

Regime-Based Mixture Ambiguity Sets for Distributionally Robust Chance-Constrained Optimal Load Shedding

Qiaoyi Wang

Chongqing University of Education

Wei Liu

Beijing Normal University at Zhuhai

Minxin Shu

Chongqing University of Education

Feiyu Chen

Chongqing University of Education

Zongxin Zhang

Chongqing University of Education

Bo Yu

yubo@dlut.edu.cn

Chongqing University of Education

Research Article

Keywords: mixture ambiguity sets, distributionally robust optimization, optimal load shedding, error regimes, box approximation

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10129227/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Regime-Based Mixture Ambiguity Sets for Distributionally Robust Chance-Constrained Optimal Load Shedding

Qiaoyi Wang^{1†}, Wei Liu^{2†}, Minxin Shu^{1†}, Feiyu Chen^{1*},
Zongxin Zhang¹, Bo Yu^{1*}

¹National Center for Applied Mathematics in Chongqing, Chongqing
Normal University, Chongqing, China.

²Beijing Normal University at Zhuhai, Zhuhai, China.

*Corresponding author(s). E-mail(s): chenfeiyu@cqu.edu.cn;
yubo@dlut.edu.cn;

Contributing authors: 2024110518099@stu.cqu.edu.cn;
liuwei@bnuz.edu.cn; shuxin2357@gmail.com; 2484686122@qq.com;

[†]These authors contributed equally to this work.

Abstract

We develop a regime-based mixture-ambiguity model for distributionally robust chance-constrained optimal load shedding. Historical nodal net-injection errors are clustered into operational error regimes, and each regime is assigned a local Wasserstein ambiguity set centered at a representative empirical distribution and supported on a regime-conditional compact box. The mixture construction reduces the conservatism caused by enclosing heterogeneous errors in a single global ambiguity set and avoids the worst-case saturation induced by a common support under nonzero Wasserstein radii. Predictive and rolling responsibilities update the nominal regime weights online, and an outer $\mathbf{L}_1$ ambiguity set protects against weight misspecification. Under an $\mathbf{L}_1$ ground metric, the robust CVaR approximation admits a linear deterministic reformulation, while representative atoms control the number of sample-dependent constraints. In daily rolling-origin backtests on RBTS, RTS-79, and a screened RTS-GMLC, the framework restores out-of-sample security where deterministic dispatch violates the chance constraint in a majority of hours, while the regime structure recovers much of the conservatism that a single global Wasserstein ball imposes—up to a quarter of the scheduled load shedding on RBTS—at matched reliability. A sweep over the Wasserstein radius confirms the predicted saturation effect, and a limited-data

study shows the regime-localized ball to be necessary for out-of-sample reliability under heavy-tailed errors. The large-system study demonstrates the tractability of the screened implementation and a diagnostics-driven path toward full-network security. Benchmarked against moment-based and Gaussian chance-constraint paradigms, the framework is the most economical secure policy on the heavy-tailed and data-scarce regimes, where a moment-only set is infeasibly conservative and the Gaussian rule is reliable only under its parametric assumption.

Keywords: mixture ambiguity sets, distributionally robust optimization, optimal load shedding, error regimes, box approximation

1 Introduction

The increasing penetration of renewable energy sources, particularly wind power, has increased operational uncertainty in modern power systems. During severe weather events or large forecast deviations, available reserves may be depleted, making load shedding an unavoidable last-resort recourse for maintaining system integrity. These operating conditions have motivated research into optimal load shedding (OLS), reliability assessment, and uncertainty-aware dispatch formulations [1–6]. Chance-constrained (CC) models are widely used to limit the occurrence of physical constraint violations [4, 5, 7]. Their effectiveness, however, relies on accurate knowledge of the uncertainty distribution, which is rarely available in practice [8, 9]. Distributionally robust optimization (DRO) addresses this limitation by optimizing against a family of plausible distributions and leads naturally to distributionally robust chance-constrained (DRCC) models [10–14]. In power-system applications, data-driven DRCC and DRO formulations have been developed for approximate AC-OPF, economic dispatch, energy-and-reserve scheduling, and stochastic OPF under renewable uncertainty [15–18]. Among various DRO methodologies, Wasserstein-based ambiguity sets have gained prominence due to their optimal-transport interpretation, theoretical guarantees, and out-of-sample performance [19–21]. Most existing DRCC models construct a single global ambiguity set, implicitly assuming that the uncertainty follows a homogeneous distribution. This global abstraction can lead to overly conservative dispatch decisions and unnecessary load shedding, as the model prepares for extreme scenarios that are physically implausible under the specific prevailing conditions. The central modeling issue addressed here is that a single Wasserstein ball may mix heterogeneous error regimes and thereby exaggerate the worst-case uncertainty faced in a particular operating window. Although error generation is influenced by weather conditions, we do not explicitly learn the complex mapping from weather to error distributions. Instead, we identify statistical operational error regimes directly from historical net-injection errors using a Gaussian mixture model, following the classical mixture-model and EM-estimation literature [22, 23]. This choice avoids introducing an additional black-box forecasting layer and preserves the clarity of the DRO reformulation. It is also consistent with contextual and data-adaptive robust optimization viewpoints, where side information or regime information is used to localize uncertainty descriptions [24].

The main contributions of this paper are summarized as follows:

- **A regime-wise mixture ambiguity set that removes the structural conservatism of a single global Wasserstein ball.** A global ball must simultaneously hedge against all heterogeneous error regimes and is therefore forced to prepare for worst cases that are physically implausible under the prevailing operating condition. By attaching local Wasserstein sets to statistically identified net-injection error regimes, the proposed model recovers a meaningful part of this lost economic efficiency—scheduling at or below the global benchmark in essentially every hour and materially below it where regime heterogeneity is pronounced—while preserving the out-of-sample security guarantee.
- **Identification and proof of a saturation failure mode of mixture ambiguity sets.** We prove (Proposition 1) that under a *common* support and a positive Wasserstein radius the regime mixture collapses back to the single global worst case, nullifying regime separation, and that *regime-conditional* support boxes are precisely what preserves the mixture’s advantage as a strictly positive, radius-stable margin. A sweep over the Wasserstein radius confirms the predicted collapse-versus-plateau behavior. To our knowledge this saturation effect and its support-based remedy have not been articulated before.
- **A tractable and scalable exact reformulation.** We derive an exact linear reformulation of the robust joint-CVaR approximation under regime-conditional box supports and an L_1 Wasserstein ground metric, and introduce regime-wise representative atoms that bound the number of sample-dependent constraints. The resulting models solve in seconds per dispatch and extend, via constraint screening, to a large system (RTS-GMLC).
- **Multi-system daily rolling-origin validation.** On RBTS, RTS-79, and RTS-GMLC we show that the framework restores out-of-sample security where deterministic dispatch violates the chance constraint in a majority of hours, operates at matched security with reduced shedding relative to the global benchmark, and remains computationally tractable at scale. A limited-data study isolates the value of the robust layer itself: under heavy tails and scarce data the purely empirical (zero-radius) model violates the reliability target out of sample, whereas the regime-localized Wasserstein model preserves it. A direct comparison against moment-based and Gaussian chance-constraint paradigms shows the moment-only ambiguity set to be infeasibly conservative and the Gaussian rule to be economical only where its parametric assumption holds, while the regime mixture is the cheaper secure policy across the realistic heavy-tailed and data-scarce regimes. Predictive and rolling weight updates with an outer L_1 weight-ambiguity set provide an adaptive, misspecification-robust refinement.

Related positioning.

Most data-driven DRCC formulations for power-system dispatch hedge against a *single* ambiguity set—a moment-based set [10], a global Wasserstein ball [15–18], or a likelihood region [11]—constructed from the pooled error sample. Such a set implicitly treats the forecast-error distribution as homogeneous and must therefore

widen to cover every operating condition simultaneously. Contextual and data-adaptive robust optimization localize the uncertainty description using side information [24], but typically still yield one context-conditioned set per decision. Our formulation differs in two respects: it partitions the *error space* itself into statistically identified regimes and equips each with its own Wasserstein ball over a regime-conditional support, and it makes explicit—and then removes—a saturation failure mode in which a shared support collapses the regime mixture back onto the single global ball. This retains the optimal-transport guarantees of Wasserstein DRO [20, 21, 25] while recovering the economy that a global ball forgoes.

The remainder of this paper is organized as follows. Section 2 introduces the physical affine-control DC-OLS architecture and the mixture ambiguity formulation. Section 3 provides the deterministic reformulations. Section 4 details the data-driven parameter generation. Section 5 reports the numerical results, and Section 6 concludes the paper.

Notation

Throughout this paper, bold lowercase letters denote vectors and bold uppercase letters denote matrices. For a vector $\mathbf{v}$, $\|\mathbf{v}\|_1$ and $\|\mathbf{v}\|_\infty$ denote the L_1 and L_∞ norms, respectively, and $\|\cdot\|_*$ denotes the dual norm of the ground norm used in the Wasserstein distance. The symbol $\mathbf{1}$ denotes an all-one vector whose dimension is determined by context. When ambiguity may arise, we write $\mathbf{1}_B$, $\mathbf{1}_G$, or $\mathbf{1}_K$. For a finite set $\mathcal{A}$, $|\mathcal{A}|$ denotes its cardinality.

Sets and indices

$\mathcal{B}$	Set of buses.
$\mathcal{G}$	Set of conventional generators.
$\mathcal{W}$	Set of wind farms or renewable units.
$\mathcal{L}$	Set of transmission lines.
$\mathcal{M}$	Index set of all affine reserve and line-flow security constraints, $\{1, \dots, 2 \mathcal{G} + 2 \mathcal{L} \}$.
$\mathcal{M}_{\text{def}}$	Deficit-side subset of $\mathcal{M}$ (upward reserve and line-flow limits) on which the chance constraint is imposed.
$\mathcal{M}_+$	Augmented index set $\mathcal{M}_{\text{def}} \cup \{0\}$ used in the CVaR positive-part representation.
$\mathcal{K}$	Set of forecast-error regimes.
$\mathcal{I}_k$	Historical sample index set assigned to regime k by the hard GMM responsibility rule.
$\mathcal{I}_k^{\text{rep}}$	Representative sample index set retained for regime k .

