p}_{t,i_{\text{CL}}}^{\text{CL},\psi} \le p_{t,i_{\text{CL}}}^{\text{CL},\psi} \le \sigma_{\max}\ddot{p}_{t,i_{\text{CL}}}^{\text{CL},\psi} \tag{9}$$
$$\sum_{t}^{T} p_{t,i_{\text{CL}}}^{\text{CL},\psi} = \sum_{t}^{T} \ddot{p}_{t,i_{\text{CL}}}^{\text{CL},\psi} \tag{10}$$
$$q_{t,i_{\text{CL}}}^{\text{CL},\psi} = \alpha \cdot p_{t,i_{\text{CL}}}^{\text{CL},\psi} \tag{11}$$
where $\ddot{p}_{t,i_{\text{CL}}}^{\text{CL},\psi}$ is the uncertain base load of the CL without any intervention. Eq. (10) ensures that the total consumed energy equals the original energy consumption. We assume the CL has a fixed power factor α as Eq. (11). There are also uncertain Non-controllable Loads (NL) in the network with $\sigma_{\min} = \sigma_{\max} = 1$. Note that this paper focuses on the case where a single aggregator collects all the DER information and the network information of a distribution network. However, it is possible to extend to the multiple-aggregator-single-network case, where other aggregators' power scheduling can be regarded as part of the uncertain NLs modelled here.

B. Network Model

Following Refs. [32] and [33], we consider a generic three-phase distribution network with one slack bus (the PCC in Fig. 1) and N^B three-phase PQ buses. Denote by $\boldsymbol{s}_t^{\text{Agg}}$ the three-phase complex power injection of the slack bus, by $\boldsymbol{s}_t^Y$ the three-phase complex power injection of all the PQ buses from Wye sources, by $\boldsymbol{s}_t^D$ the three-phase complex power injection of all the PQ buses from Delta sources, by $\boldsymbol{v}_t^0$ the three-phase complex nodal voltage of the slack bus, by $\boldsymbol{v}_t$ the three-phase complex nodal voltages of all the PQ buses, by $\boldsymbol{i}_t$ the phase net current injections of all the PQ buses, and by $\boldsymbol{i}_t^D$ the phase-

to-phase currents of all the PQ buses. Furthermore, denote the complex three-phase bus admittance matrix as:

$$\boldsymbol{Y} := \begin{bmatrix} \boldsymbol{Y}_{00} & \boldsymbol{Y}_{0L} \\ \boldsymbol{Y}_{L0} & \boldsymbol{Y}_{LL} \end{bmatrix} \in \mathbb{C}^{3(N^B+1) \times 3(N^B+1)} \qquad (12)$$

where $\mathbb{C}$ is the complex domain. $\boldsymbol{Y}_{00}$, $\boldsymbol{Y}_{0L}$, $\boldsymbol{Y}_{L0}$, and $\boldsymbol{Y}_{LL}$ indicate self-/mutual- admittance matrices of/between slack/non-slack buses. The bus admittance matrix $\boldsymbol{Y}$ can be formed from the network topology, the π-model of the three-phase power lines, and other passive elements, representing the physical coupling of a three-phase distribution network. Based on [32], the AC load-flow equations can be expressed as:

$$\mathrm{diag}(\boldsymbol{H}^\top \overline{\overline{\boldsymbol{i}_t^D}}) \boldsymbol{v}_t + \boldsymbol{s}_t^Y = \mathrm{diag}(\boldsymbol{v}_t) \overline{\overline{\boldsymbol{i}_t}}, \qquad (13)$$

$$\boldsymbol{s}_t^D = \mathrm{diag}(\boldsymbol{H}\boldsymbol{v}_t) \overline{\overline{\boldsymbol{i}_t^D}}, \qquad (14)$$

$$\boldsymbol{i}_t = \boldsymbol{Y}_{L0}\boldsymbol{v}_t^0 + \boldsymbol{Y}_{LL}\boldsymbol{v}_t, \qquad (15)$$

$$\boldsymbol{s}_t^{\mathrm{Agg}} = \mathrm{diag}(\boldsymbol{v}_t^0)(\overline{\overline{\boldsymbol{Y}_{00}}}\,\overline{\overline{\boldsymbol{v}_t^0}} + \overline{\overline{\boldsymbol{Y}_{0L}}}\,\overline{\overline{\boldsymbol{v}_t}}) \qquad (16)$$

where $\overline{\overline{\boldsymbol{i}_t^D}}$ indicates conjugate, and $\mathrm{diag}(\cdot)$ constructs a square matrix with the elements of the vector inside as its diagonal. $\boldsymbol{H}$ is a coefficient matrix. Refer to [32], [33] for a detailed interpretation. Like [3] and [6], we apply the linear distribution network model in [32]. By interpreting the fixed-point function derived from the AC load-flow equations, [32] showed that the three-phase voltage magnitudes $|\boldsymbol{v}_t|$, the three-phase line current magnitudes $|\boldsymbol{i}_t^L|$, and the aggregated active power at the PCC $\boldsymbol{p}_t^{\mathrm{Agg}}$ (summing over the three phases) can be expressed by linear functions w.r.t. the three-phase active (the real parts of $\boldsymbol{s}_t^Y$ and $\boldsymbol{s}_t^D$) and reactive (the imaginary parts of $\boldsymbol{s}_t^Y$ and $\boldsymbol{s}_t^D$) power injections of the N^B PQ buses. Let $\boldsymbol{x}_t$ collect the active and reactive power of DERs (decision variables) plus NLs (uncertain parameters) at all the connected phases and buses at time step t:

$$\boldsymbol{x}_t := [p_{t,i_d}^{d,\psi}, q_{t,i_d}^{d,\psi}, \ddot{p}_{t,i_{\mathrm{NL}}}^{\mathrm{NL},\psi}, \alpha\ddot{p}_{t,i_{\mathrm{NL}}}^{\mathrm{NL},\psi}] \qquad (17)$$

$$i_d \in [N^d], d \in \{\mathrm{PV}, \mathrm{WT}, \mathrm{ES}, \mathrm{CL}\}, i_{\mathrm{NL}} \in [N^{\mathrm{NL}}], \psi \in \Psi$$

By combining the linear relationship between $\boldsymbol{x}_t$ and the real and imaginary parts of $\boldsymbol{s}_t^Y$ and $\boldsymbol{s}_t^D$, we have the following linear network model:

$$\forall t \in [T], \quad |\boldsymbol{i}_t^L| = \boldsymbol{J}_t\boldsymbol{x}_t + \boldsymbol{j}_t^0 \qquad (18)$$

$$|\boldsymbol{v}_t| = \boldsymbol{K}_t\boldsymbol{x}_t + \boldsymbol{k}_t^0 \qquad (19)$$

$$p_t^{\mathrm{Agg}} = \boldsymbol{g}_t^\top\boldsymbol{x}_t + g_t^0 \qquad (20)$$

$$|\boldsymbol{i}_{\mathrm{min}}^L| \leq |\boldsymbol{i}_t^L| \leq |\boldsymbol{i}_{\mathrm{max}}^L| \qquad (21)$$

$$|\boldsymbol{v}_{\mathrm{min}}| \leq |\boldsymbol{v}_t| \leq |\boldsymbol{v}_{\mathrm{max}}| \qquad (22)$$

where $\boldsymbol{J}_t$, $\boldsymbol{j}_t^0$, $\boldsymbol{K}_t$, $\boldsymbol{k}_t^0$, $\boldsymbol{g}_t$, g_t^0 are linear coefficients derived from a given operational point, a no-load case, and the three-phase bus admittance matrix $\boldsymbol{Y}$, leading to better global performance than the first-order Taylor's approximation [32]. Refer to [32] for a detailed mathematical deduction. Its global performance is also theoretically analysed and verified through extensive case studies in [32], [33], showing a less than 0.6% (1.4%) voltage relative error for a 2000-node distribution feeder (IEEE-8500 feeder, resp.) under a range of load injections. Note that, in general, the accuracy of a single linear network model only holds when the optimised operating point is close to the given operating point used to derive the linear model coefficients [33], [34], given the nonlinear nature of the AC load flow equations. Iterative approaches can be necessary

to find accurate network solutions with updated linear model coefficients in each iteration [34].

Another promising track in network modelling is convex relaxation. However, existing convex relaxation is only exact for specific conditions that are unfortunately difficult to predict in practice; when the conditions do not hold, there may be a large optimality gap and it is difficult to reconstruct a feasible solution [34]–[36]. In contrast, although there is no guarantee of exactness as well, the linear network model is much more computationally efficient given the current mature and reliable linear programming solvers [35].

The line current and nodal voltage need to be constrained within a given range as Eq. (21) and (22). Real-world distribution networks are typically voltage-constrained and the line current limits can be disregarded [4], especially for distribution networks with long power lines [37], so we only consider the three-phase nodal voltage limits as Eq. (22). Incorporating the line current constraints is straightforward as it is also linear.

C. Problem Target: Jointly Reliable AFAPR

The outcome of a DER aggregation problem is the AFAPR Ω^{Agg}, which is a feasible region for the aggregated active power $\boldsymbol{p}^{\mathrm{Agg}} := [p_{t=0}^{\mathrm{Agg}}, \cdots, p_{t=T}^{\mathrm{Agg}}]$. The aggregated power is coupled with the individual DER power and NL power $\boldsymbol{x} := [\boldsymbol{x}_{t=0}, \cdots, \boldsymbol{x}_{t=T}]$ by Eq. (20), which are further subject to the DER operational constraints and network constraints. Therefore, we can define the AFAPR as:

$$\Omega^{\mathrm{Agg}} = \{\boldsymbol{p}^{\mathrm{Agg}} \mid \exists \boldsymbol{x}, \text{ Eq. (1)-(22) } \forall \psi, i^D, t \} \qquad (23)$$

Fig. 3(a) visualises our AFAPR definition. The feasible region defined by all the DER constraints and network constraints (1)-(22) for the $[\boldsymbol{p}^{\mathrm{Agg}}, \boldsymbol{x}]$ space can be represented as the high-dimension yellow polytope. Because the AFAPR Ω^{Agg} in (23) focuses on the aggregated power only, it can be represented as the red projected polytope in the $\boldsymbol{p}^{\mathrm{Agg}}$ space in Fig. 3(a). Note that (23) defines the AFAPR across T time steps rather than for one time step only, which ensures the time-coupling constraints of ESs and the energy requirements of CLs.