System variables and parameters

$\mathbf{L}_t$	Baseline nodal load forecast at time t .
$\mathbf{f}_t$	Baseline renewable generation forecast at time t .
$\mathbf{g}$	Baseline conventional generation dispatch.
β	AGC participation-factor vector.
$\mathbf{r}^D, \mathbf{r}^U$	Downward and upward reserve-capacity vectors.
$\Delta \mathbf{L}$	Load-shedding vector.

$\mathbf{c}$	Scheduled renewable-curtailment vector, $0 \leq \mathbf{c} \leq \mathbf{f}_t$.
$\overline{\mathbf{F}}$	Transmission-line capacity vector.
M_G, M_W, M_B	PTDF matrices for generators, renewable units, and buses.
Uncertainty and ambiguity-set notation	
$\tilde{\boldsymbol{\omega}}$	Random nodal net-injection forecast-error vector.
$\boldsymbol{\omega}$	Generic realization or optimization variable in a support set.
$\tilde{\boldsymbol{\omega}}_i$	Historical forecast-error sample.
$\boldsymbol{\omega}_t$	Realized post-dispatch error at time t .
$\hat{\boldsymbol{\omega}}_t$	Point estimate of the upcoming forecast error at time t .
Ω	System-wide aggregate mismatch, $\Omega = \mathbf{1}_B^\top \tilde{\boldsymbol{\omega}}$.
Ξ_k	Regime-conditional compact support set of forecast errors in regime k .
Ξ^{glob}	Pooled compact support set used by the single global Wasserstein benchmark.
$\mathbf{c}_k, \mathbf{d}_k$	Center and half-width of the regime-conditional support box Ξ_k .
$\hat{P}_k^{\text{rep}}$	Representative empirical distribution of regime k .
$\mathcal{P}_k$	Local Wasserstein ambiguity set of regime k , supported on Ξ_k .
ϱ_k	Wasserstein radius of regime k .
$\mathbf{w}$	Candidate mixture-weight vector.
$\boldsymbol{\pi}_t$	Online nominal mixture-weight center.
δ_w	Radius of the outer L_1 weight ambiguity set.
Risk and reformulation notation	
ϵ	Chance-constraint risk level.
ε_c	Negligible curtailment-penalty weight ($\varepsilon_c \ll 1$) in the objective.
τ	CVaR auxiliary threshold variable.
$W_1(\cdot, \cdot)$	Type-1 Wasserstein distance.
$S_{\Xi_k}(\cdot)$	Support function of the regime-conditional support set Ξ_k .
$\ \cdot\ _1, \ \cdot\ _\infty$	L_1 and L_∞ norms.

The notation $\mathcal{P}(\Xi_k)$ denotes the set of all probability distributions supported on Ξ_k , and $S_{\Xi_k}(\mathbf{v}) = \sup_{\boldsymbol{\omega} \in \Xi_k} \mathbf{v}^\top \boldsymbol{\omega}$ denotes the support function of Ξ_k . The symbol $\mathbb{P}$ denotes the unknown true distribution, P_k denotes a candidate local distribution in regime k , $\mathcal{P}_k$ denotes the corresponding local ambiguity set, and $P = \sum_{k=1}^K w_k P_k$ denotes a candidate mixture distribution. Each historical sample is assigned to a unique regime by the hard responsibility rule

$$\kappa(i) = \arg \max_{k \in \mathcal{K}} \gamma_{ik}, \quad \mathcal{I}_k = \{i : \kappa(i) = k\}.$$

Consequently, all retained atoms $i \in \mathcal{I}_k^{\text{rep}} \subseteq \mathcal{I}_k$ are associated with the regime-conditional support Ξ_k . In the static model derivation, time subscripts are omitted for brevity. During rolling implementation, terms $\mathbf{L}$, $\mathbf{f}$, and $\boldsymbol{\pi}$ are replaced by their time-specific values $\mathbf{L}_t$, $\mathbf{f}_t$, and $\boldsymbol{\pi}_t$.

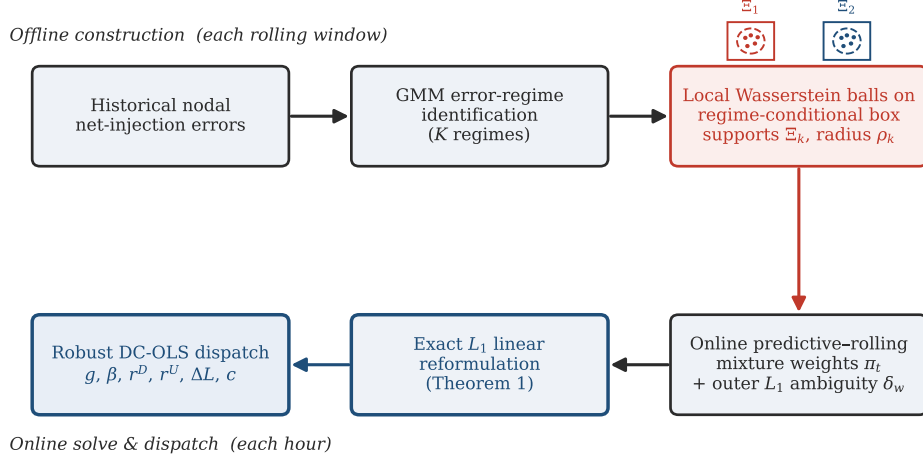


Fig. 1 Overview of the regime-mixture DRCC-OLS framework. *Offline* (top row, per rolling window): historical nodal net-injection errors are clustered into K error regimes by a Gaussian mixture model, and each regime receives a local Wasserstein ball over its own regime-conditional box support Ξ_k of radius ρ_k . *Online* (bottom row, per dispatch hour): predictive-rolling mixture weights π_t , protected by an outer L_1 weight-ambiguity set of radius δ_w , feed the exact L_1 linear reformulation of Theorem 1, which is solved for the robust DC-OLS dispatch. The regime-conditional supports $\Xi_1, \Xi_2, \dots$ prevent the mixture from saturating onto a single global ball (Proposition 1).

2 Problem Formulation: DC-OLS Architecture

Figure 1 gives a roadmap of the framework: the physical dispatch architecture and the mixture ambiguity set are introduced in this section, the deterministic reformulation in Section 3, and the data-driven construction of the regimes and weights in Section 4. Consider a power system with conventional generator set $\mathcal{G}$, wind-farm set $\mathcal{W}$, bus set $\mathcal{B}$, and transmission-line set $\mathcal{L}$. Let $\mathbf{L} \in \mathbb{R}^{|\mathcal{B}|}$ and $\mathbf{f} \in \mathbb{R}^{|\mathcal{W}|}$ denote the deterministic baseline forecasts of nodal load and wind-farm generation, respectively. The random nodal net-injection error vector $\tilde{\omega} \in \Xi \subset \mathbb{R}^{|\mathcal{B}|}$ jointly captures generation and load fluctuations. The component $\tilde{\omega}_b$ is defined as the aggregate wind-generation forecast error minus the load forecast error at bus b . Wind-generation surplus contributes positively and load increase contributes negatively. The system-wide aggregate mismatch Ω is defined by

$$\Omega = \mathbf{1}_B^\top \tilde{\omega}. \quad (1)$$

The primary decision vector is

$$x = (\mathbf{g}, \boldsymbol{\beta}, \mathbf{r}^D, \mathbf{r}^U, \Delta \mathbf{L}, \mathbf{c}), \quad (2)$$

containing the baseline generation $\mathbf{g}$, AGC participation factors $\boldsymbol{\beta}$, downward/upward reserve capacities $\mathbf{r}^D$ and $\mathbf{r}^U$, the load-shedding vector $\Delta \mathbf{L}$, and the scheduled renewable-curtailment vector $\mathbf{c} \in \mathbb{R}^{|\mathcal{W}|}$. The OLS objective minimizes the total shed

load $\mathbf{1}_B^\top \Delta \mathbf{L}$, augmented by a negligible penalty $\varepsilon_c \mathbf{1}_W^\top \mathbf{c}$ with $\varepsilon_c \ll 1$; scheduled curtailment is thus invoked only to preserve power-balance feasibility under forecast surplus (low-load, high-renewable hours) and does not distort the load-shedding objective. The deterministic feasible set $\mathcal{X}$ enforces the physical limits:

$$x \in \mathcal{X} \iff \begin{cases} \mathbf{g} - \mathbf{r}^D \geq \mathbf{g}^{\min}, & \mathbf{g} + \mathbf{r}^U \leq \mathbf{g}^{\max}, \\ \mathbf{1}_G^\top \mathbf{g} + \mathbf{1}_W^\top (\mathbf{f} - \mathbf{c}) = \mathbf{1}_B^\top (\mathbf{L} - \Delta \mathbf{L}), & \mathbf{1}_G^\top \boldsymbol{\beta} = 1, \\ \boldsymbol{\beta} \geq 0, & \mathbf{r}^D, \mathbf{r}^U \geq 0, & 0 \leq \Delta \mathbf{L} \leq \mathbf{L}, & 0 \leq \mathbf{c} \leq \mathbf{f}. \end{cases} \quad (3)$$

In real time, the affine recourse policy is $\tilde{g}_j(\tilde{\omega}) = g_j - \beta_j \Omega$ for $j \in \mathcal{G}$.

2.1 From Physical Limits to Unified Affine Constraints

The affine recourse policy must ensure that generator limits and transmission capacities are respected for any realization of $\tilde{\omega}$. The physical constraints can be expressed as follows.