The uncertain DERs make the AFAPR Ω^{Agg} uncertain in the scheduling phase. Here, the uncertain parameters include the maximum active power capability of PV $\bar{p}_{t,i_{\mathrm{PV}}}^{\mathrm{PV},\psi}$ and WT $\bar{p}_{t,i_{\mathrm{WT}}}^{\mathrm{WT},\psi}$, the base load of CL $\ddot{p}_{t,i_{\mathrm{CL}}}^{\mathrm{CL},\psi}$ and NL $\ddot{p}_{t,i_{\mathrm{NL}}}^{\mathrm{NL},\psi}$, and thus a part of $\boldsymbol{x}$. For managing the risk, it is desired to find a "jointly reliable AFAPR" $\Omega^{\mathrm{Agg}}(\epsilon)$ such that any $\boldsymbol{p}^{\mathrm{Agg}}$ within it can be implementable without violating any constraints for the entire scheduling horizon $[T]$ with the desired probability $1 - \epsilon$. JCC provides an explicit modelling of such probability. Accordingly, we can define $\Omega^{\mathrm{Agg}}(\epsilon)$ as:

$$\Omega^{\mathrm{Agg}}(\epsilon) = \{\boldsymbol{p}^{\mathrm{Agg}} | \mathbb{P}(\boldsymbol{p}^{\mathrm{Agg}} \in \Omega^{\mathrm{Agg}}) \geq 1 - \epsilon\} =$$

$$\left\{ \boldsymbol{p}^{\mathrm{Agg}} \left| \mathbb{P}\left(\exists \boldsymbol{x}, \begin{array}{c} (1),(9),(10),(20),(21),(22) \\ \forall t, \psi, i_{\mathrm{PV}}, i_{\mathrm{WT}}, i_{\mathrm{CL}} \\ (2), (3)\text{-}(8), (11), \forall t, \psi, i_{\mathrm{ES}} \end{array} \right) \geq 1 - \epsilon, \right. \right\} \qquad (24)$$

where the probability function $\mathbb{P}$ defines a JCC requiring that all DER constraints and network constraints involving uncertain parameters need to be met simultaneously with a probability of at least $1 - \epsilon$. The power balance constraint (20) is included as it involves $\boldsymbol{x}$, which contains the uncertain NL defined in (17). Constraints without uncertain parameters

are listed out of the JCC. Eqs. (18) and (19) can be absorbed to their line current limit (21) and voltage limit (22) by changing variables and are thus omitted. Note that the equality constraints (10) and (20) need further processing and will be discussed in our benchmark setting in Section IV-B. Also, (21) and (22) are in the vector expression so they are a set of constraints for all phases and nodes.

The size of the JCC in the jointly reliable AFAPR $\Omega^{\text{Agg}}(\epsilon)$ (24) is large as it contains constraints for all the time steps, uncertain DERs, phases and nodes, so the exact reformulation methods like [24] become less tractable. Note that the ICC model used in existing AFPR or AFAPR work [6], [16], [18], [19] is tractable, however, less reliable. Finding the analytical expression of the jointly reliable AFAPR $\Omega^{\text{Agg}}(\epsilon)$ is challenging, and thus approximations are necessary. For the safety preference in power system operations, we specifically focus on an inner approximation of the original $\Omega^{\text{Agg}}(\epsilon)$. In other words, our objective is to find an approximated AFAPR $\Omega^{\text{Agg}}_{\text{Approx}}(\epsilon)$ such that:

$$\Omega^{\text{Agg}}_{\text{Approx}}(\epsilon) \subseteq \Omega^{\text{Agg}}(\epsilon) \text{ in (24)} \qquad (25)$$

We also require the derived approximation $\Omega^{\text{Agg}}_{\text{Approx}}(\epsilon)$ to have a mild complexity when it is integrated into the upper-level scheduling and be capable of most of the scheduling problems. In addition, the inner approximation should not be overly conservative: the Bonferroni approximation [21] is a tractable inner approximation but it tends to be overly conservative especially under the large-size JCC in $\Omega^{\text{Agg}}(\epsilon)$ (24) [20], [21].

Our AFAPR approximation problem can be alternatively defined as maximising the volume of the approximator, with (25) included in the constraints [3], [13], [14]. However, the volume maximisation is hard to solve here, as our target, namely the true jointly reliable AFAPR $\Omega^{\text{Agg}}(\epsilon)$ in (24), is generally nonconvex and it is hard to find the analytic form due to the joint probabilistic constraint [38]. In addition, even though our target can be convex in rare cases, volume maximisation problems are only solvable for specific approximators like boxes [3], ellipsoids [14], or shifting and scaling prototype polytopes, i.e., the homothet with the maximum volume [13].

Remark 1: It is important to ensure the reliability of AFAPR, i.e., high probability of implementability, which is our main focus. When a scheduled p^{Agg} is not implementable, there can be two implications: 1) DERs have enough power flexibility but network constraints are violated, and/or 2) DERs do not have enough power flexibility. The former case threatens the network operation. The latter result may be compensated by the transmission grid, which, however, would become less viable when in the future the DSO plays a larger share in the whole system.

Remark 2: Depending on specific problems, an AFAPR containing less reliable power trajectories may still lead to acceptably accurate scheduling; thus, one may explore the application-oriented reliable AFAPR based on information about the specific scheduling problems that will be carried out. Nonetheless, our current focus $\Omega^{\text{Agg}}(\epsilon)$ is more universal as any scheduling problems based on it can provide solutions that are reliable. This is particularly beneficial in TSO-DSO interactions where a DSO may not know the exact upper-level
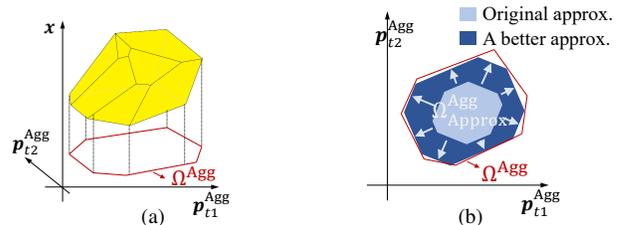
Figure 3. (a): The high-dimension yellow polytope represents the feasible region of $[p^{\text{Agg}}, x]$ defined by all the DER constraints and network constraints. Because the AFAPR Ω^{Agg} in (23) focuses on the aggregated power only, it can be represented as the red projected polytope in the p^{Agg} space. (b): A better inner approximation can be achieved by tuning the intercept of $\Omega^{\text{Agg}}_{\text{Approx}}$ with the same slope.

scheduling problem formulation after submitting the AFAPR.

Remark 3: Recently [39] proposed the concept of non-anticipativity to handle uncertainties in multistage problems. However, we aim to identify the region of the aggregated power that is available with a high probability for the future scheduling horizon. Chance constraints provide explicit modelling of this probability and are therefore naturally applied.

Remark 4: Reactive power is another important dimension in DER aggregation, that is, the AFPR or the PQ chart. Ref. [2] provided a systematic framework for the aggregated PQ chart at each time step. Note that, since the time-coupling constraints of storage-like DERs couple the active power across time and the active power couples the reactive power, it may be necessary to have a time-coupled AFPR beyond the common single-time-step PQ aggregation [2]. The time-coupled AFPR can be defined by slightly modifying our AFAPR definition in (23) and (24), where p^{Agg} will be replaced with $[p^{\text{Agg}}, q^{\text{Agg}}]$, and $q^{\text{Agg}} := [q^{\text{Agg}}_{t=0}, \cdots, q^{\text{Agg}}_{t=T}]$, which leads to a more complicated setting. Future work can investigate whether our proposed polytope-based method in Section III can be adaptable. In this paper, our main focus is on the AFAPR of the active power.

III. PROPOSED METHOD

It is challenging to find a good inner approximation under the large JCC in $\Omega^{\text{Agg}}(\epsilon)$ (24). However, starting from the surrogate polytope inner approximation in a deterministic setting, we instead apply the JCC to the surrogate polytope and finally derive a tractable analytical expression for the inner approximation of the jointly reliable AFAPR $\Omega^{\text{Agg}}(\epsilon)$.

A. Surrogate Polytope in the Deterministic Setting

When there is no uncertainty, we notice that the AFAPR Ω^{Agg} in Eq. (23) admits a polytope inner approximation $\Omega^{\text{Agg}}_{\text{Approx}}$ with a pre-selected shape [22], such that:

$$\Omega^{\text{Agg}}_{\text{Approx}} \subseteq \Omega^{\text{Agg}}, \ \Omega^{\text{Agg}}_{\text{Approx}} = \{p^{\text{Agg}} | Ap^{\text{Agg}} \leq b\} \qquad (26)$$

where the matrix A controls the slope of each facet of the polytope, i.e., shape, and is pre-selected, and the vector b determines the intercept of each facet and is to be inferred. Fig. 3(b) shows that adjusting the intercept b while fixing the shape A is able to achieve a good inner approximation of the true AFAPR Ω^{Agg}. The parameters of b for such an inner polytope approximation can be inferred in a computationally affordable manner by methods in [6], [22], which also empirically showed the good approximation accuracy. As the network constraints and DER constraints are included in Ω^{Agg}, the polytope approximation also implicitly includes these low-level constraints due to $\Omega^{\text{Agg}}_{\text{Approx}} \subseteq \Omega^{\text{Agg}}$.

1) Virtual Battery Approximator: Here we specifically introduce the Virtual Battery (VB) approximator in [6] as we use this setting in our case studies. Note that our proposed method works for any polytope approximator with pre-selected slopes. A VB model is essentially a set of power constraints and energy constraints making it look like a battery:

$$\Omega_{\text{Approx}}^{\text{Agg}} = \{\boldsymbol{p}^{\text{Agg}} | \boldsymbol{A}^{\text{VB}} \boldsymbol{p}^{\text{Agg}} \leq \boldsymbol{b}^{\text{VB}}\} \tag{27}$$

$$\boldsymbol{A}^{\text{VB}} = [\boldsymbol{I}, -\boldsymbol{I}, \boldsymbol{\Lambda}, -\boldsymbol{\Lambda}]^{\top} \tag{28}$$

$$\boldsymbol{b}^{\text{VB}} = [\bar{p}_{t=1}^{\text{VB}}, \cdots, \bar{p}_{t=T}^{\text{VB}}, -\underline{p}_{t=1}^{\text{VB}}, \cdots, -\underline{p}_{t=T}^{\text{VB}}, \\ \bar{e}_{t=1}^{\text{VB}}, \cdots, \bar{e}_{t=T}^{\text{VB}}, -\underline{e}_{t=1}^{\text{VB}}, \cdots, -\underline{e}_{t=T}^{\text{VB}}] \tag{29}$$

where $\boldsymbol{I} \in \mathbb{R}^{T \times T}$ is the identity matrix and $\boldsymbol{\Lambda} \in \mathbb{R}^{T \times T}$ is a lower-triangle matrix with all non-zero elements equal to 1. $\boldsymbol{b}^{\text{VB}}$ collects the VB parameters including the maximum power $\bar{p}_t^{\text{VB}}$, minimum power $\underline{p}_t^{\text{VB}}$, maximum energy $\bar{e}_t^{\text{VB}}$, and minimum energy levels $\underline{e}_t^{\text{VB}}$ at each time step $t \in [T]$.

The surrogate polytope constraints in (26) or the special case (27) apply to $\boldsymbol{p}^{\text{Agg}}$ only. The number of constraints in (26) can be smaller than that in Ω^{Agg} (23). Specifically, for the VB polytope, the number of constraints is only $4T$, leading to tractable JCC reformulation as introduced next.