Reserve constraints: For each generator $j \in \mathcal{G}$, the real-time output deviation $-\beta_j \Omega$ must lie within the scheduled reserve capacities:

$$-r_j^D \leq -\beta_j \Omega \leq r_j^U \iff \begin{cases} \beta_j \Omega - r_j^D \leq 0, \\ -\beta_j \Omega - r_j^U \leq 0. \end{cases} \quad (4)$$

Transmission constraints: Let $M_G \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{G}|}$, $M_W \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{W}|}$, and $M_B \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{B}|}$ be the power transfer distribution factor matrices. The line flows under the affine policy are

$$\mathbf{F}(\tilde{\omega}) = M_G(\mathbf{g} - \boldsymbol{\beta} \Omega) + M_B \tilde{\omega} + M_W(\mathbf{f} - \mathbf{c}) - M_B(\mathbf{L} - \Delta \mathbf{L}). \quad (5)$$

The flow limits $-\bar{\mathbf{F}} \leq \mathbf{F}(\tilde{\omega}) \leq \bar{\mathbf{F}}$ must hold almost surely. By substituting $\Omega = \mathbf{1}_B^\top \tilde{\omega}$ and collecting terms in $\tilde{\omega}$, each line $e \in \mathcal{L}$ yields two scalar inequalities:

$$(M_{B,e,*} - (M_{G,e,*} \boldsymbol{\beta}) \mathbf{1}_B^\top) \tilde{\omega} + M_{G,e,*} \mathbf{g} + M_{W,e,*}(\mathbf{f} - \mathbf{c}) - M_{B,e,*}(\mathbf{L} - \Delta \mathbf{L}) - \bar{F}_e \leq 0, \quad (6)$$

$$- (M_{B,e,*} - (M_{G,e,*} \boldsymbol{\beta}) \mathbf{1}_B^\top) \tilde{\omega} - (M_{G,e,*} \mathbf{g} + M_{W,e,*}(\mathbf{f} - \mathbf{c}) - M_{B,e,*}(\mathbf{L} - \Delta \mathbf{L})) - \bar{F}_e \leq 0. \quad (7)$$

All these random inequalities are affine in $\tilde{\omega}$ with coefficients that depend linearly on the decision variables x . Therefore they can be uniformly written as

$$\mathbf{a}_m(x)^\top \tilde{\omega} + b_m(x) \leq 0, \quad m \in \mathcal{M}, \quad (8)$$

where $\mathcal{M} = \{1, \dots, 2|\mathcal{G}| + 2|\mathcal{L}|\}$ indexes all reserve and transmission constraints.

2.2 Unified Random Constraint Definition

To ensure that unit capacities and transmission limits are respected under affine recourse, all random operational limits are written in the unified affine form:

$$\mathbf{a}_m(x)^\top \tilde{\boldsymbol{\omega}} + b_m(x) \leq 0, \quad m \in \mathcal{M}. \quad (9)$$

Here, the index set is $\mathcal{M} = \{1, \dots, 2|\mathcal{G}| + 2|\mathcal{L}|\}$. The parameters $\mathbf{a}_m(x) \in \mathbb{R}^{|\mathcal{B}|}$ and $b_m(x) \in \mathbb{R}$ are affine mappings of the decision variables x , explicitly defined across four categories:

For the downward reserve constraints ($j \in \mathcal{G}$):

$$\mathbf{a}_j(x) = \beta_j \mathbf{1}_B, \quad b_j(x) = -r_j^D. \quad (10)$$

For the upward reserve constraints ($j \in \mathcal{G}$):

$$\mathbf{a}_{|\mathcal{G}|+j}(x) = -\beta_j \mathbf{1}_B, \quad b_{|\mathcal{G}|+j}(x) = -r_j^U. \quad (11)$$

For the upper transmission limit of line $e \in \mathcal{L}$:

$$\mathbf{a}_{2|\mathcal{G}|+e}(x) = (M_{B,e,*} - (M_{G,e,*} \boldsymbol{\beta}) \mathbf{1}_B^\top)^\top, \quad (12)$$

$$b_{2|\mathcal{G}|+e}(x) = M_{G,e,*} \mathbf{g} + M_{W,e,*} (\mathbf{f} - \mathbf{c}) - M_{B,e,*} (\mathbf{L} - \Delta \mathbf{L}) - \bar{F}_e. \quad (13)$$

For the lower transmission limit of line $e \in \mathcal{L}$:

$$\mathbf{a}_{2|\mathcal{G}|+|\mathcal{L}|+e}(x) = - (M_{B,e,*} - (M_{G,e,*} \boldsymbol{\beta}) \mathbf{1}_B^\top)^\top, \quad (14)$$

$$b_{2|\mathcal{G}|+|\mathcal{L}|+e}(x) = - (M_{G,e,*} \mathbf{g} + M_{W,e,*} (\mathbf{f} - \mathbf{c}) - M_{B,e,*} (\mathbf{L} - \Delta \mathbf{L})) - \bar{F}_e. \quad (15)$$

Note that $M_{G,e,*} \boldsymbol{\beta}$ is a scalar, making $(M_{G,e,*} \boldsymbol{\beta}) \mathbf{1}_B^\top$ a $|\mathcal{B}|$ -dimensional row vector.

2.3 Error-Regime-Based Mixture Ambiguity Set

Since load shedding is driven by the under-supply (deficit) direction, we focus the distributional robustness on the deficit-critical components. Scheduled curtailment $\mathbf{c}$ absorbs baseline forecast surplus deterministically through the power balance in $\mathcal{X}$, so the over-generation (downward-reserve) direction is governed by deterministic scheduling and real-time curtailment rather than by the robust chance constraint. Accordingly, the chance constraint is imposed on the deficit-critical index set $\mathcal{M}_{\text{def}} = \{|\mathcal{G}| + 1, \dots, 2|\mathcal{G}| + 2|\mathcal{L}|\} \subset \mathcal{M}$, comprising the upward-reserve constraints and the two-sided line-flow limits, which govern the under-supply security that load shedding is meant to protect. The joint chance constraint is $\mathbb{P}(\max_{m \in \mathcal{M}_{\text{def}}} \{\mathbf{a}_m(x)^\top \tilde{\boldsymbol{\omega}} + b_m(x)\} \leq 0) \geq 1 - \epsilon$. To construct the robust CVaR approximation [26], we introduce the augmented index set $\mathcal{M}_+ = \mathcal{M}_{\text{def}} \cup \{0\}$ to encode the positive-part function. Define:

$$\tilde{\mathbf{a}}_0(x) = \mathbf{0}, \quad \tilde{b}_0(x, \tau) = 0, \quad (16)$$

$$\tilde{\mathbf{a}}_m(x) = \mathbf{a}_m(x), \quad \tilde{b}_m(x, \tau) = b_m(x) - \tau, \quad \forall m \in \mathcal{M}_{\text{def}}. \quad (17)$$

The positive-part maximum can then be written as:

$$\left(\max_{m \in \mathcal{M}_{\text{def}}} \{ \mathbf{a}_m(x)^\top \tilde{\boldsymbol{\omega}} + b_m(x) \} - \tau \right)^+ = \max_{m \in \mathcal{M}_+} \{ \tilde{\mathbf{a}}_m(x)^\top \tilde{\boldsymbol{\omega}} + \tilde{b}_m(x, \tau) \}. \quad (18)$$

The underlying distribution is modeled as a mixture $P = \sum_{k=1}^K w_k P_k$, where K is the number of error regimes and $\mathcal{K} = \{1, \dots, K\}$ is the regime index set. Each historical sample is assigned to a unique regime by

$$\kappa(i) = \arg \max_{k \in \mathcal{K}} \gamma_{ik}, \quad \mathcal{I}_k = \{i : \kappa(i) = k\},$$

and the retained atoms satisfy $\mathcal{I}_k^{\text{rep}} \subseteq \mathcal{I}_k$. Hence every representative atom $\hat{\boldsymbol{\omega}}_i$, $i \in \mathcal{I}_k^{\text{rep}}$, is evaluated under the corresponding regime-conditional support set Ξ_k . For each regime $k \in \mathcal{K}$, the local ambiguity set is a Wasserstein ball centered at a representative empirical distribution $\hat{P}_k^{\text{rep}}$ and supported on Ξ_k :

$$\mathcal{P}_k = \left\{ P_k \in \mathcal{P}(\Xi_k) : W_1(P_k, \hat{P}_k^{\text{rep}}) \leq \varrho_k \right\}, \quad (19)$$

where $\varrho_k \geq 0$ is the Wasserstein radius and $\hat{P}_k^{\text{rep}} = \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \delta_{\hat{\boldsymbol{\omega}}_i}$.

The mixture weights $\mathbf{w}$ belong to an outer L_1 ambiguity set $\mathcal{W}_t$ centered at $\bar{\boldsymbol{\pi}}$:

$$\mathcal{W}_t = \left\{ \mathbf{w} \in \mathbb{R}_+^K : \mathbf{1}_K^\top \mathbf{w} = 1, \|\mathbf{w} - \bar{\boldsymbol{\pi}}\|_1 \leq \delta_w \right\}. \quad (20)$$

Here, $\bar{\boldsymbol{\pi}}$ serves as a generic nominal center. In online operations, $\bar{\boldsymbol{\pi}}$ is replaced by $\boldsymbol{\pi}_t = (\pi_{1,t}, \dots, \pi_{K,t})$. The MA-DRCC-OLS formulation is:

$$\min_{x \in \mathcal{X}, \tau \in \mathbb{R}} \mathbf{1}_B^\top \Delta \mathbf{L} + \varepsilon_c \mathbf{1}_W^\top \mathbf{c} \quad (21)$$

$$\text{s.t.} \quad \tau + \frac{1}{\epsilon} \sup_{\mathbf{w} \in \mathcal{W}_t} \sum_{k=1}^K w_k \sup_{P_k \in \mathcal{P}_k} \mathbb{E}_{P_k} \left[\max_{m \in \mathcal{M}_+} \{ \tilde{\mathbf{a}}_m(x)^\top \tilde{\boldsymbol{\omega}} + \tilde{b}_m(x, \tau) \} \right] \leq 0. \quad (22)$$

3 Deterministic Reformulation

Theorem 1 *Given (x, τ) , define*

$$f(x, \tau, \boldsymbol{\omega}) = \max_{m \in \mathcal{M}_+} \{ \tilde{\mathbf{a}}_m(x)^\top \boldsymbol{\omega} + \tilde{b}_m(x, \tau) \}.$$

For each $k \in \mathcal{K}$, let $\Xi_k \subset \mathbb{R}^{|\mathcal{B}|}$ be a compact convex regime-conditional support set and let

$$\hat{P}_k^{\text{rep}} = \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \delta_{\hat{\boldsymbol{\omega}}_i}, \quad \mathcal{I}_k^{\text{rep}} \subseteq \mathcal{I}_k.$$

Define the local ambiguity set

$$\mathcal{P}_k = \left\{ P_k \in \mathcal{P}(\Xi_k) : W_1(P_k, \hat{P}_k^{\text{rep}}) \leq \varrho_k \right\}.$$