B. Extension to Jointly Reliable AFAPR Approximation

When considering the DERs' uncertainty, similar to what we did for the target jointly reliable AFAPR $\Omega^{\text{Agg}}(\epsilon)$ (24), we can define a reliable polytope approximator $\Omega_{\text{Approx}}^{\text{Agg}}(\epsilon)$ as:

$$\Omega_{\text{Approx}}^{\text{Agg}}(\epsilon) = \{\boldsymbol{p}^{\text{Agg}} | \mathbb{P}_{\tilde{\boldsymbol{b}}}(\boldsymbol{p}^{\text{Agg}} \in \Omega_{\text{Approx}}^{\text{Agg}}) \geq 1 - \epsilon\} \\ = \{\boldsymbol{p}^{\text{Agg}} | \mathbb{P}_{\tilde{\boldsymbol{b}}}(\boldsymbol{A}\boldsymbol{p}^{\text{Agg}} \leq \tilde{\boldsymbol{b}}) \geq 1 - \epsilon\} \tag{30}$$

where $\mathbb{P}_{\tilde{\boldsymbol{b}}}(\boldsymbol{A}\boldsymbol{p}^{\text{Agg}} \leq \tilde{\boldsymbol{b}}) \geq 1 - \epsilon$ is called "surrogate polytope JCC". $\mathbb{P}_{\tilde{\boldsymbol{b}}}$ is the probability function for the true distribution of $\tilde{\boldsymbol{b}}$: since the slope $\boldsymbol{A}$ is pre-selected, our uncertainty towards the approximated AFAPR (26) reduces to $\boldsymbol{b}$ only, which can be modelled as a random vector $\tilde{\boldsymbol{b}}$, highlighted by the overhead "tilde". However, the true distribution of $\tilde{\boldsymbol{b}}$ is unknown and a subjective assumption can lead to bias. To represent our ambiguity towards the distribution of $\tilde{\boldsymbol{b}}$, we replace the polytope JCC in (30) with a distributionally robust form by WDRJCC; hence our WDRJCC-based approximation to $\Omega^{\text{Agg}}(\epsilon)$, denoted as $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$, can be defined as:

$$\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta) = \{\boldsymbol{p}^{\text{Agg}} | \mathbb{P}_{\tilde{\boldsymbol{b}}}'(\boldsymbol{p}^{\text{Agg}} \in \Omega_{\text{Approx}}^{\text{Agg}}) \geq 1 - \epsilon, \forall \mathbb{P}_{\tilde{\boldsymbol{b}}}' \in \mathcal{F}(\theta)\} \\ = \{\boldsymbol{p}^{\text{Agg}} | \mathbb{P}_{\tilde{\boldsymbol{b}}}'(\boldsymbol{A}\boldsymbol{p}^{\text{Agg}} \leq \tilde{\boldsymbol{b}}) \geq 1 - \epsilon, \ \forall \mathbb{P}_{\tilde{\boldsymbol{b}}}' \in \mathcal{F}(\theta)\} \tag{31}$$

where $\mathcal{F}(\theta)$ is the Wasserstein ambiguity set capturing all the probability functions $\mathbb{P}_{\tilde{\boldsymbol{b}}}'$ for the distributions within the desired Wasserstein distance (radius) θ to the empirical distribution, whose probability function $\mathbb{P}_{\tilde{\boldsymbol{b}}}^{E}$ is formed up by the historical data. $\mathcal{F}(\theta)$ is thus defined as $\mathcal{F}(\theta) := \{\mathbb{P}_{\tilde{\boldsymbol{b}}}' : d_W(\mathbb{P}_{\tilde{\boldsymbol{b}}}^{E}, \mathbb{P}_{\tilde{\boldsymbol{b}}}') \leq \theta\}$, with the distance function:

$$d_W(\mathbb{P}_{\tilde{\boldsymbol{b}}}^{E}, \mathbb{P}_{\tilde{\boldsymbol{b}}}') \\ := \inf_{\Pi} \left\{ \begin{array}{l} \mathbb{E}_{(\boldsymbol{\xi}, \boldsymbol{\xi}') \sim \Pi}[\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|] \\ \text{s.t. } \Pi \text{ has marginal distributions} \\ \text{with probability functions } \mathbb{P}_{\tilde{\boldsymbol{b}}}^{E}, \mathbb{P}_{\tilde{\boldsymbol{b}}}'. \end{array} \right\} \tag{32}$$

Other forms of ambiguity set exist. We choose the Wasserstein set for its rigorous out-of-sample guarantees [15].

C. Exact Reformulation

We next show our approximation $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ admits a tractable reformulation. Suppose we have collected N real-

isations for the random vector $\tilde{\boldsymbol{b}} \in \mathbb{R}^D$ and denote the dataset for these realisations as $\mathcal{D} := \{\boldsymbol{b}^1, \cdots, \boldsymbol{b}^N\}$, where D is the dimension of the random vector $\tilde{\boldsymbol{b}}$ in Eq. (30). Let q_d denote the $(\lfloor \epsilon N \rfloor + 1)^{\text{th}}$ greatest value of $\{-b_d^1, \cdots, -b_d^N\}$, where $\lfloor \cdot \rfloor$ is the floor function and b_d^i denotes the dth entry of the ith realisation $\boldsymbol{b}^i$ of $\tilde{\boldsymbol{b}}$. We further introduce the index set $[N]_d$ as $[N]_d := \{i \in [N] : -b_d^i > q_d\}$ with $d \in [D]$.

Based on [23], the WDRJCC in (31) admits an exact reformulation; thus our approximated reliable AFAPR $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ can be equivalently written as a set of MILP constraints:

$$\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$$

$$= \left\{ \boldsymbol{p}^{\text{Agg}} \in \mathcal{X} : \begin{array}{ll} \exists \boldsymbol{z} \in \{0,1\}^N, s \geq 0, \boldsymbol{r} \geq \boldsymbol{0}, & \text{(a)} \\ \epsilon s \geq \theta + \dfrac{1}{N} \sum_{i \in [N]} r_i, & \text{(b)} \\ M(1 - z_i) \geq s - r_i, i \in [N], & \text{(c)} \\ \sum_{i \in [N]} z_i \leq \lfloor \epsilon N \rfloor, & \text{(d)} \\ b_d^i - \boldsymbol{a}_d^{\top} \boldsymbol{p}^{\text{Agg}} + (-b_d^i - q_d)z_i & \\ \quad \geq s - r_i, \quad i \in [N]_d, d \in [D], & \text{(e)} \\ -q_d - \boldsymbol{a}_d^{\top} \boldsymbol{p}^{\text{Agg}} \geq s, \quad d \in [D] & \text{(f)} \end{array} \right\} \tag{33}$$

where $\boldsymbol{z} \in \mathbb{Z}^N$, $s \in \mathbb{R}$, $\boldsymbol{r} \in \mathbb{R}^N$ are auxiliary decision variables, and $\boldsymbol{a}_d$ is the dth row of the slope matrix $\boldsymbol{A}$ in $\Omega_{\text{Approx}}^{\text{Agg}}$ (26). $\mathcal{X} \subset \mathbb{R}^T$ is a compact domain for $\boldsymbol{p}^{\text{Agg}}$ and can be defined by the maximum and minimum aggregated power of the network based on the installed capacity. M is a sufficiently large constant and is suggested in [23] to be

$$M = \max_{\boldsymbol{p}^{\text{Agg}} \in \mathcal{X}, d \in [D]} \{|b_d^i - \boldsymbol{a}_d^{\top} \boldsymbol{p}^{\text{Agg}}|\} \tag{34}$$

Comparing the derived $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ in Eq. (33) to the polytope approximation under no uncertainty in Eq. (26), we can see that the MILP constraints introduce additional ancillary terms in Eq. (33)(e) and (f) that tighten the original polytope constraints, which reflects the risk-averse attitude (ϵ) and hedges the worst-case distribution in WDRJCC (θ) [23]. Since θ is the radius of the Wasserstein ambiguity set $\mathcal{F}(\theta)$, a larger θ leads to a smaller reliable AFAPR by definition.

Furthermore, we can show the derived $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ in (33) achieves our objective in Eq. (25):

Lemma 1. *For our $\Omega_{WDR}^{Agg}(\epsilon, \theta)$ defined in (33), if $\mathbb{P}_{\tilde{\boldsymbol{b}}} \in \mathcal{F}(\theta)$, then $\Omega_{WDR}^{Agg}(\epsilon, \theta) \subseteq \Omega^{Agg}(\epsilon)$ in (24).*

Proof. Note that $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ in (33) is equivalent to (31); thus for $\forall \boldsymbol{p}^{\text{Agg}'} \in \Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$, we have

$$\mathbb{P}_{\tilde{\boldsymbol{b}}}'(\boldsymbol{p}^{\text{Agg}'} \in \Omega_{\text{Approx}}^{\text{Agg}}) \geq 1 - \epsilon, \ \forall \mathbb{P}_{\tilde{\boldsymbol{b}}}' \in \mathcal{F}(\theta) \tag{35}$$

Since $\mathbb{P}_{\tilde{\boldsymbol{b}}} \in \mathcal{F}(\theta)$, we further have $\mathbb{P}_{\tilde{\boldsymbol{b}}}(\boldsymbol{p}^{\text{Agg}'} \in \Omega_{\text{Approx}}^{\text{Agg}}) \geq 1 - \epsilon$. Finally, because $\Omega_{\text{Approx}}^{\text{Agg}} \subseteq \Omega^{\text{Agg}}$, a solution $\boldsymbol{p}^{\text{Agg}'}$ being within $\Omega_{\text{Approx}}^{\text{Agg}}$ with at least a probability $1 - \epsilon$ will also belong to Ω^{Agg} with at least a probability $1 - \epsilon$, i.e.,

$$\mathbb{P}(\boldsymbol{p}^{\text{Agg}'} \in \Omega^{\text{Agg}}) \geq 1 - \epsilon \Leftrightarrow \boldsymbol{p}^{\text{Agg}'} \in \Omega^{\text{Agg}}(\epsilon) \text{ in (24)} \tag{36}$$

We have proved $\forall \boldsymbol{p}^{\text{Agg}'} \in \Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$, $\boldsymbol{p}^{\text{Agg}'} \in \Omega^{\text{Agg}}(\epsilon)$, then equivalently we have $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta) \subseteq \Omega^{\text{Agg}}(\epsilon)$. $\square$

In other words, when our Wasserstein ambiguity set $\mathcal{F}(\theta)$ covers the probability function $\mathbb{P}_{\tilde{\boldsymbol{b}}}$ for the true distribution of $\tilde{\boldsymbol{b}}$, Lemma 1 states that our proposed $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ (33) achieves

our objective in (25). Lemma 1 also indicates that our proposed $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon, \theta)$ (33) considers the numerous network constraints and DER constraints including the large JCC in (24), even though these constraints are not explicitly modelled in (33). Finally, as the WDRJCC in (31) is exactly reformulated, our inner approximation $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon, \theta)$ to the jointly reliable AFAPR could achieve a lower conservativeness compared to the Bonferroni approximation.

The low complexity of the approximation is also a part of our objective. The reformulation in Eq. (33) introduces N binary variables and is composed of at most $D + D \lfloor \epsilon N \rfloor + N + 1$ constraints (exclusive of the constraints for the type of decision variables like integrity and non-negativity). The complexity of (33) is low due to the use of surrogate polytope JCC and the Wasserstein metric: as discussed in Section III-A, for the VB polytope used in the case studies, the dimension of $\tilde{b}$, i.e., D, is only $4T$, not increasing with the number of DERs nor with the size of the network. Also, the use of the Wasserstein metric leads to a good out-of-sample performance even under a small number of data samples (N). Our case studies in Section IV demonstrate the low complexity and the good performance for scheduling problems based on the derived $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon)$ (33) with $N = 50$ and $T = 24$. Note that modelling a single generation unit needs T binaries or $3T$ binaries when the shut-down and start-up statuses are explicitly modelled [40], which is close to the $N = 50$ binaries of our proposed MILP-based AFAPR.

It is also worth noting the possible computing challenges when integrating our proposed MILP-based jointly reliable AFAPR into the upper-level scheduling, although the number of binaries introduced by our AFAPR is comparable to a single generator. The upper-level scheduling problem may be a large-scale multi-period AC optimal power flow problem to accurately model network constraints, storage devices, and generation ramping [41], which is already a computationally challenging non-linear non-convex problem. The additional binaries introduced by our AFAPR, especially when there are several AFAPRs for several distribution networks, may amplify the issue. Furthermore, our proposed AFAPR makes scheduling problem at least a MILP, which is in theory NP-hard and can be hard to solve in the worst cases even at a small scale. Future work may explore the conditional value-at-risk method [42] to derive a convex or even linear inner approximation to the WDRJCC in (31), which in turn achieves an inner approximation to our MILP-based jointly reliable AFAPR in (33). A convex inner approximation would improve tractability, but also bring conservativeness.

Finally, the proposed MILP-based AFAPR in (33) can be integrated into any non-linear non-convex upper-level scheduling problems performed on a closed and bounded domain [23].

Remark 5: The proposed MILP-based AFAPR (33) is based on the inner polytope approximation in Section III-A. One may also explore the inner ellipsoid $\{p^{\mathrm{Agg}} | p^{\mathrm{Agg}} = E\xi + e, \|\xi\| \leq 1\}$ to approximate the AFAPR in a deterministic setting [14]. The vector ξ collects auxiliary decision variables. Both E and e are parameters to be inferred [14] and are viewed as random variables under our proposed method to derive the reliable AFAPR as in Section III-B. However, since the random E is

multiplied with decision variables in ξ, the exact analytical WDRJCCO reformulation in (33) is not applicable.

Remark 6: It is possible to integrate more general non-linear but convex DER models and network models into our proposed method. The key part of our proposed method that can be affected by the general convex modelling is the derivation of the inner polytope approximation in a deterministic setting (Section III-A). To derive the VB-based polytope inner approximation, our applied method in [6] is demonstrated only in the linear setting in the original paper, but could be extended to general convex cases since the derivation is based on KKT conditions that hold for general convex programmes. However, the method in [6] will finally introduce binaries that make the solution process more complicated if a convex but nonlinear network model is applied. The box-based polytope inner approximation is demonstrated to be capable of general convex constraints [3], but it can be less accurate in approximating the AFAPR.

Remark 7: Lemma 1 indicates that our proposed AFAPR (33) achieves an inner approximation to the true jointly reliable AFAPR, indicating the under-utilisation of DERs' flexibility to some extent. In contrast, outer approximation [43] as another paradigm in AFAPR approximation can fully exploits the DERs' flexibility. However, a downside of outer approximation is the coverage of infeasible power profiles. As discussed in Remark 1, the infeasibility may trigger severe impacts.

Remark 8: Although Wasserstein Distributionally Robust optimisation that minimises the worst-case expectation has a linear reformulation for some cases [44], our reliable AFAPR is defined as a set of reliable power profiles, which requires a more complicated WDRJCC leading to our MILP reformulation (33). To the best of our knowledge, there are no other exact reformulations for our case that are linear or convex.

Remark 9: Our proposed method may tackle cases with more significant changes including network reconfiguration and the DER faults. Corresponding $\tilde{b}$ scenarios can be generated for these cases and added to $\mathcal{D}$ in deriving our reliable AFAPR (33) to improve the robustness. These require more case study validations and are left as future work. Also, in this paper the probability distribution of uncertain parameters are independent on the DER scheduling (the decision variables), which is commonly assumed in existing literature [6], [20]. There may be specific cases with decision-dependent uncertainty where our formulations would need to be extended to handle.

D. Cost and Disaggregation

The derived AFAPR only characterises the feasibility region and is insufficient for economic upper-level system scheduling [45], which requires a monetization function that maps the aggregated power to the total DER operational costs for providing flexibility. Existing work derived the cost mapping by characterising the cost zone [2], [26], in which the flexibility cost is smaller than a certain threshold, or evaluating the cost for each aggregated power sample with certain intervals [46]. However, these works did not derive the analytical cost curve, which is important to integrate efficiently into optimisation-based scheduling tools [45]. In fact, because of

the complicated constraints of heterogeneous DERs and the network, the analytical cost curve is difficult to derive. Refs. [16] and [45] proposed using a piecewise-linear approximation to derive the analytical cost curve. However, as pointed out in [47], there is still a gap in considering the intertemporality. In general, monetization in DER aggregation remains a vital but underexplored topic.

This work adopted the piecewise-linear approximation in [16] to derive the analytical aggregated cost model associated with the derived AFAPR for each time step respectively, which neglects the intertemporality. However, it should be noted that our case studies evaluate the proposed method using the exact low-level DER operational costs instead of the approximated aggregated cost. This paper considers the following DER operational cost:

$$
\begin{aligned}
\mathcal{C}(\boldsymbol{x}) =\,&\omega^{\mathrm{ES}}\Delta t \sum_{t,\psi,i^{\mathrm{ES}}} (\widehat{p}_{t,i_{\mathrm{ES}}}^{\mathrm{ES},\psi} - \breve{p}_{t,i_{\mathrm{ES}}}^{\mathrm{ES},\psi}) \\
&+ \omega^{\mathrm{CL}}\Delta t \sum_{t,\psi,i^{\mathrm{CL}}} \max\{(\ddot{p}_{t,i_{\mathrm{CL}}}^{\mathrm{CL},\psi} - p_{t,i_{\mathrm{CL}}}^{\mathrm{CL},\psi}),0\}
\end{aligned}
\tag{37}
$$

where the first term is the charging and discharging costs of ES, and the second term models the load curtailment costs. ω^{ES} and ω^{CL} are the corresponding cost coefficients. Renewable generation is considered cost-free. Then, for each time step t we find the piecewise-linear function that fits the minimum achievable total DER cost for each aggregated power p_t^{Agg}, with the fitting procedure detailed in [16]. We use the forecast mean values of uncertain DERs in calculating the aggregated cost model denoted as $\mathcal{C}'(\boldsymbol{p}^{\mathrm{Agg}})$, which will be submitted together with the AFAPR for upper-level system scheduling to respect the DER operation costs.

After the upper-level scheduling, the aggregator will be informed of the scheduled aggregated power $\boldsymbol{p}^{\mathrm{Agg}*}$, which is disaggregated to each DER. For each realisation of uncertain DERs, the disaggregation process can be formulated as:

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \mathcal{C}(\boldsymbol{x}) + \tau(\boldsymbol{x}) \\
\text{s.t.} \quad & \text{Eq. (1)} - (22)
\end{aligned}
\tag{38}
$$

where $\mathcal{C}(\boldsymbol{x})$ is the DER operation cost in (37), and $\tau(\boldsymbol{x})$ penalises the nodal voltage estimated by the linear network model being close to the limit, which is to offset the inaccuracy of the linear network constraints. An alternative could be a re-dispatch based on the updated linear network model. We do not additionally include a penalty for the ES simultaneous charging and discharging, because our modelled ES charging cost in $\mathcal{C}(\boldsymbol{x})$ has the same form of the proposed penalty term in [30], [31] explained in Section II-A2. Note that, as stated in Section II, the aggregated cost and the disaggregation are not the main scope of this paper, and the gradually revealed uncertainty over the disaggregation phase is not considered and is left for future work.

E. Overall Framework

Algorithm 1 summarises the steps for using our proposed method with a detailed description given below. The aggregator derives the jointly reliable AFAPR $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon,\theta)$ and submits it to the upper-level system for the central scheduling. The aggregator first collects the dataset $\mathcal{D}$ that contains possible realisations of the random vector $\tilde{\boldsymbol{b}}$ in (30) in an offline process, which is then used in the scheduling stage.

Algorithm 1 Using the Proposed Jointly Reliable AFAPR

Offline Dataset Gathering Stage

  Aggregator:
1: **for** each historical horizon i **do**
2:     Infer $\boldsymbol{b}_{\mathrm{fore,hist}}^i$ and $\boldsymbol{b}_{\mathrm{hist}}^i$ based on historical DER forecasts and true values;
3:     Push $\boldsymbol{b}_{\mathrm{error}}^i = \boldsymbol{b}_{\mathrm{hist}}^i - \boldsymbol{b}_{\mathrm{fore, hist}}^i$ to $\mathcal{D}_{\mathrm{error}}$;
4: **end for**

Scheduling Stage

  Aggregator:
1: Infer $\boldsymbol{b}_{\mathrm{fore}}$ based on forecasts of uncertain DERs in the horizon to be scheduled;
2: Get $\mathcal{D}$ as Eq. (40);
3: Derive $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon,\theta)$ as Eq. (33) and submit;
  Upper-level System:
4: Solve scheduling problems like Eq. (43) and the solution will meet the desired $1 - \epsilon$ reliability level.

---

1) Offline Dataset Gathering: The dataset $\mathcal{D}$ should contain possible realisations of $\tilde{\boldsymbol{b}}$ for the future horizon. We decompose $\tilde{\boldsymbol{b}}$ into a deterministic part $\boldsymbol{b}_{\mathrm{fore}}$ plus a random part $\tilde{\boldsymbol{b}}_{\mathrm{error}}$:

$$
\tilde{\boldsymbol{b}} = \boldsymbol{b}_{\mathrm{fore}} + \tilde{\boldsymbol{b}}_{\mathrm{error}}
\tag{39}
$$

Such decomposition captures the random variable's dependency on the scheduled day by $\boldsymbol{b}_{\mathrm{fore}}$, which is commonly used in current research [20]. During the ith historical operation horizon, the decision-maker runs a method to find $\boldsymbol{b}$ (e.g., [6]) based on forecasts of uncertain DERs and the real values after the uncertainty is revealed. We denote the one based on forecasts as $\boldsymbol{b}_{\mathrm{fore, hist}}^i$ and the other as $\boldsymbol{b}_{\mathrm{hist}}^i$. The quantity $\boldsymbol{b}_{\mathrm{error}}^i = \boldsymbol{b}_{\mathrm{hist}}^i - \boldsymbol{b}_{\mathrm{fore, hist}}^i$ is added to $\mathcal{D}_{\mathrm{error}}$ as the ith scenario of the random part $\tilde{\boldsymbol{b}}_{\mathrm{error}}$.

In real-world applications, the data collection process can run parallel to the operation. A new data sample will be added to $\mathcal{D}_{\mathrm{error}}$ every time an operation horizon ends. Additionally, our proposed method does not have stringent requirements for the size of the dataset, reducing the efforts on data collection. Our case studies in Section IV-F2 demonstrate that a $\mathcal{D}_{\mathrm{error}}$ with data size 50 can achieve good performance for a day-ahead 24-hour scheduling problem with hourly resolution.

2) Scheduling Stage: In the operational stage, the aggregator first makes forecasts of all the uncertain DERs for the horizon to be scheduled, which is a common step in power system scheduling problems. $\boldsymbol{b}_{\mathrm{fore}}$ is derived based on these forecasts, after which we add each element in $\mathcal{D}_{\mathrm{error}}$ to $\boldsymbol{b}_{\mathrm{fore}}$ to get the desired $\mathcal{D}$ as

$$
\mathcal{D} = \{\boldsymbol{b}_{\mathrm{fore}} + \boldsymbol{b}_{\mathrm{error}}^i | \boldsymbol{b}_{\mathrm{error}}^i \in \mathcal{D}_{\mathrm{error}}\}
\tag{40}
$$

based on which the MILP expression (33) for $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon,\theta)$ can be derived, which is then submitted for upper-level system scheduling (a simple example is (43)). The scheduled aggregated power will be implementable without violating any DER constraints or network constraints for the entire horizon $[T]$ with joint probability $1-\epsilon$. Note that the derived $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon,\theta)$ is for the entire horizon $[T]$ rather than for a single time step. The aggregator can also calculate the aggregated cost associated with the AFAPR and submit them to the upper-level system; the dispatched aggregated power is finally disaggregated to

individual DERs as described in Section III-D.