Let the outer weight ambiguity set be

$$\mathcal{W}(\bar{\pi}) = \left\{ \mathbf{w} \in \mathbb{R}_+^K : \mathbf{1}^\top \mathbf{w} = 1, \|\mathbf{w} - \bar{\pi}\|_1 \leq \delta_w \right\},$$

where $\bar{\pi} \in \mathbb{R}_+^K$, $\mathbf{1}^\top \bar{\pi} = 1$, and $\delta_w \geq 0$. Then the worst-case expectation evaluation problem

$$\Phi(x, \tau) := \sup_{\mathbf{w} \in \mathcal{W}(\bar{\pi})} \sum_{k=1}^K w_k \sup_{P_k \in \mathcal{P}_k} \mathbb{E}_{P_k}[f(x, \tau, \tilde{\omega})]$$

is equal to the optimal value of the following finite-dimensional problem:

$$\begin{aligned} \min_{\eta, \nu, \mathbf{t}, \boldsymbol{\lambda}, \mathbf{s}, \boldsymbol{\zeta}} \quad & \eta + \nu(\delta_w - 1) + \sum_{k=1}^K \bar{\pi}_k t_k \\ \text{s.t.} \quad & t_k + \eta - \nu \geq \lambda_k \varrho_k + \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} s_{ik}, \quad \forall k \in \mathcal{K}, \\ & s_{ik} \geq \tilde{b}_m(x, \tau) + S_{\Xi_k}(\tilde{\mathbf{a}}_m(x) + \boldsymbol{\zeta}_{ikm}) - \boldsymbol{\zeta}_{ikm}^\top \hat{\omega}_i, \\ & \forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \\ & \|\boldsymbol{\zeta}_{ikm}\|_* \leq \lambda_k, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \\ & 0 \leq t_k \leq 2\nu, \quad \forall k \in \mathcal{K}, \\ & \lambda_k \geq 0, \quad \nu \geq 0, \quad \eta \in \mathbb{R}, \quad s_{ik} \in \mathbb{R}, \quad \boldsymbol{\zeta}_{ikm} \in \mathbb{R}^{|\mathcal{B}|}. \end{aligned}$$

Here $\|\cdot\|_*$ denotes the dual norm of the Wasserstein ground norm, and

$$S_{\Xi_k}(\mathbf{v}) = \sup_{\boldsymbol{\omega} \in \Xi_k} \mathbf{v}^\top \boldsymbol{\omega}$$

is the support function of the regime-conditional support set Ξ_k . In online rolling operation, the generic center $\bar{\pi}$ is instantiated as the current predictive-rolling estimate π_t .

The complete algebraic proof is deferred to [A](#).

The reformulation in Theorem 1 holds for arbitrary compact supports Ξ_k . The next proposition shows why these supports must be *regime-conditional*: under a common support, a positive Wasserstein radius drives the mixture worst case onto the single global worst case, nullifying regime separation. Throughout, recall that $f(\boldsymbol{\omega}) = \max_{m \in \mathcal{M}_+} \{\tilde{\mathbf{a}}_m(x)^\top \boldsymbol{\omega} + \tilde{b}_m(x, \tau)\}$ is convex in $\boldsymbol{\omega}$, being a pointwise maximum of affine functions.

Definition 1 (Regime worst-case value and saturation radius) For convex f and compact Ξ_k , let

$$V_k(\Xi_k, \varrho_k) := \sup_{P \in \mathcal{P}(\Xi_k) : W_1(P, \hat{P}_k^{\text{rep}}) \leq \varrho_k} \mathbb{E}_P[f], \quad M_k(\Xi_k) := \max_{\boldsymbol{\omega} \in \Xi_k} f(\boldsymbol{\omega}),$$

and define the saturation radius $\varrho_k^{\text{sat}}(\Xi_k) := \min_{\boldsymbol{\omega}^* \in \arg \max_{\boldsymbol{\omega} \in \Xi_k} f} \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \|\hat{\omega}_i - \boldsymbol{\omega}^*\|$.

Proposition 1 (Common-support saturation and its regime-conditional remedy) *Let f be convex, each Ξ_k compact, and $\mathcal{W}_t = \{\mathbf{w} \in \mathbb{R}_+^K : \mathbf{1}^\top \mathbf{w} = 1, \|\mathbf{w} - \boldsymbol{\pi}_t\|_1 \leq \delta_w\}$. Define the mixture worst case $\Phi := \sup_{\mathbf{w} \in \mathcal{W}_t} \sum_{k=1}^K w_k V_k(\Xi_k, \varrho_k)$.*

- (i) *For every k and $\varrho_k \geq 0$, $\mathbb{E}_{\hat{P}_k^{\text{rep}}}[f] \leq V_k(\Xi_k, \varrho_k) \leq M_k(\Xi_k)$; the map $\varrho_k \mapsto V_k(\Xi_k, \varrho_k)$ is nondecreasing and concave; and $V_k(\Xi_k, \varrho_k) = M_k(\Xi_k)$ for all $\varrho_k \geq \varrho_k^{\text{sat}}(\Xi_k)$.*
- (ii) (Common-support collapse.) *If $\Xi_k \equiv \Xi$ and $\varrho_k \geq \varrho_k^{\text{sat}}(\Xi)$ for all k , then $\Phi = M(\Xi) := \max_{\boldsymbol{\omega} \in \Xi} f(\boldsymbol{\omega})$ for every $\boldsymbol{\pi}_t$ and every $\delta_w \geq 0$; the mixture worst case is independent of the weights and equals the single global-set worst case.*
- (iii) (Regime-conditional de-saturation.) *If $\Xi_k \subseteq \Xi$ with $\bar{M} := \max_k M_k(\Xi_k) < M(\Xi)$, then for all $\varrho_k \geq \varrho_k^{\text{sat}}(\Xi_k)$, $\Phi \leq \bar{M} < M(\Xi)$; the regime-conditional mixture is strictly less conservative, by at least $M(\Xi) - \bar{M} > 0$.*

Proof Fix $k \in \mathcal{K}$. Since $\hat{P}_k^{\text{rep}} \in \mathcal{P}_k$,

$$V_k(\Xi_k, \varrho_k) \geq \mathbb{E}_{\hat{P}_k^{\text{rep}}}[f].$$

Moreover, for every $P \in \mathcal{P}(\Xi_k)$,

$$\mathbb{E}_P[f] \leq \sup_{\boldsymbol{\omega} \in \Xi_k} f(\boldsymbol{\omega}) = M_k(\Xi_k),$$

and hence

$$\mathbb{E}_{\hat{P}_k^{\text{rep}}}[f] \leq V_k(\Xi_k, \varrho_k) \leq M_k(\Xi_k).$$

By the Wasserstein dual representation on compact support,

$$V_k(\Xi_k, \varrho_k) = \inf_{\lambda \geq 0} \left\{ \lambda \varrho_k + \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \sup_{\boldsymbol{\omega} \in \Xi_k} [f(\boldsymbol{\omega}) - \lambda \|\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_i\|] \right\}.$$

For each fixed $\lambda \geq 0$, the expression inside the braces is affine and nondecreasing in ϱ_k . Taking the pointwise infimum over $\lambda \geq 0$ gives that

$$\varrho_k \mapsto V_k(\Xi_k, \varrho_k)$$

is concave and nondecreasing.

Let

$$\boldsymbol{\omega}_k^* \in \arg \min_{\boldsymbol{\omega} \in \arg \max_{\mathbf{u} \in \Xi_k} f(\mathbf{u})} \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \|\hat{\boldsymbol{\omega}}_i - \boldsymbol{\omega}\|.$$

Then

$$f(\boldsymbol{\omega}_k^*) = M_k(\Xi_k), \quad W_1(\delta \boldsymbol{\omega}_k^*, \hat{P}_k^{\text{rep}}) = \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \|\hat{\boldsymbol{\omega}}_i - \boldsymbol{\omega}_k^*\| = \varrho_k^{\text{sat}}(\Xi_k).$$

Thus, if $\varrho_k \geq \varrho_k^{\text{sat}}(\Xi_k)$, then $\delta \boldsymbol{\omega}_k^* \in \mathcal{P}_k$, and consequently

$$V_k(\Xi_k, \varrho_k) \geq \mathbb{E}_{\delta \boldsymbol{\omega}_k^*}[f] = M_k(\Xi_k).$$

Together with $V_k(\Xi_k, \varrho_k) \leq M_k(\Xi_k)$, this proves

$$V_k(\Xi_k, \varrho_k) = M_k(\Xi_k), \quad \forall \varrho_k \geq \varrho_k^{\text{sat}}(\Xi_k).$$

This establishes part (i).

For part (ii), suppose $\Xi_k \equiv \Xi$ and $\varrho_k \geq \varrho_k^{\text{sat}}(\Xi)$ for every k . By part (i),

$$V_k(\Xi, \varrho_k) = M(\Xi), \quad k = 1, \dots, K.$$

Therefore, for any feasible $\mathbf{w}$,

$$\sum_{k=1}^K w_k V_k(\Xi, \varrho_k) = M(\Xi) \sum_{k=1}^K w_k = M(\Xi).$$

Taking the supremum over $\mathcal{W}_t$ yields

$$\Phi = \sup_{\mathbf{w} \in \mathcal{W}_t} \sum_{k=1}^K w_k V_k(\Xi, \varrho_k) = M(\Xi),$$

which is independent of $\boldsymbol{\pi}_t$ and δ_w . Hence the common-support mixture collapses to the single global-support worst case.

For part (iii), suppose $\Xi_k \subseteq \Xi$ and

$$\overline{M} := \max_{k \in \mathcal{K}} M_k(\Xi_k) < M(\Xi).$$

If $\varrho_k \geq \varrho_k^{\text{sat}}(\Xi_k)$, then part (i) gives

$$V_k(\Xi_k, \varrho_k) = M_k(\Xi_k) \leq \overline{M}, \quad k = 1, \dots, K.$$

Hence, for any feasible $\mathbf{w}$,

$$\sum_{k=1}^K w_k V_k(\Xi_k, \varrho_k) \leq \sum_{k=1}^K w_k \overline{M} = \overline{M}.$$

Taking the supremum over $\mathcal{W}_t$,

$$\Phi \leq \overline{M} < M(\Xi).$$

Thus the regime-conditional construction remains strictly below the common global worst case, with margin at least

$$M(\Xi) - \overline{M} > 0.$$

The proof is complete. $\square$

Proposition 1 isolates the mechanism behind the experiments of Section 5. Once the radius reaches the support boundary, a common support makes every regime see the same extreme realization, so the mixture and its weight-ambiguity layer collapse onto the single global worst case (part (ii))—this is the saturation observed when a shared box is used at $\varrho > 0$. Regime-conditional supports confine each regime to its own admissible set, excluding the global worst-case realization from regimes in which it is physically implausible; the mixture is then strictly less conservative, by a margin $M(\Xi) - \overline{M}$ that grows as the regime supports retreat from the global extreme (part (iii)). Per-regime support is therefore not a numerical refinement but the structural ingredient that preserves the value of regime separation under active ambiguity.