3) Complexity Analysis: Over the online Steps 1-3 to derive the proposed jointly reliable AFAPR $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ (33) in Algorithm 1, the only computationally demanding step is the derivation of b_{fore} (Step 1) based on forecasts of uncertain DERs, by using existing methods that derive the inner polytope approximation of the AFAPR in a deterministic setting. The main aim of our paper is to improve these existing methods to have a jointly reliable probabilistic guarantee, namely Steps 2 and 3, which involve only basic algebra operations given the collected $\tilde{b}_{\text{error}}$ scenarios offline and are computationally trivial. Therefore, the proposed online process to derive the jointly reliable AFAPR does not change the order of complexity compared with existing methods that derive the inner polytope approximation of the AFAPR in a deterministic setting.

The existing methods in Step 1 involve optimisation problems with complexity dependent on the number of DERs and the network sizes. These methods are shown to be computationally tractable in the respective references [22], [48], and our Algorithm 1 is capable of more efficient methods to perform Step 1. In addition, another option is to derive a data-driven way of generating the $\tilde{b}$ scenarios in $\mathcal{D}$ directly, bypassing Step 1 in calculating b_{fore} and possibly leading to improved accuracy given a sufficient amount of historical data. Our later visualisation of $\tilde{b}$ scenarios in Fig. 7 explained in Section IV-C suggests good predictability, possibly because the high stochasticity of individual DERs offset each other in reaching the aggregation parameters in b.

IV. CASE STUDY

A. Network and DER Settings

Our case studies are based on the IEEE-123 distribution feeder [49] modelled in OPEN [50]. DERs include 10 ESs, 20 PVs, 20 WTs, and 30 CLs. There are also 65 NLs. The DER location and connected phases are shown in Fig. 4. The red colour is for phase a or ab, the green colour represents phase b or bc, and the blue colour is for the rest. The UK power demand is estimated to have 20%-40% curtailment potential by 2050 [1]. We use an intermediate and symmetric setting: we assume the CLs are controllable with a ±30% range. We set the voltage limits in Eq. (22) to be 0.95 to 1.05 p.u. as [16]. The DER cost coefficients $\omega_{i_{\text{ES}}}^{\text{ES}}$ and $\omega_{i_{\text{CL}}}^{\text{CL}}$ are both set to 8p/kWh as [24]. The sum of the peak load for the whole network is about 3500 kW. To fully validate the proposed framework, we consider three case studies with different renewable energy source (RES) settings: 1) LowRES with an 800 kW total PV capacity and an 800 kW WT capacity; 2) HighWT, in which the total WT capacity increases to 5600 kW; and 3) HighPV, in which the total PV capacity increases to 7000 kW. For all three case studies, the total ES size is 1800 kWh and the setting for CLs and NLs is kept unchanged. All three case studies have stressful initial network conditions. Fig. 5 illustrates the nodal voltages for the three case studies under an 'unregulated' profile, where RESs are fully utilised with unity power factors, CLs are maintained at the base levels, and ESs are controlled to minimise their own energy costs. Under-voltage issues occur in the 'LowRES' case due to the peak network load. Both over-voltage and under-voltage issues exist in the 'HighPV'
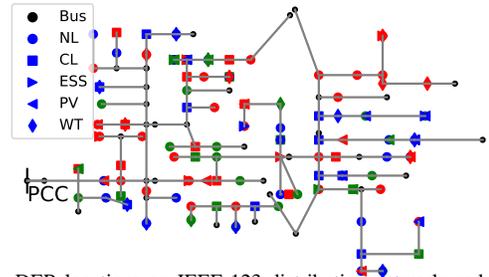
Figure 4. DER locations on IEEE-123 distribution network: red for phase a or ab, green for phase b or bc, and blue for the remaining phases.
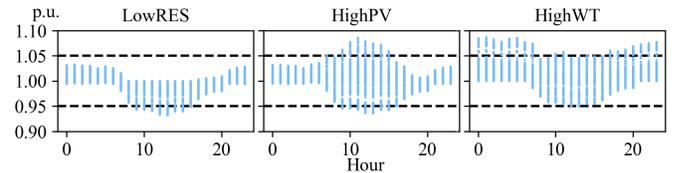
Figure 5. Nodal voltages for all phases at all nodes under an unregulated baseline profile, where RESs are fully utilized with unity power factors, CLs are maintained at the base levels, and ESs are controlled to minimize their own energy costs. Dashed lines are the voltage limits set in the case studies.

case due to the peak solar generation at some nodes and the high demand at other nodes far away from PVs. Finally, in 'HighWT', severe over-voltage issues occur in the morning and at night due to the low demand but high wind availability.

As mentioned in Section II-C, the base load for CL and NL, and the maximum PV and WT generation are considered uncertain. To verify the solution reliability of the proposed method, we generate a set of possible DER power scenarios. Each of the scenarios is generated from a deterministic base curve (forecast) plus a forecasting error scenario with hourly granularity. The base curve for RES comes from the real-world dataset in [51], and the base curve for network load is scaled from the load data for US commercial buildings [52].

The RES forecasting error scenarios come from real-world forecasting error statistics in [53]–[55], which is set the same for all RES in our case study considering a relatively small coverage of a distribution grid. We plot the WT and PV error histogram data in Fig. 6, which shows a non-Gaussian pattern. The error scenario for the network load is the only one synthesised from Gaussian distribution with zero mean and a standard deviation 5% of the base curves, which is at a magnitude in day-ahead commercial or industrial building load forecasting errors [56]. Ref. [57] showed that the Gaussian distribution can fit the load forecasting error well. We set the correlation of the load forecasting error among buses as 0.5. In the 'LowRES' case the Gaussian error of load can be dominant, but the real-world non-Gaussian RES error data will be dominant in the 'HighPV' and 'HighWT' cases.

Based on the forecasting error statistics, we generate 500 samples per time step of the scheduling horizon $[T]$ as our training set, which is for constructing our proposed jointly reliable AFAPR and other benchmarks. Note that our proposed method ($\mathcal{D}$ in the 'Offline Dataset Gathering Stage') is implemented based on part of the training set, e.g., only randomly picked 50 in Section IV-F2. We further generate a test set containing 300 samples per time step for the out-of-sample cost and reliability evaluation of different models.
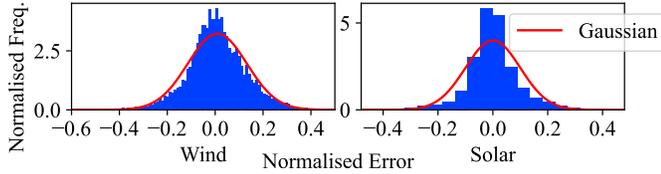
Figure 6. Histogram of the real-world wind and solar forecasting error.

Figure 7. Visualisation of VB parameters of 5 of the $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenarios in $\mathcal{D}$ in the 'LowRES' case.

B. Benchmark Models

The benchmarks start from the original form of the reliable AFAPR $\Omega^{\text{Agg}}(\epsilon)$ in (24). The JCC in (24) has a large size, so the known exact reformulation technique like [24] becomes intractable. Bonferroni approximation is commonly applied in the existing literature [4], [20] to approximate the intractable JCC by ICCs with a lower risk level $\epsilon_I = \epsilon/N^J$, where N^J is the number of constraints in the JCC. Note that the equality constraints (10) and (20) in the JCC will lead to a zero constraint satisfaction probability when the random variables have a probability density function (PDF) with respect to the Lebesgue measure (e.g., the Gaussian PDF). To deal with the ill-posed equality constraints in the JCC, we replace the involved uncertain variables with their deterministic forecasts and hereby lift (10) and (20) out of the JCC. Our benchmark 1 (B1) used this setting under the Bonferroni approximation.

Our benchmark 2 (B2) follows [16]. The equality constraint (20) is tackled the same as B1. To reduce the size of JCC and take the randomness of NL on the aggregated power into account, [16] proposed replacing the uncertain parameters in the individual DER active power constraints (1) and (9) with their forecasts and hereby lift them out of the JCC; then the following two constraints for the summed power of uncertain DERs and NLs are added into the JCC:

$$\sum_{i_{d'},\psi}^{N^{d'}} p_t^{d'} - \sum_{i_{\text{NL}},\psi}^{N^{\text{NL}}} \ddot{p}_t^{\text{NL,fore}} \leq \sum_{i_{d_{\text{R}}},\psi}^{N^{D_{\text{R}}}} \overline{p}_t^{d_{\text{R}}} - \sum_{i_{\text{CL}},\psi}^{N^{\text{CL}}} \sigma_{\min}\ddot{p}_t^{\text{CL}} - \sum_{i_{\text{NL}},\psi}^{N^{\text{NL}}} \ddot{p}_t^{\text{NL}}$$
(41)

$$\sum_{i_{d'},\psi}^{N^{d'}} p_t^{d'} - \sum_{i_{\text{NL}},\psi}^{N^{\text{NL}}} \ddot{p}_t^{\text{NL,fore}} \geq \sum_{i_{\text{CL}},\psi}^{N^{\text{CL}}} \sigma_{\max}\ddot{p}_t^{\text{CL}} - \sum_{i_{\text{NL}},\psi}^{N^{\text{NL}}} \ddot{p}_t^{\text{NL}}$$
(42)

where we have $d' \in [\text{PV}, \text{WT}, \text{CL}]$ and $d_{\text{R}} \in [\text{PV}, \text{WT}]$. The phase index ψ and bus index i are hidden for simplicity. Constraint (41) means the summed power scheduling of all the uncertain DERs plus the forecasted NL should be smaller than the case when the generation reaches the maximum and the load reaches the minimum (these maximum and minimum are uncertain parameters); constraint (42) is for the other side. The Bonferroni approximation is then applied to convert the JCC with ICCs under ϵ_I, which is less conservative than B1 since the size of JCC in (24) is reduced due to the removal of individual uncertain DERs' active power constraints.

We also include benchmarks converting the JCC in (24) into ICCs with the same risk level ϵ, which could lead to unreliable results. B1 and B2 under the direct ICC modelling with the risk level ϵ are denoted as B1unsafe and B2unsafe.

Among all benchmark models, the Gaussian distribution is used to model the uncertain parameters (even for our non-Gaussian RES distribution). This Gaussian setting is common as it enables an efficient transformation of ICCs into solvable counterparts based on the corresponding percentiles. All 500 training data are used to estimate the Gaussian distribution for different uncertain parameters in the benchmark models.
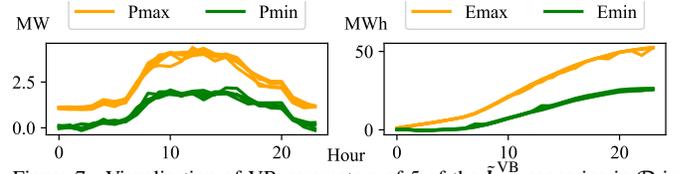
C. Settings of the Proposed Method

As our DERs only have energy and power constraints, the VB polytope approximator is applied. The bound shrinking method in [6] is used to obtain the VB parameters as described in Section III-A1. We then collect $\mathcal{D}$ based on the process in Algorithm 1: 50 randomly selected data samples from the training set are used to generate $\boldsymbol{b}_{\text{hist}}^{\text{VB},i}$ with $i \in \{1, \cdots, 50\}$, simulating those collected over 50 historical operation horizons. We use the forecasts of uncertain DERs to derive $\boldsymbol{b}_{\text{fore}}^{\text{VB}}$ for the scheduled horizon and $\boldsymbol{b}_{\text{fore, hist}}^{\text{VB},i}$ for historical horizons. The subtraction between those $\boldsymbol{b}_{\text{hist}}^{\text{VB},i}$ and $\boldsymbol{b}_{\text{fore, hist}}^{\text{VB},i}$ lead to the 50 historical $\tilde{\boldsymbol{b}}_{\text{error}}^{\text{VB}}$ scenarios ($\mathcal{D}_{\text{error}}$), which, combined with $\boldsymbol{b}_{\text{fore}}^{\text{VB}}$, results in $\mathcal{D}$. Fig. 7 displays 5 of the collected $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenarios in $\mathcal{D}$ for the 'LowRES' case. A $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenario consists of power and energy boundary parameters as (29). Each pair of the 'Pmax' and 'Pmin' are the power boundaries of aggregated active power that can be adjusted within for each time step, while each pair of 'Emax' and 'Emin' represents the energy bounds for the accumulated aggregated active power. Unlike a real battery, the energy boundaries of VB are derived to closely capture the accumulated energy generation and consumption of aggregated DERs; thus display an increasing pattern.

The bound shrinking method [6] requires solving a series of MILPs, which could be time-consuming when the problem scale is large. In addition, our case studies simulate the entire scenario generation process (mentioned above) that is carried out offline in the real-world implementation. To reduce simulation efforts, we relax the stopping bound and set the maximum solution time of each bound shrinking MILP to 40s in case studies for scheduling evaluation after Section IV-F, where we have $T = 24$ and a large number of DERs.

The radius θ in our proposed WDRJCC reformulation needs to be set prior to solving the problem. A larger θ leads to a larger ambiguity set and thus a more conservative AFAPR, so the selection criterion can be finding the smallest θ while providing a reliable solution, by 5-fold cross-validation (CV). For each fold, 20% data in $\mathcal{D}$ is held out and a solution is derived based on the remaining 80% data. The reliability score in this fold is estimated by the proportion of the held-out $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenarios whose corresponding polytope $\Omega_{\text{Approx}}^{\text{Agg}}$ (27) contains the solution. The 20% percentile of the 5 reliability scores after the CV run for a θ candidate is its final reliability score. The smallest θ with a score no less than the desired reliability level is selected as the best radius. Finally, to reduce the effect of randomness, all the subsequent results of our proposed method are averaged over 5 random runs.
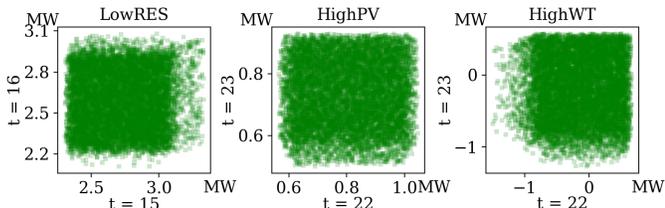
Figure 8. The projected 10,000 Monte-Carlo uniform samples of the proposed MILP-based AFAPR in (33).
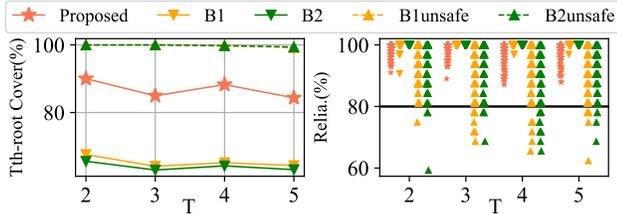
Figure 9. Evaluation for the whole AFAPR in conservativeness and inner-approximation. The black horizontal line is the desired reliability level $1 - \epsilon$.

D. Visualisation of the AFAPR

To facilitate understanding, we uniformly sampled 10,000 points from the proposed AFAPR in Eq. (33) with $T = 24$ and visualised them in Fig. 8. Each point corresponds to a 24-dim vector $\boldsymbol{p}^{\text{Agg}}$ representing the aggregated power across $T = 24$ time steps. To visualise these high-dimension points, we project them to two adjacent time steps in Fig. 8. The varying density of points suggest that our proposed MILP-based AFAPR defines a region more complicated than boxes. This shape can be attributed to the VB polytope (Section III-A1) we applied, which involves energy constraints that cut the high-dimensional boxes defined by the power constraints. At the same time, all these projections seem to be convex, partially explaining the computational tractability of our proposed MILP-based AFAPR in Eq. (33).

E. Evaluation of the Whole AFAPR

Recall our objective (25) is to find a less conservative inner approximation to $\Omega^{\text{Agg}}(\epsilon)$; thus, this case study evaluates the conservativeness and whether the inner approximation is achieved from the perspective of the whole AFAPR. We first use the Monte Carlo method to uniformly sample 100 points from $\Omega^{\text{Agg}}(\epsilon)$ (constructed by out-of-sample DER scenarios) and 100 points from the proposed AFAPR $\Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta)$ in (33) and the other four benchmarks. Each point corresponds to a possible $\boldsymbol{p}^{\text{Agg}}$, i.e., the aggregated power across the scheduling horizon $[T]$. We then define two metrics: 1)'Coverage': the proportion of samples from the true $\Omega^{\text{Agg}}(\epsilon)$ that are covered by the proposed method and benchmarks. A higher coverage indicates a lower conservativeness; 2) 'Reliability': if the proposed method and benchmarks contain "unsafe points" outside the true $\Omega^{\text{Agg}}(\epsilon)$. The inner approximation is achieved only if there are no "unsafe points".

The applied linear network model may have limited accuracy, especially because the AFAPR represents the power range under various operating conditions. Therefore, the Z-Bus nonlinear power flow simulation [58] is used to check the violation of network constraints over the sampling from $\Omega^{\text{Agg}}(\epsilon)$ and the evaluation of "Reliability". Results under 'HighWT' and ϵ=0.2 for various T are given in Fig. 9. Note

that the coverage (corresponding to volume) changes exponentially with the increase of T, so a decreasing trend can still be observed even when there is a high approximation accuracy for each dimension; we therefore plot the Tth-root of the coverage. We can see that the proposed method has a Tth-root of the coverage rate around 90% for all T, significantly higher than B1 and B2; from the right column of the figure, we see that all samples from the proposed method have reliability above the desired 80% (ϵ=0.2), meaning that our proposed method achieves the objective: a less conservative inner approximation to $\Omega^{\text{Agg}}(\epsilon)$. Although B1unsafe and B2unsafe lead to around 100% coverage, they cover less reliable $\boldsymbol{p}^{\text{Agg}}$ samples, i.e., "unsafe points" outside $\Omega^{\text{Agg}}(\epsilon)$.

F. Scheduling Evaluation

1) An Alternative Metric: In many power system scheduling problems, the scheduling horizon can be long, e.g., 24 time steps with an hourly resolution for day-ahead scheduling. The evaluation method in the previous section cannot be used for long time horizons due to the excessive complexity of Monte Carlo sampling. Note that the reliable AFAPR is used for scheduling problems, and a better AFAPR approximation tends to yield scheduling solutions with better optimality (less conservative) while still ensuring the desired solution reliability. Therefore, to assess the approximation quality, we can evaluate the performance of the optimal solution $\boldsymbol{p}^{\text{Agg}*}$ to the following DA cost minimisation problem with hourly resolution and $T = 24$ under the UK DA market price $\boldsymbol{\pi}^{\text{DA}}$:

$$\boldsymbol{p}^{\text{Agg}*} = \arg\min \quad \boldsymbol{\pi}^{\text{DA}\top} \boldsymbol{p}^{\text{Agg}} + \mathcal{C}'(\boldsymbol{p}^{\text{Agg}})$$
$$\text{s.t.} \quad \boldsymbol{p}^{\text{Agg}} \in \Omega_{\text{WDR}}^{\text{Agg}}(\epsilon, \theta) \text{ in (33)} \tag{43}$$

where $\mathcal{C}'(\boldsymbol{p}^{\text{Agg}})$ is the aggregated DER cost derived by a piecewise-linear method in Section III-D. Based on [3], one can also co-optimise the profit on reserve provision by adding $\overline{\boldsymbol{r}}^{\top}(\overline{\boldsymbol{p}}^{\text{Agg}} - \boldsymbol{p}^{\text{Agg}}) + \underline{\boldsymbol{r}}^{\top}(\boldsymbol{p}^{\text{Agg}} - \underline{\boldsymbol{p}}^{\text{Agg}})$ to the objective and $\underline{\boldsymbol{p}}^{\text{Agg}}, \overline{\boldsymbol{p}}^{\text{Agg}} \in \Omega_{\text{WDR}}^{\text{Agg}}$ to the constraints. $\overline{\boldsymbol{r}}$ and $\underline{\boldsymbol{r}}$ are the reward coefficients for the upward and downward flexibility reserve for $[T]$. $\overline{\boldsymbol{p}}^{\text{Agg}}$ and $\underline{\boldsymbol{p}}^{\text{Agg}}$ are the maximum and minimum available power defining the size of the reserve.

We evaluate the performance of the optimal DA scheduling $\boldsymbol{p}^{\text{Agg}*}$ from two perspectives: 1) Out-of-sample reliability: The disaggregation problem (38) is performed under each of the 300 out-of-sample DER scenarios. The nonlinear Z-Bus power flow simulation is implemented to check the network constraints of the disaggregated results. The out-of-sample reliability is defined as the proportion of the 300 out-of-sample DER scenarios where the disaggregation is successful and no network constraints are violated. As our scheduling horizon is for the next whole day, 300 scenarios correspond to roughly one-year running. The DA scheduling $\boldsymbol{p}^{\text{Agg}*}$ is considered reliable as long as the out-of-sample reliability is greater than or equal to the desired level $1-\epsilon$. 2) Out-of-sample cost: After solving the disaggregation problem (38) under each of the 300 out-of-sample DER scenarios, we can get the actual cost as $\boldsymbol{\pi}^{\text{DA}\top} \boldsymbol{p}^{\text{Agg}*} + \mathcal{C}(\boldsymbol{x})$. The out-of-sample cost is the average of these actual costs over the 300 out-of-sample scenarios.