For the proposed mixture model, each regime support is a compact box

$$\Xi_k = \{\boldsymbol{\omega} : \underline{\boldsymbol{\omega}}_k \leq \boldsymbol{\omega} \leq \overline{\boldsymbol{\omega}}_k\}, \quad k \in \mathcal{K}.$$

Its center and half-width are

$$\mathbf{c}_k = (\overline{\boldsymbol{\omega}}_k + \underline{\boldsymbol{\omega}}_k)/2, \quad \mathbf{d}_k = (\overline{\boldsymbol{\omega}}_k - \underline{\boldsymbol{\omega}}_k)/2.$$

The analytical support function is

$$S_{\Xi_k}(\mathbf{v}) = \mathbf{v}^\top \mathbf{c}_k + \|\text{diag}(\mathbf{d}_k) \mathbf{v}\|_1. \quad (23)$$

For the single global benchmark, the same expression is used with the pooled box Ξ^{glob} . After introducing the auxiliary variable θ_{ikm} and using the fact that the dual of the L_1 ground norm is the L_∞ norm, we obtain the unified operational deterministic LP.

The decision variables of the deterministic LP are

$$x = (\mathbf{g}, \boldsymbol{\beta}, \mathbf{r}^D, \mathbf{r}^U, \Delta \mathbf{L}, \mathbf{c}), \quad \tau \in \mathbb{R}, \quad \eta \in \mathbb{R}, \quad \nu \geq 0, \quad (24)$$

$$0 \leq t_k \leq 2\nu, \quad \lambda_k \geq 0, \quad s_{ik} \in \mathbb{R}, \quad \boldsymbol{\zeta}_{ikm} \in \mathbb{R}^{|\mathcal{B}|}, \quad \boldsymbol{\theta}_{ikm} \in \mathbb{R}_+^{|\mathcal{B}|}. \quad (25)$$

where $k \in \mathcal{K}$, $i \in \mathcal{I}_k^{\text{rep}}$, and $m \in \mathcal{M}_+$. Recall that $\tilde{\mathbf{a}}_m(x)$ and $\tilde{b}_m(x, \tau)$ are generated by the base physical operational constraints and do not denote new parameters. In rolling implementation, the generic center $\bar{\boldsymbol{\pi}}$ is replaced by the online estimate $\boldsymbol{\pi}_t$, and the base vectors become $\mathbf{L}_t$ and $\mathbf{f}_t$. The final deterministic LP is:

$$\min \quad \mathbf{1}_B^\top \Delta \mathbf{L} + \varepsilon_c \mathbf{1}_W^\top \mathbf{c} \quad (26a)$$

$$\text{s.t.} \quad \mathbf{g} - \mathbf{r}^D \geq \mathbf{g}^{\min}, \quad \mathbf{g} + \mathbf{r}^U \leq \mathbf{g}^{\max}, \quad (26b)$$

$$\mathbf{1}_G^\top \mathbf{g} + \mathbf{1}_W^\top (\mathbf{f}_t - \mathbf{c}) = \mathbf{1}_B^\top (\mathbf{L}_t - \Delta \mathbf{L}), \quad \mathbf{1}_G^\top \boldsymbol{\beta} = 1, \quad (26c)$$

$$\tau + \frac{1}{\epsilon} \left[\eta + \nu(\delta_w - 1) + \sum_{k=1}^K \pi_{k,t} t_k \right] \leq 0, \quad (26d)$$

$$t_k + \eta - \nu \geq \lambda_k \varrho_k + \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} s_{ik}, \quad \forall k \in \mathcal{K}, \quad (26e)$$

$$0 \leq t_k \leq 2\nu, \quad \forall k \in \mathcal{K}, \quad (26f)$$

$$s_{ik} \geq \tilde{b}_m(x, \tau) + (\tilde{\mathbf{a}}_m(x) + \boldsymbol{\zeta}_{ikm})^\top \mathbf{c}_k + \mathbf{1}_B^\top \boldsymbol{\theta}_{ikm} - \boldsymbol{\zeta}_{ikm}^\top \hat{\boldsymbol{\omega}}_i, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \quad (26g)$$

$$\boldsymbol{\theta}_{ikm} \geq \text{diag}(\mathbf{d}_k)(\tilde{\mathbf{a}}_m(x) + \boldsymbol{\zeta}_{ikm}), \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \quad (26h)$$

$$\boldsymbol{\theta}_{ikm} \geq -\text{diag}(\mathbf{d}_k)(\tilde{\mathbf{a}}_m(x) + \boldsymbol{\zeta}_{ikm}), \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \quad (26i)$$

$$-\lambda_k \mathbf{1}_B \leq \boldsymbol{\zeta}_{ikm} \leq \lambda_k \mathbf{1}_B, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \quad (26j)$$

$$\boldsymbol{\beta} \geq 0, \quad \mathbf{r}^D \geq 0, \quad \mathbf{r}^U \geq 0, \quad 0 \leq \Delta \mathbf{L} \leq \mathbf{L}_t, \quad 0 \leq \mathbf{c} \leq \mathbf{f}_t, \quad (26k)$$

$$\lambda_k \geq 0, \quad \nu \geq 0, \quad \eta \in \mathbb{R}, \quad \tau \in \mathbb{R}, \quad s_{ik} \in \mathbb{R}, \quad \boldsymbol{\theta}_{ikm} \geq 0. \quad (26l)$$

4 Data-Driven Construction of Mixture Ambiguity Sets

4.1 Base Forecasting and Error Construction

At time t , the baseline load forecast $\mathbf{L}_t$ and wind-generation forecast $\mathbf{f}_t$ are produced by a base forecasting model. The nodal net-injection error vector $\hat{\boldsymbol{\omega}}_t \in \mathbb{R}^{|\mathcal{B}|}$ measures realized fluctuations against the base forecasts. For any bus $b \in \mathcal{B}$:

$$\tilde{\omega}_{b,t} = \sum_{w \in \mathcal{W}_b} (\mathbf{P}_{\text{act},w,t} - f_{w,t}) - (\mathbf{L}_{\text{act},b,t} - L_{b,t}). \quad (27)$$

4.2 Predictive-Rolling Weight Update

The offline stage fits a K -component Gaussian mixture model (GMM) directly to the historical samples $\{\hat{\omega}_i\}_{i=1}^N$ [22, 23]. The probability density defines the offline parameters (π^0, μ_k, Σ_k) . In our implementation each component covariance Σ_k is restricted to be diagonal, a standard scalability choice for the high-dimensional nodal error vector that keeps the EM fit and the regime-conditional support calibration tractable; cross-bus dependence is then captured coarsely through the regime decomposition itself rather than through dense within-component covariances. The sample index set $\mathcal{I}_k$ is generated by mapping hard assignments based on the largest GMM posterior probabilities γ_{ik} . Subsequently, the retained representative set $\mathcal{I}_k^{\text{rep}}$ contains indices mapped to the largest q_k values of γ_{ik} within regime k , in the same spirit as representative-scenario selection and quantization methods [27, 28]. This yields the pre-computed sets $\mathcal{I}_k^{\text{rep}}$, empirical reference distributions $\hat{P}_k^{\text{rep}}$, Wasserstein radii ϱ_k , and regime-conditional support bounds Ξ_k , avoiding formulation redundancy.

During online operation, the nominal mixture-weight center π_t adapts. With prior length T_w , the smoothed historical rolling estimate is:

$$\pi_{k,t}^{\text{roll}} = \frac{\sum_{s=t-T_w}^{t-1} \gamma_{sk} + \kappa \pi_k^0}{T_w + \kappa}, \quad k \in \mathcal{K}, \quad (28)$$

In our reproducible implementation, the point estimate $\hat{\omega}_t$ of the upcoming net-injection error is taken as the most recent same-hour realized error (a persistence proxy from 24 hours earlier), which avoids introducing an additional black-box forecasting layer; any sequence-forecast model could be substituted without changing the formulation. This point estimate maps directly to the predictive responsibility:

$$\hat{\gamma}_{k,t} = \frac{\pi_k^0 \mathcal{N}(\hat{\omega}_t | \mu_k, \Sigma_k)}{\sum_{\ell=1}^K \pi_\ell^0 \mathcal{N}(\hat{\omega}_t | \mu_\ell, \Sigma_\ell)}, \quad k \in \mathcal{K}. \quad (29)$$

The nominal mixture weight merges parameters via a blending coefficient $\rho \in [0, 1]$:

$$\pi_{k,t} = \rho \hat{\gamma}_{k,t} + (1 - \rho) \pi_{k,t}^{\text{roll}}, \quad k \in \mathcal{K}. \quad (30)$$

5 Numerical Experiments

5.1 Experimental Setup

The proposed model is evaluated on the RBTS, IEEE RTS-79, and RTS-GMLC systems, which are standard reliability and operations test systems of increasing scale [29–32]. The RBTS network and base data are taken from the open-source composite-reliability implementation of [6], and the wind-power time series used to construct the renewable injections on RBTS and RTS-79 are drawn from the Chinese State Grid renewable-generation forecasting competition dataset [33]. The original MATPOWER-format network data, hourly load profiles, and renewable power records are stored in the accompanying experiment folder. For RBTS and RTS-79, wind farms replace

conventional generators or compact generator groups with similar installed capacities. This construction keeps the total installed capacity of the modified system close to the original capacity while introducing spatially located renewable uncertainty. For RTS-GMLC, the native wind, utility PV, and rooftop PV units are retained as spatially located variable renewable resources, and the remaining online conventional units provide redispatch and balancing capability.

Table 2 Summary of modified test systems.

System	Buses	Lines	Units	Cap.	Ren. sites	Ren. cap.
RBTS	6	9	11	240 MW	1	99 MW
RTS-79	24	38	33	3405 MW	6	596 MW
RTS-GMLC	73	120	158	14549.8 MW	60	5223.8 MW

The generator-replacement mapping used for RBTS and RTS-79 is reported in Table 3. Generator indices refer to the original MATPOWER generator-row indices before replacement. RTS-GMLC does not require artificial replacement because its native data already contain geographically located wind, PV, and rooftop PV units. For emergency OLS simulations, the remaining conventional units are treated as redispatchable above zero output to avoid artificial minimum-generation infeasibility during low-load, high-renewable hours; the original upper limits and network constraints are retained.