2) Results: We plot the out-of-sample cost and out-of-sample reliability results under four desired reliability levels $(1 - \epsilon)$ in Fig. 10. For all three RES settings and all the
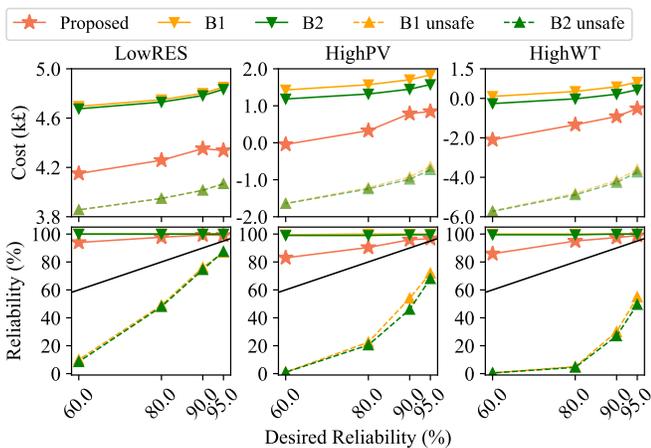
Figure 10. First row: minimised day-ahead costs; second row: test reliability, where the black diagonal line represents the desired reliability levels $1 - \epsilon$.
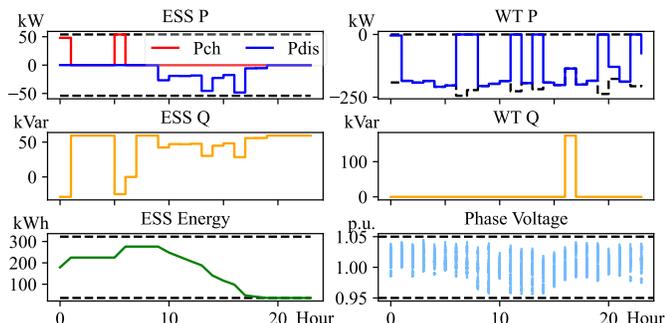
Figure 11. Successfully disaggregated DER power and nodal voltages (by Z-Bus nonlinear power flow) for the next T=24 scheduling horizon.

Figure 12. Sensitivity analysis for reliability of our proposed method for total ES sizes being 500kWh (E500), 1800kWh (E1800), and 3600kWh (E3600). The black diagonal lines represent the desired reliability levels $1 - \epsilon$.

Figure 13. Sensitivity analysis of the out-of-sample costs for total ES sizes being 500kWh (E500), 1000kWh (E1800), and 3600kWh (E3600).

desired reliability levels, our proposed method achieves lower costs than B1 and B2, and our proposed method meets the reliability requirement. On the other hand, B1unsafe and B2unsafe have lower costs than our proposed method but fail to meet reliability requirements for all cases. Overall, we see that our proposed method leads to less conservativeness while meeting the desired reliability.

We also plot a successful disaggregation result, i.e., the DER trajectories across $[T]$ reaching the scheduled $\boldsymbol{p}^{\text{Agg}*}$, in Fig. 11: there is no simultaneous charging and discharging power for ES. Also, all active power, reactive power, and network voltages (by the Z-Bus power flow) are within the limits.

G. Sensitivity Analysis

1) Size of Storage: The time-coupling DERs increase the approximation difficulties of $\Omega_{\text{Approx}}^{\text{Agg}}$ [6], which may affect the performance of our proposed method. Therefore, we evaluate our proposed method under three ES sizes: total size being 500 kWh (E500), 1800 kWh (E1800), and 3600 kWh (E3600). The maximum stored energy in the last setting can roughly support our evaluated IEEE-123 grid for one hour under the peak load time, so it is considered a large ES setting.

The out-of-sample reliability results are given in Fig. 12, which indicates that our proposed method is reliable under all ES sizes. Note that we use the inner polytope approximation, so a larger ES size leads to a more conservative approximation, thus increasing the test reliability. However, this increased conservativeness is minor compared to the reduced conservativeness by exactly reformulating the JCC in (33): Fig. 13 displays the out-of-sample costs of our proposed method, B1,

and B2 when increasing the ES size. Results for B1unsafe and B2unsafe are omitted as they are unreliable. We can see that our proposed method is still superior for all ES sizes.

2) Size of $\mathcal{D}$: This section evaluates the impact of the number of $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenarios in $\mathcal{D}$, i.e., the size of $\mathcal{D}$ or equivalently the number of binaries N of our proposed MILP-based AFAPR (33). Fig. 14 shows the out-of-sample cost, reliability, and the computing time of the scheduling problem (43) for the 'HighWT' case when the desired reliability is 80%. The costs of the reliable B1 and B2 are plotted for comparison. We see that an overly small number of $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenarios, i.e., 5, can lead to low cost but can be less reliable as shown in the reliability plot. Increasing the number of $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenarios from 80 to 120 results in no cost reduction (even a 2% increase) but a $1.7\times$ increase in computing time. Overall, based on Fig. 14, setting the number of $\tilde{\boldsymbol{b}}^{\text{VB}}$ scenarios (the number of binaries N) to within 30-80 may be more practical in use. We observe the same conclusion in both 'HighPV' and 'LowRES' cases as well.

It is worth noting the theoretical result in [44] that provides a priori lower-bound estimate of the probability that the true data distribution can be covered by the ambiguity set $\mathcal{F}(\theta)$ in (31), as a function of N (the size of $\mathcal{D}$) and the Wasserstein radius θ. However, in practice, setting N and θ based on the theoretical result can lead to over-conservativeness. It is still suggested to use hold-out method or the cross-validation method (as we did in Section IV-C) to make a prudent setting [44].

There is a also trade-off in choosing the polytope approximation in (26). A polytope with more edges, i.e., a larger $\text{Dim}(\boldsymbol{b}) = D$, can have a higher approximation accuracy to the true AFAPR but also lead to more constraints in our MILP-based AFAPR 33. A sensitivity analysis for the performance
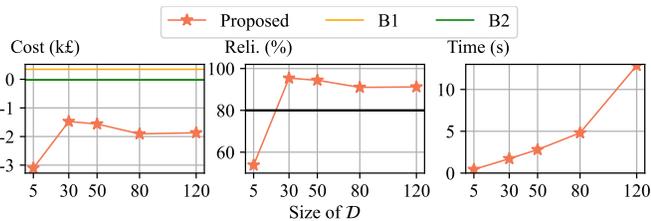
Figure 14. Sensitivity analysis for out-of-sample costs and reliability of the proposed method with increasing number of $\tilde{\boldsymbol{b}}^{\mathrm{VB}}$ scenarios in $\mathcal{D}$, i.e., the size of $\mathcal{D}$ or equivalently the number of binaries N of our proposed MILP-based AFAPR. The dark line represents the desired reliability.
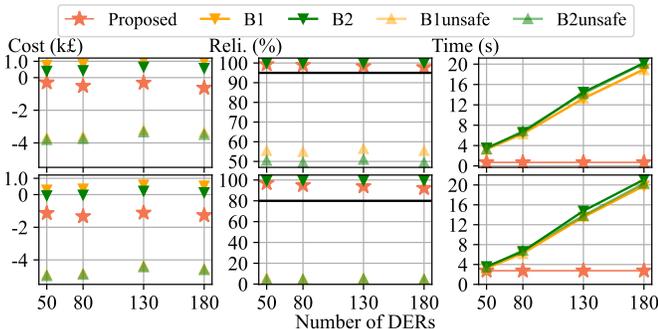
Figure 15. Out-of-sample costs, reliability, and scheduling time under various DER numbers. The black horizontal line is the desired reliability level $1 - \epsilon$.

versus the computing time can be performed to select the best polytope for specific applications. In general, our selected VB approximator has a good performance and low complexity for DERs with only power and energy constraints as demonstrated in our case studies and other related literature [6].

H. Scalability Test

We test the model performance under varying numbers of DERs from 50 to 180. We carry out the same cost-reliability analysis for 'highWT' under $\epsilon = 0.05$ (first row) and $\epsilon = 0.2$ (second row) in Fig. 15. The total capacity for each type of DERs is kept consistent for a better comparison. For the out-of-sample cost in the first column, our method consistently outperforms B1 and B2. Although B1unsafe and B2unsafe display lower costs, they fail to meet the desired reliability lines (dark horizontal lines) in the middle-column plot. Conversely, our method consistently yields reliable solutions, suggesting that our proposed method retains its superiority across different problem scales. We also plot the scheduling time, namely the time of solving the scheduling problem (43), using our proposed method and the benchmarks in the right-most column of Fig. 15. We can see that the complexity of our method is not affected by the number of DERs and the scheduling time is within seconds. The results suggest that our proposed jointly reliable AFAPR inner-approximation $\Omega_{\mathrm{WDR}}^{\mathrm{Agg}}(\epsilon)$ (33) can lead to computationally efficient upper-level scheduling problems.

I. Computing Time of the AFAPR Derivation

This section displays the computing time of the online process used to derive the reliable AFAPR (33) as shown in Algorithm 1. As discussed in Section III-E3, the only computationally demanding step is the derivation of $\boldsymbol{b}_{\mathrm{fore}}$ (Step 1) based on forecasts of uncertain DERs, by using existing methods that derive the inner polytope approximation of the AFAPR in a deterministic setting; other steps involve only simple algebra operations given the collected $\tilde{\boldsymbol{b}}_{\mathrm{error}}$ scenarios offline and are

computationally trivial. The bound shrinking method in [6] is applied to obtain $\boldsymbol{b}_{\mathrm{fore}}$, which involves solving a series of MILPs. Based on our relaxed MILP criterion described in Section IV-C, the total computing time in Step 1 is 150 to 300 seconds for DER numbers of 50 to 180 with an AMD Ryzen 9 5900X CPU. Note that our Algorithm 1 is capable of more efficient methods to perform Step 1; a possible choice is Ref. [48] that proposed a decomposition-based method that significantly accelerated the process in obtaining the inner polytope approximation with better performance than [6]. More importantly, as analysed in Section III-E3, our proposed online process to derive the jointly reliable AFAPR does not change the order of complexity compared with existing methods used to derive the inner polytope approximation of the AFAPR in a deterministic setting.

As discussed in Section III-D, it is important to derive the aggregated cost function. The piecewise-linear method [16] applied to derive the aggregate cost mapping involves solving $T \times N^P$ independent (parallelisable) linear programmes (LP), where $N^P = 5$ is the number of segments in the piecewise-linear function. Each LP only takes 1.5-4 seconds for DER numbers of 50 to 180, which is computationally tractable.

V. CONCLUSION

This paper proposes a novel surrogate polytope method for modelling the jointly reliable AFAPR over the entire scheduling horizon, taking into account linear network constraints and DER constraints including time-coupling effects. The proposed method starts from the polytope inner approximation with a fixed slope in a deterministic setting. The Wasserstein Distributional Robust Joint Chance Constraint (WDRJCC) is applied to the surrogate polytope rather than the low-level DER constraints and network constraints, leading to an exact reformulation as a tractable set of MILP constraints. We prove that the derived MILP constraints are an inner approximation to the original intractable jointly reliable AFAPR; we further show that our proposed method leads to a less conservative but still reliable AFAPR, by comparisons with four benchmarks under various case studies, with the nonlinear Z-Bus power flow simulation applied to validate the satisfaction of network constraints. The historical data size required is small, making the proposed method easy to deploy. The complexity of our MILP formulation is low, leading to computationally efficient upper-level scheduling. The size of the MILP does not increase with the network size nor with the number of DERs. Considering the possible intractability of MILPs in the worst cases, future work may explore convex but inner-approximation of the MILP formulation to further improve tractability. Finally, the linear network model is only accurate for certain operating conditions. There is space for an AFAPR that fully captures nonlinear network constraints or incorporates linearisations for multiple network conditions.

REFERENCES

[1] National Grid ESO, "Uk future energy scenarios." [Online]. Available: https://www.nationalgrideso.com/future-energy/future-energy-scenarios

[2] S. Riaz and P. Mancarella, "Modelling and characterisation of flexibility from distributed energy resources," *IEEE Trans on Power Systems*, 2021.