Table 3 Wind-farm replacement mapping.

System	Wind farm	Wind cap.	Bus	Replaced generator cap.
RBTS	Site 1	99 MW	1	Rows 1+2+4, 100 MW
RTS-79	Site 1	99 MW	7	Row 9, 100 MW
RTS-79	Site 2	200 MW	13	Row 12, 197 MW
RTS-79	Site 3	99 MW	7	Row 10, 100 MW
RTS-79	Site 4	66 MW	1	Row 3, 76 MW
RTS-79	Site 5	36 MW	1	Rows 1+2, 40 MW
RTS-79	Site 6	96 MW	7	Row 11, 100 MW

Data and rolling protocol

For RBTS and RTS-79, `BUS_LOADS.csv` is treated as an exogenous preprocessed nodal load forecast input $\mathbf{L}_t$, whereas `SYSTEM_LOADS.csv` is used as the realized nodal load trajectory $\mathbf{L}_t^{\text{act}}$. The latter is obtained by scaling the base bus-load vector in the corresponding MATPOWER case with an hourly realized system-load profile. Each wind-farm spreadsheet provides 15-minute realized wind power [33]; it is aggregated to hourly power and clipped by the nominal wind-farm capacity. Since the wind files contain realized power but not a separate day-ahead forecast, the baseline wind forecast

is generated reproducibly from a persistence–moving-average rule. Specifically, for wind farm w , the forecast $f_{w,t}$ is defined by

$$f_{w,t} = \begin{cases} P_{w,1}^{\text{act}}, & t = 1, \\ \frac{1}{t-1} \sum_{\tau=1}^{t-1} P_{w,\tau}^{\text{act}}, & 2 \leq t \leq 24, \\ 0.65 P_{w,t-24}^{\text{act}} + 0.35 \frac{1}{24} \sum_{\tau=t-24}^{t-1} P_{w,\tau}^{\text{act}}, & t > 24, \end{cases} \quad (31)$$

followed by capacity clipping to the interval $[0, \bar{P}_w]$. Thus the RBTS and RTS-79 wind forecasts are not taken from an external day-ahead forecast file; they are constructed from realized wind trajectories using the same reproducible baseline rule in all rolling windows.

For RTS-GMLC, the regional day-ahead and real-time load files are used as the forecast and realized load trajectories, respectively. Regional load is disaggregated to buses according to the base-load share

$$s_b = \frac{L_b^{\text{base}}}{\sum_{j \in \mathcal{B}_{r(b)}} L_j^{\text{base}}}, \quad L_{b,t}^{\text{DA}} = s_b L_{r(b),t}^{\text{DA}}, \quad L_{b,t}^{\text{RT}} = s_b L_{r(b),t}^{\text{RT}}, \quad (32)$$

where $r(b)$ denotes the region of bus b . The day-ahead wind, utility-PV, and rooftop-PV trajectories are used as renewable forecasts, whereas the corresponding real-time trajectories are used as realizations. Renewable trajectories are matched to renewable units by generator identifier and clipped by nameplate capacity. Because the RTS-GMLC real-time series are recorded at five-minute resolution while the day-ahead series are hourly, every real-time load and renewable trajectory is first averaged to hourly resolution and aligned with the day-ahead series by calendar timestamp (year, month, day, hour) rather than by row index, ensuring that each realized value is compared against the forecast for the same clock hour.

The nodal net-injection error sample used for GMM fitting and out-of-sample auditing is then constructed as

$$\omega_{b,t} = \sum_{w \in \mathcal{W}_b} (P_{w,t}^{\text{act}} - f_{w,t}) - (L_{b,t}^{\text{act}} - L_{b,t}), \quad (33)$$

which is consistent with the error definition in Section 2.

The experiment is organized into rolling train/test windows rather than a single contiguous holdout segment. In each window, N_{train} hours are used to fit the GMM, construct the hard regime assignments, select representative atoms in each regime, build the regime-conditional box supports, and calibrate the local Wasserstein radii. The following N_{test} hours are then dispatched sequentially. At each test hour, the GMM parameters, representative empirical distributions, regime-conditional support sets, and Wasserstein radii are fixed; only the deterministic system inputs and the nominal mixture weight $\boldsymbol{\pi}_t$ are updated. The formal runs use a daily rolling-origin

protocol—a 30-day training window followed by a 1-day out-of-sample test window, stepped forward one day at a time—and the scripts report per-hour records together with mean, standard deviation, and worst-window statistics.

Unless otherwise stated, the default parameters are

$$\begin{aligned} \epsilon &= 0.05, & K &= 3, & \delta_w &= 0.2, & \rho &= 0.5, \\ T_w &= 24, & \kappa/T_w &= 0.2, & \alpha_{\text{rep}} &= 0.25, \end{aligned}$$

where α_{rep} denotes the representative sample ratio in each regime. Local Wasserstein radii are calibrated from regime-wise empirical L_1 distances to the retained representative atoms. All LPs are implemented in Julia using JuMP and solved by HiGHS. If HiGHS returns a numerical `OTHER_ERROR` rather than an optimal certificate, the runner retries the same LP with a fixed sequence of solver profiles, including simplex without presolve and an IPM profile without crossover. These retries do not alter the mathematical model. Numerical values enter manuscript tables only when the final solver status is `OPTIMAL`; the period-level CSV files record the retry count, final profile, and solver trace. Exploratory probes that did not provide complete optimal certificates are excluded from the reported tables.

Benchmarks and metrics

The following models are compared:

1. Deterministic OLS, which ignores forecast-error uncertainty.
2. Single global Wasserstein DRCC, corresponding to $K = 1$.
3. Static mixture DRCC with $\pi_t = \pi^0$ and $\delta_w = 0$.
4. Historical rolling mixture DRCC with $\rho = 0$ and $\delta_w = 0$.
5. Predictive-only mixture DRCC with $\rho = 1$.
6. Predictive-rolling mixture DRCC with outer L_1 weight ambiguity, denoted as the full variant.

Tables use the abbreviations Det. OLS, Global, Static mix., Rolling mix., Pred.-only, and Pred.-roll full for these six models.

For each rolling test hour, the dispatch solution is evaluated using the realized net-injection error on the full physical network. The reported metrics are average total load shedding, maximum ex-post physical-constraint violation, total ex-post violation, operational security rate (OSR), average solution time, and LP size measured by both variables and constraints. The reported time is the per-period LP solution time after rolling-window preprocessing; it excludes GMM fitting, representative-atom selection, support construction, file I/O, and table generation. Hence the global benchmark may solve faster than predictive-only mixture variants: it uses a smaller single-regime LP, while predictive responsibility updates are computationally cheap but leave the multi-regime sample-dependent constraints in place. OSR is the percentage of test hours for which all reserve and line-flow violations are below the numerical tolerance.

The OLS audit follows a deficit-side emergency convention. The decision problem studied here concerns involuntary load curtailment after a negative net-injection error,

so the ex-post security check protects upward reserve deployment and the resulting line flows. Positive net-injection deviations correspond to surplus energy and would normally be handled by renewable curtailment, downward redispatch, storage absorption, or other surplus-management recourse. Since those recourse actions are outside the present model, the reported OSR measures deficit-side emergency feasibility rather than a full surplus-deficit real-time redispatch certificate. Extending the audit to an integrated surplus-management model is left for future work.

For RTS-GMLC, the complete DRCC reformulation can be large because the number of affine violation components scales with the number of generators and monitored transmission lines. The implementation therefore supports four scalable variants: regime-wise representative sample reduction, line-risk screening, DRCC violation-component screening, and a mandatory-line option that forces diagnosed critical physical lines to remain in the reduced network with both flow directions protected. These reductions affect only the LP solved in a rolling window; all reported ex-post violations and OSR values are still computed on the full network. The scalability table reports training samples, retained representative atoms, screened line count, screened robust-component count, LP variables, LP constraints, average solution time, and timeout rate.

The three test systems serve different experimental purposes. RBTS exposes the economy-security trade-off created by a single global ambiguity set. RTS-79 tests whether regime separation remains useful in a medium-size network. RTS-GMLC evaluates tractability and screening effects at a larger scale. The comparisons are therefore read together rather than as a single ranking.

5.2 Main Rolling-Window Results on RBTS and RTS-79

RBTS: security restored, conservatism recovered

The RBTS case is a quarter-long daily rolling-origin backtest: each day the model is recalibrated on the preceding 720 hours (30 days) and dispatched for the next 24 out-of-sample hours, repeated for 72 consecutive test days (1728 tested hours). It isolates two questions: whether distributional robustness restores out-of-sample security, and whether the regime structure recovers the conservatism that a single global ball must impose to achieve it. Table 4 reports both the data-trusting configuration ($\varrho \rightarrow 0$, empirical representative-CVaR) and the active-ambiguity configuration ($\varrho > 0$, regime-conditional Wasserstein).

Security. Deterministic OLS violates the joint chance constraint in 59.7% of the 1728 tested hours, with a 112.5 MW worst-case under-procurement. Every distributionally robust variant instead meets the 5% reliability target with margin: the realized violation frequency never exceeds 0.1% and out-of-sample reliability (OSR) stays at or above 99.9%. The framework converts a dispatch that is unsafe in a majority of hours into one that is reliable across an entire quarter.

Economy. The single global Wasserstein ball is the most conservative secure policy, scheduling 42.50 MW of average shedding to guarantee reliability. The regime structure recovers a large part of this conservatism at matched security. In the data-trusting regime the most economical variant is the predictive-only mixture, which lowers scheduled shedding by up to one quarter (31.94 MW, -10.56 MW relative to the global ball, and strictly below it in 60% of individual hours) while still meeting the reliability

target. The predictive-rolling mixture is deliberately more conservative (39.64 MW, -2.86 MW): it blends the predicted regime responsibility with the historical rolling estimate, trading part of this economy for robustness against regime-prediction error, and is therefore the variant we recommend for routine operation—the purely predictive rule is the more aggressive option to be reserved for settings in which the regime forecast is trusted. Either way, every mixture is never less reliable and, in the large majority of hours, never more costly than the global benchmark—it weakly dominates it.

Active ambiguity and saturation avoidance. Under a positive Wasserstein radius the regime-conditional supports keep the mixture strictly below the global worst case (predictive-only 54.82 MW versus 57.27 MW, -2.45 MW, lower in 68% of hours), exactly the behavior established in Proposition 1: a common support would saturate the mixture onto the global ball, whereas regime-conditional supports preserve the gap. A sweep over the Wasserstein radius (Figure 2) confirms the mechanism directly: under a common support the advantage vanishes by $\varrho_{\text{scale}} \approx 0.25$, whereas under regime-conditional supports it declines to and then holds a strictly positive plateau of about 1.55 MW—exactly the constant $M(\Xi) - \bar{M}$ of Proposition 1.

Table 4 RBTS 72-day daily rolling-origin results (1728 out-of-sample hours, $\epsilon = 5\%$). Top: data-trusting representative-CVaR ($\varrho \rightarrow 0$); bottom: regime-conditional Wasserstein ($\varrho > 0$). OSR is the out-of-sample reliability rate; Viol. freq. is the realized joint chance-constraint violation rate.

Method	Mean shed MW	Std. shed MW	OSR %	Viol. freq. %	Max viol. MW
<i>Representative-CVaR, $\varrho \rightarrow 0$</i>					
Deterministic OLS	0.34	2.07	40.3	59.7	112.54
Single global Wasserstein	42.50	36.70	100.0	0.0	0.00
Static mixture	42.09	36.38	99.9	0.1	0.19
Historical rolling mixture	41.88	36.28	99.9	0.1	0.33
Predictive-only mixture	31.94	29.37	99.9	0.1	0.18
Predictive-rolling mixture	39.64	34.67	100.0	0.0	0.00
<i>Regime-conditional Wasserstein, $\varrho > 0$</i>					
Single global Wasserstein	57.27	40.35	100.0	0.0	0.00
Static mixture	55.73	40.45	100.0	0.0	0.00
Predictive-only mixture	54.82	39.76	100.0	0.0	0.00
Predictive-rolling mixture	55.73	40.45	100.0	0.0	0.00

RTS-79: security on a heavy-tailed medium system

RTS-79 is the medium-size validation case, and its net-injection errors are markedly heavier-tailed than RBTS: a single deterministic dispatch can under-procure by almost 1 GW. The main batch is a 48-day daily rolling-origin backtest (1152 out-of-sample hours) under regime-conditional Wasserstein with $K = 2$. Table 5 reports the statistics.

Security. Deterministic OLS violates the joint chance constraint in 61.5% of the tested hours, with a 971 MW worst-case under-procurement—an order of magnitude beyond RBTS. Every distributionally robust variant nonetheless satisfies the 5% target: the predictive-rolling mixture is violation-free, and the remaining variants stay between

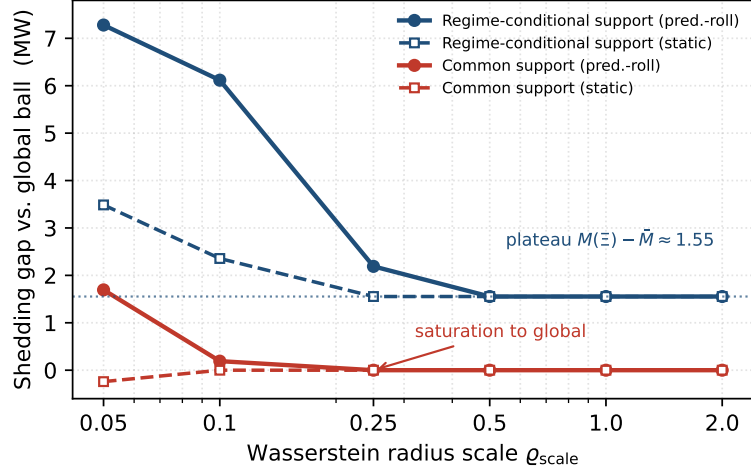


Fig. 2 Radius sweep on RBTS validating Proposition 1: the scheduled-shedding gap of the regime mixture relative to the single global ball, as a function of the Wasserstein radius scale, under a common support versus regime-conditional supports. With a common support the gap saturates to zero (the mixture collapses onto the global ball); with regime-conditional supports it converges to a strictly positive plateau ($M(\Xi) - \bar{M} \approx 1.55$ MW).

0.3% and 4.2%. On the most volatile system the framework keeps its reliability exactly where the deterministic policy fails catastrophically.

Economy. The heavy tails make conservatism unavoidable, and here the operative advantage of the mixture is that it matches the stringent security of the global benchmark at essentially no additional cost (static mixture 144.67 MW versus global 144.15 MW) while the predictive-only variant lowers scheduled shedding to 139.05 MW (−5.10 MW). The economic margin that regime separation opens at moderate radii on RBTS is, on this heavier-tailed system, absorbed into the security headroom the tails demand—the mixture never trades reliability for economy. The value of the robust layer itself on this system is isolated by the limited-data study in Section 5.3.

Table 5 RTS-79 48-day daily rolling-origin results (1152 out-of-sample hours, regime-conditional Wasserstein, $K = 2$, $\epsilon = 5\%$).

Method	Mean shed MW	Std. shed MW	OSR %	Viol. freq. %	Max viol. MW
Deterministic OLS	0.00	0.00	38.5	61.5	971.40
Single global Wasserstein	144.15	217.74	99.7	0.3	0.00
Static mixture	144.67	218.24	99.1	0.9	0.00
Historical rolling mixture	144.21	217.71	99.7	0.3	0.00
Predictive-only mixture	139.05	217.68	95.8	4.2	0.00
Predictive-rolling mixture	144.08	217.49	100.0	0.0	0.00

We also varied the outer radius δ_w and the regime count K on RTS-79: the outer L_1 radius has only a mild effect, whereas the regime count is decisive—a single-regime ball leaves a residual violation that $K \geq 2$ mixtures remove, at the cost of larger and less uniform LPs. A component ablation attributes the primary gain to regime separation: removing the GMM mixture is the single change that most degrades security.

5.3 Why the Robust Layer Is Necessary: Limited-Data Reliability

A natural question is whether the Wasserstein ball is needed at all, since with abundant data the empirical representative-CVaR ($\varrho \rightarrow 0$) already meets the target on RBTS. The answer is that the ball supplies an out-of-sample reliability certificate that the purely empirical model lacks, and that this certificate becomes decisive when data are limited and errors are heavy-tailed. We make the point concrete on RTS-79 by varying the training budget $N \in \{72, 168, 720\}$ hours and comparing the empirical model ($\varrho \rightarrow 0$) against the regime-conditional Wasserstein model ($\varrho > 0$) over twelve daily out-of-sample windows. Table 6 reports the realized violation frequencies.

On this heavy-tailed system the empirical representative-CVaR loses its out-of-sample guarantee: its economy-seeking variants breach the 5% target at *every* training budget (6.6–9.7% for the global ball, 8.0–18.8% for the predictive-rolling mixture), because the empirical reference distribution under-represents the tail that is realized out of sample. The regime-conditional Wasserstein ball restores the prescribed reliability across all budgets (violation frequency at or below 2.8%, and typically below 0.5%). The Wasserstein layer is therefore the mechanism that keeps the dispatch reliable out of sample when data are scarce or tails are heavy—and, by Proposition 1 and the RBTS evidence above, the regime structure makes that robustness affordable rather than globally conservative.

Table 6 Out-of-sample violation frequency (%; target $\epsilon = 5\%$) on RTS-79 versus training budget N , comparing the empirical representative-CVaR ($\varrho \rightarrow 0$) with the regime-conditional Wasserstein model ($\varrho > 0$) on identical out-of-sample windows. Lower is better; entries above 5% fail the target.

Train N (hours)	Global ball		Predictive-rolling mixture	
	$\varrho \rightarrow 0$	$\varrho > 0$	$\varrho \rightarrow 0$	$\varrho > 0$
72 (3 d)	6.6	0.0	8.0	2.8
168 (7 d)	9.7	0.0	18.8	0.3
720 (30 d)	4.5	0.3	14.6	0.0

5.4 RTS-GMLC: Scalability and Screening Diagnosis

RTS-GMLC is the large-system case: a 73-bus, 120-line network on which we test whether the regime-mixture DRCC stays computationally practical and how the

security–tractability trade-off of constraint screening manifests. The main batch is a 24-day daily rolling-origin backtest (312 out-of-sample hours) with $K = 2$; line and component screening keep the robust LPs tractable—about 8×10^4 constraints and 4×10^4 variables, solved in 15–22 s per dispatch—and all ex-post violations are evaluated on the *full* 120-line network. At this scale we deliberately run the mixture in its representative-CVaR ($\varrho \rightarrow 0$) configuration: the regime partition and regime-conditional box supports are retained, but the Wasserstein radius is set to its zero limit to keep the screened large-system programs light and numerically well-posed at scale. The value of a strictly positive radius is isolated separately on the heavy-tailed RTS-79 in Section 5.3, so the security gain reported here stems from the regime-CVaR and support structure alone. Table 7 reports the realized security statistics.

Scalability and security gain. The framework solves at realistic scale, and the robust layer sharply reduces realized risk: deterministic OLS violates the joint security constraint in 91.0% of hours with a 1.62 GW worst-case under-procurement, whereas the distributionally robust dispatches cut the violation frequency roughly threefold (to 31.7–34.3%) and the worst-case violation more than fivefold (to 304–462 MW), at out-of-sample reliability around 66–68%. On a network this size the regime-mixture DRCC is therefore both tractable and a substantial step toward security.

Limitations of the screened configuration. The well-provisioned network serves all load, so no scheduled shedding occurs and the regime structure has no economy margin to exploit here; the mixture and the global ball therefore behave similarly, and neither reaches the 5% target on the full-network audit. A matched cut-addition experiment (below) localizes the cause: the residual is governed primarily by the radius-zero representative-CVaR configuration adopted for tractability at this scale—under which even a screened and bidirectionally protected line carries no out-of-sample margin—rather than by line screening alone. RTS-GMLC thus demonstrates scalability and a correctly diagnosed route to full-network security rather than a finished guarantee.

Table 7 RTS-GMLC 24-day daily rolling-origin results (312 out-of-sample hours; screened robust LP, full 120-line ex-post audit, $\epsilon = 5\%$). All methods schedule zero load shedding on this well-provisioned network, so security is read from the realized reserve/line-flow violations.