[3] X. Chen, E. Dall'Anese, C. Zhao, and N. Li, "Aggregate power flexibility in unbalanced distribution systems," *IEEE Trans on Smart Grid*, 2020.

[4] J. Zhai *et al.*, "Distributionally robust joint chance-constrained dispatch for integrated transmission-distribution systems via distributed optimization," *IEEE Transactions on Smart Grid*, 2022.

[5] Y. Wen, Z. Hu, S. You, and X. Duan, "Aggregate feasible region of ders: Exact formulation and approximate models," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4405–4423, 2022.

[6] S. Wang and W. Wu, "Aggregate flexibility of virtual power plants with temporal coupling constraints," *IEEE Trans on Smart Grid*, 2021.

[7] M. S. Nazir *et al.*, "Inner approximation of minkowski sums: A union-based approach and applications to aggregated energy resources," in *2018 IEEE Conference on Decision and Control*, 2018, pp. 5708–5715.

[8] M. Heleno, R. Soares, J. Sumaili, R. Bessa, L. Seca, and M. A. Matos, "Estimation of the flexibility range in the transmission-distribution boundary," in *2015 IEEE Eindhoven PowerTech*, 2015, pp. 1–6.

[9] W. Lin, Z. Yang, J. Yu, L. Jin, and W. Li, "Tie-line power transmission region in a hybrid grid: Fast characterization and expansion strategy," *IEEE Transactions on Power Systems*, vol. 35, no. 3, 2019.

[10] H. Ma, R. Jiang, and Z. Yan, "Distributionally robust co-optimization of power dispatch and do-not-exceed limits," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 887–897, 2019.

[11] D. M. Gonzalez *et al.*, "Determination of the time-dependent flexibility of active distribution networks to control their tso-dso interconnection power flow," in *Power Systems Computation Conference (PSCC)*, 2018.

[12] J. M. Morales, S. Pineda, and Y. Dvorkin, "Learning the price response of active distribution networks for tso-dso coordination," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 2858–2868, 2021.

[13] S. Taheri, V. Kekatos, H. Veeramachaneni, and B. Zhang, "Data-driven modeling of aggregate flexibility under uncertain and non-convex device models," *IEEE Transactions on Smart Grid*, pp. 1–1, 2022.

[14] B. Cui, A. Zamzam, and A. Bernstein, "Network-cognizant time-coupled aggregate flexibility of distribution systems under uncertainties," in *2021 American Control Conference (ACC)*, 2021, pp. 4178–4183.

[15] A. Arrigo *et al.*, "Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation," *Europ. Journal of Operational Research*, 2022.

[16] S. Wang, W. Wu, Q. Chen, J. Yu, and P. Wang, "Stochastic flexibility evaluation for virtual power plant by aggregating distributed energy resources," *CSEE Journal of Power and Energy Systems*, 2022.

[17] X. Geng and L. Xie, "Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization," *Annual reviews in control*, vol. 47, 2019.

[18] M. Zhang, X. Zhu, and N. Lu, "A data-driven probabilistic-based flexibility region estimation method for aggregated distributed energy resources," *arXiv preprint arXiv:2110.07406*, 2021.

[19] Z. Yi, Y. Xu, W. Gu, L. Yang, and H. Sun, "Aggregate operation model for numerous small-capacity distributed energy resources considering uncertainty," *IEEE Transactions on Smart Grid*, 2021.

[20] Y. Ding, T. Morstyn, and M. D. McCulloch, "Distributionally robust joint chance-constrained optimization for microgrids considering contingencies and renewable uncertainty," *IEEE Trans on Smart Grid*, 2022.

[21] M. Zhang, Y. Xu, X. Shi, and Q. Guo, "A fast polytope-based approach for aggregating large-scale electric vehicles in the joint market under uncertainty," *IEEE Transactions on Smart Grid*, pp. 1–1, 2023.

[22] Y. Wen and Z. Hu, "A novel inner approximation method for the aggregated temporally-coupled power flexibility of distributed energy resources," *arXiv preprint arXiv:2303.01691*, 2023.

[23] N. Ho-Nguyen, F. Kılınç-Karzan, S. Küçükyavuz, and D. Lee, "Distributionally robust chance-constrained programs with rhs uncertainty under wasserstein ambiguity," *Mathematical Programming*, 2021.

[24] H. Liu, J. Qiu, and J. Zhao, "A data-driven scheduling model of virtual power plant using wasserstein distributionally robust optimization," *International Journal of Electrical Power & Energy Systems*, 2022.

[25] H. Früh *et al.*, "Coordinated vertical provision of flexibility from distribution systems," *IEEE Trans. on Power Systems*, 2023.

[26] J. Silva *et al.*, "Estimating the active and reactive power flexibility area at the tso-dso interface," *IEEE Trans. on Power Systems*, 2018.

[27] M. Zhang, Y. Xu, and H. Sun, "Optimal coordinated operation for a distribution network with virtual power plants considering load shaping," *IEEE Transactions on Sustainable Energy*, 2023.

[28] Z. Yi, Y. Xu, J. Zhou, W. Wu, and H. Sun, "Bi-level programming for optimal operation of an active distribution network with multiple virtual power plants," *IEEE Transactions on Sustainable Energy*, 2020.

[29] M. Braun, *Provision of ancillary services by distributed generators: Technological and economic perspective*. Fraunhofer IWES, 01 2009.

[30] X. Chen and N. Li, "Leveraging two-stage adaptive robust optimization for power flexibility aggregation," *IEEE Trans. on Smart Grid*, 2021.

[31] K. Garifi, K. Baker, D. Christensen, and B. Touri, "Convex relaxation of grid-connected energy storage system models with complementarity constraints in dc opf," *IEEE Trans. on Smart Grid*, 2020.

[32] A. Bernstein and E. Dall'Anese, "Linear power-flow models in multi-phase distribution networks," in *ISGT-Europe*. IEEE, 2017.

[33] A. Bernstein *et al.*, "Load flow in multiphase distribution networks: Existence, uniqueness, non-singularity and linear models," *IEEE Transactions on Power Systems*, 2018.

[34] M. Cai, R. Yang, and Y. Zhang, "Iteration-based linearized distribution-level locational marginal price for three-phase unbalanced distribution systems," *IEEE Transactions on Smart Grid*, 2021.

[35] S. Mhanna and P. Mancarella, "An exact sequential linear programming algorithm for the optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 666–679, 2021.

[36] Z. Yang *et al.*, "A linearized opf model with reactive power and voltage magnitude: A pathway to improve the mw-only dc opf," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1734–1745, 2018.

[37] C. Heinrich, P. Fortenbacher, A. Fuchs, and G. Andersson, "Pv-integration strategies for low voltage networks," in *ENERGYCON*, 2016.

[38] A. Prékopa, *Stochastic programming*. Springer, 2013, vol. 324.

[39] Q. Zhai, X. Li, X. Lei, and X. Guan, "Transmission constrained uc with wind power: An all-scenario-feasible milp formulation with strong nonanticipativity," *IEEE Trans on Power Systems*, 2016.

[40] Q. Gao, Z. Yang, W. Yin, W. Li, and J. Yu, "Internally induced branch-and-cut acceleration for unit commitment based on improvement of upper bound," *IEEE Transactions on Power Systems*, 2022.

[41] A. Agarwal and L. Pileggi, "Large scale multi-period optimal power flow with energy storage systems using differential dynamic programming," *IEEE Transactions on Power Systems*, vol. 37, no. 3, 2022.

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

[43] S. Barot and J. A. Taylor, "A concise, approximate representation of a collection of loads described by polytopes," *International Journal of Electrical Power & Energy Systems*, vol. 84, pp. 55–63, 2017.

[44] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, 2018.

[45] F. Capitanescu, "Computing cost curves of active distribution grids aggregated flexibility for tso-dso coordination," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 2381–2384, 2024.

[46] C. Feng, Q. Chen, Y. Wang, P.-Y. Kong, H. Gao, and S. Chen, "Provision of contingency frequency services for virtual power plants with aggregated models," *IEEE Transactions on Smart Grid*, 2023.

[47] I.-I. Avramidis, "A first "stab" at defining temporal flexibility-to-cost mapping," *IEEE Transactions on Sustainable Energy*, 2023.

[48] Z. Li, M. Liu, M. Xie, and J. Zhu, "Robust optimization approach with acceleration strategies to aggregate an active distribution system as a virtual power plant," *International Journal of Electrical Power & Energy Systems*, vol. 142, p. 108316, 2022.

[49] K. P. Schneider *et al.*, "Analytic considerations and design basis for the ieee distribution test feeders," *IEEE Trans on power systems*, 2017.

[50] T. Morstyn *et al.*, "Open: An open-source platform for developing smart local energy system applications," *Applied Energy*, vol. 275, 2020.

[51] S. Pfenninger and I. Staffell, "Renewables.ninja." [Online]. Available: https://www.renewables.ninja/

[52] NREL, "End-Use Load Profiles for the U.S. Building Stock," 10 2021. [Online]. Available: https://data.openei.org/submissions/4520

[53] B.-M. Hodge *et al.*, "Wind power forecasting error distributions: An international comparison," NREL, Tech. Rep., 2012.

[54] M. David, M. A. Luis, and P. Lauret, "Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data," *International Journal of Forecasting*, vol. 34, no. 3, pp. 529–547, 2018.

[55] Y. Zhang *et al.*, "Day-ahead power output forecasting for small-scale solar pv electricity generators," *IEEE Trans. on Smart Grid*, 2015.

[56] J. G. Jetcheva, M. Majidpour, and W.-P. Chen, "Neural network model ensembles for building-level electricity load forecasts," *Energy and Buildings*, vol. 84, pp. 214–223, 2014.

[57] J. Cheng *et al.*, "Evaluating the spatial correlations of multi-area load forecasting errors," in *PMAPS*, 2016, pp. 1–6.

[58] M. Bazrafshan and N. Gatsis, "Comprehensive modeling of three-phase distribution systems via the bus admittance matrix," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2015–2029, 2018.

Yihong Zhou received the B.Eng. degree in Electrical Engineering and its Automation from North China Electric Power University, Beijing, China, in 2020, and the M.S. degree in Artificial Intelligence from the University of Edinburgh, Edinburgh, U.K., in 2021. He is currently a PhD student with the University of Edinburgh, Edinburgh, U.K. His research interests include power system flexibility, aggregation of distributed energy resources, local electricity markets, energy forecasting, and machine learning and optimisation methods in smart grids.

Chaimaa Essayeh received the M.Sc. degree in communication systems and Ph.D. degree in computer science from l'Institut National des Postes et Telecommunications (INPT), Rabat, Morocco, in 2014 and 2019. She was a Research Associate at the School of Engineering, University of Edinburgh, U.K. She is now a Research Fellow at the Department of Engineering, Nottingham Trent University, U.K. Her research interests focus on market designs for smart local energy systems.

Thomas Morstyn (Senior Member, IEEE) received the B.Eng. degree (Hon.) in electrical engineering from the University of Melbourne, Melbourne, Australia, in 2011, and the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, NSW, Australia, in 2016. He is Associate Professor in Power Systems with the Department of Engineering Science at the University of Oxford. His research is focused on the design of control systems and markets to enable the large-scale integration of distributed power system flexibility. He is an Associate Editor for the IEEE Transactions on Power Systems.