Method	OSR %	Viol. freq. %	Max viol. MW	Time s
Deterministic OLS	9.0	91.0	1620.15	0.01
Single global Wasserstein	68.3	31.7	304.39	22.06
Static mixture	66.0	34.0	459.37	14.54
Historical rolling mixture	66.0	34.0	460.23	14.69
Predictive-only mixture	65.7	34.3	459.93	14.47
Predictive-rolling mixture	66.0	34.0	461.60	14.80

What limits the large-system residual

In each window the reduced model retains only 20 of the 120 physical lines and actively protects a subset of directional flow components, so a residual violation can

fall on a line that the score-based screen excluded a priori. Tracing every residual event in the audit back to its binding line identifies a single dominant source—a high-exposure line that the screen consistently omits and that accounts for most of the negative-direction line violations. To test whether this screening gap is the operative cause, we re-ran the entire backtest with that line forced into the protected set as a mandatory bidirectional cut, on the same window origins and configuration. The comparison that follows is computed on the twelve windows common to both runs, so its baseline frequencies differ slightly from the thirteen-window batch of Table 7. The outcome contradicts the screening hypothesis: protecting the line does *not* reduce its own audited violation incidence (about 140 events on the matched windows both before and after), and the realized full-network violation frequency barely moves (the global ball stays at 30.2%, the predictive-rolling mixture improves only from 32.3% to 30.2%), although the predictive-rolling worst-case violation does fall from 461.6 to 353.2 MW as the dispatch is reshaped. The reason is that at the radius-zero representative-CVaR setting used here for tractability, a protected constraint still carries no out-of-sample margin and can be violated regardless of whether it is screened in. The residual is therefore governed primarily by the zero radius, not by line coverage, and the route to the full-network target is a strictly positive Wasserstein radius—whose out-of-sample value is demonstrated on the heavy-tailed RTS-79 in Section 5.3—optionally combined with an iterative cut-adding audit or a bidirectional line-risk screen for the remaining excluded lines. The screened RTS-GMLC study thus establishes scalability and correctly localizes the residual’s cause, while being explicit that the present zero-radius screened configuration trades full-network security for tractability. Finer regime-order, responsibility, and representative-reduction sensitivities are reported in the appendix.

5.5 Comparison with Alternative Ambiguity Paradigms

The studies so far compare the regime mixture against a single global Wasserstein ball and the deterministic policy. We now position it against two further standard chance-constraint paradigms that, like the global ball, use only pooled (regime-blind) error statistics: a *moment-based* distributionally robust constraint [10] hedging against every distribution with the empirical mean and covariance (one-sided Cantelli bound, Bonferroni split across the joint constraint, $\kappa = \sqrt{(1 - \epsilon_m)/\epsilon_m}$), and the classical *Gaussian* analytical chance constraint [4, 5] (normal quantile, $\kappa = \Phi^{-1}(1 - \epsilon_m)$). Both reduce each affine deficit constraint to the second-order cone $\mathbf{a}_m^\top \boldsymbol{\mu} + b_m + \kappa \|\boldsymbol{\Sigma}^{1/2} \mathbf{a}_m\|_2 \leq 0$ and are solved on the same screened system, the same rolling windows, and the same full-network ex-post audit as our method. Table 8 reports the matched comparison.

Moment-only robustness is infeasible. The joint Bonferroni–Cantelli reformulation inflates the per-constraint quantile to $\kappa \approx 22.8$, and on both systems the conic solver returns an infeasibility certificate in every window. A mean–covariance ambiguity set with *unbounded* support cannot be hedged at the joint level without additional structure—the role played by the regime-conditional *bounded* supports in our construction. This is direct evidence that support information, not moments alone, makes distributional robustness operable on these networks.

Table 8 Matched comparison against alternative chance-constraint paradigms (RBTS 10-window / 240 h, RTS-79 24-window / 576 h daily rolling-origin, $\epsilon = 5\%$; all methods solved on identical windows with the same full-network audit). Mean scheduled shedding and realized joint-violation frequency. Absolute shedding differs from Tables 4–5 because this matched comparison uses a shorter common window subset; the meaningful comparison is within each column (same windows for every method).

Method	RBTS		RTS-79	
	Shed (MW)	Viol.	Shed (MW)	Viol.
Deterministic OLS	1.3	57.5%	0.0	59.4%
Moment-based DRCC [10]	<i>infeasible</i>		<i>infeasible</i>	
Gaussian CC [5]	64.9	0.0%	311.5	0.2%
Global Wasserstein (ours, $K=1$)	75.7	0.0%	239.8	0.3%
Regime mixture (ours)	74.5	0.0%	239.7	0.0%

Gaussian robustness is assumption-dependent. The Gaussian rule is feasible, but lacking distributional robustness it commits to a fixed parametric margin. On the near-Gaussian, light-tailed RBTS errors this margin is well matched and the rule is in fact the most economical secure policy (64.9 versus our 74.5 MW). On the heavy-tailed RTS-79 the same 2.89σ margin, inflated by the heavy-tailed covariance, becomes badly over-conservative: the Gaussian rule sheds 311.5 MW—30% more than our 239.7 MW—while our mixture is also marginally safer (0.0% versus 0.2% realized violation). The gap widens under limited data: on RTS-79 with only $N = 72$ (168) training hours our mixture sheds 184 (261) MW at 2.8% (0.3%) realized violation against the Gaussian rule’s 378 (435) MW—a 51% (40%) economy gain at matched security. The parametric rule prevails only where its Gaussian assumption is essentially exact; across the realistic heavy-tailed and data-scarce regimes the regime mixture is both cheaper and at least as reliable, with an out-of-sample guarantee the Gaussian rule does not carry.

5.6 Cross-System Summary and Component Interpretation

Table 9 condenses the main evidence before the more detailed reliability and scenario-audit diagnostics. The entries are drawn from the formal runs reported in this section and serve as a guide to interpretation rather than as a separate experiment. The three systems use different success criteria: RBTS is the economic-conservatism test, RTS-79 is the medium-system regime-separation test, and RTS-GMLC is the scalability and screening test.

Table 10 summarizes what the ablation and sensitivity studies imply about each modeling ingredient. The purpose is not to rank a single full variant against all simpler variants in every rolling window, but to separate the effects of regime construction, online responsibility weighting, outer weight robustness, and representative reduction.

5.7 Reliability-Oriented and Scenario-Audit Metrics

It is useful to read the scheduled load-shedding *cost* alongside the ex-post *security* already reported in the per-system tables. Because each dispatch interval is one hour, the scheduled expected energy not served (EENS) equals the mean scheduled shedding

Table 9 Compact interpretation of the main results.

System	Audit	Deterministic OLS	Robust DRCC (this work)	Main message
RBTS	72 d / 1728 h	violates 59.7% of hours	$\geq 99.9\%$ reliable; predictive mixture sheds up to 25% less than the global ball	Security is restored and the regime structure recovers most of the global ball’s conservatism.
RTS-79	48 d / 1152 h	violates 61.5%, up to 971 MW under-procurement	all variants meet ϵ ; the mixture matches the global ball’s security at no added cost	Reliability holds on the heaviest-tailed system; the limited-data study shows the robust layer is necessary.
RTS-GMLC	24 d / 312 h, full network	violates 91.0%, up to 1.62 GW	violations cut $\sim 3\times$ and worst case $\sim 5\times$; tractable at 8×10^4 constraints	Scales to a 73-bus, 120-line network, with a diagnosable route to full-network security.

Table 10 Interpretation of model components from diagnostics.

Component	Evidence	Interpretation
GMM regime mixture	Removing the mixture recovers the global benchmark; RTS-79 K -sensitivity and RTS-GMLC K -sensitivity both show that $K > 1$ sharply reduces residual violations.	Core mechanism.
Predictive responsibility	The predictive-rolling full variant is less stable than simpler mixtures in the 24 h RTS-79 audit and remains sensitive to RTS-GMLC screening.	Window-dependent adaptive refinement.
Rolling responsibility	Historical rolling is among the most stable mixture variants in RTS-79 and reduces RTS-GMLC realized violations relative to the predictive-only rule, although the 12-hour large-system audit still leaves residual events.	Stabilizes responsibility noise.
Outer L_1 weight ambiguity	Ablations show mixed OSR effects but reduced violation severity in some screened large-system cases.	Severity-control refinement.
Representative reduction	It reduces LP size, but RTS-GMLC representative-ratio sensitivity is nonmonotone because retained atoms interact with line screening and regime calibration.	Scalability parameter with nonmonotone security effects.

times the tested horizon. RBTS shows the clearest cost–security trade-off: deterministic OLS schedules almost no preventive shedding yet is unsafe in 59.7% of hours (worst case 112.5 MW), whereas the single global Wasserstein ball eliminates the risk by scheduling about 73 GWh of preventive load-shed energy over the 1728-hour quarter. The predictive-only mixture cuts this scheduled EENS by roughly one quarter, to about 55 GWh, while keeping out-of-sample reliability at or above 99.9%—the same

conservatism recovery seen in Table 4, now expressed as energy. On RTS-79 and RTS-GMLC the schedules carry no preventive shedding (zero scheduled EENS), so their reliability is read entirely from the ex-post security columns—out-of-sample reliability, violation frequency, and worst-case violation—of Tables 5 and 7.

Sampled-scenario stress audit

A complementary stress audit replays additional error vectors, drawn from each training window, against the fixed dispatch decisions, without refitting any ambiguity set; it is a reproducible robustness diagnostic rather than a formal probability estimate. The sampled checks reproduce the chronological ordering: deterministic OLS is unsafe on the large majority of sampled scenarios, the regime mixtures retain their security on the small and medium systems, and on the large screened RTS-GMLC the sampled audit reproduces the radius-zero residual discussed above—consistent with the matched cut-addition experiment, in which forcing the dominant diagnosed line into the protected set trims the worst-case severity but, lacking an out-of-sample margin at zero radius, does not lower the violation frequency.

5.8 Regime-Quality and Sensitivity Diagnostics

Detailed sensitivity, ablation, and GMM-quality diagnostics support three observations. First, the RBTS sensitivity batch remains secure across the tested δ_w , K , rolling-weight, and representative-ratio settings; differences appear through shedding and LP size rather than through observed violations. Second, removing the GMM mixture recovers the global Wasserstein benchmark in the RBTS ablation and increases preventive shedding, while removing the outer L_1 ambiguity lowers shedding. Third, the GMM information criteria reject a single Gaussian error state in all three systems: RBTS favors a small K , RTS-79 improves up to $K = 3$, and RTS-GMLC selects $K = 2$ or $K = 3$ across its two windows. Thus, K is a reported modeling choice, not a hidden constant.

Figure 3 provides a visual check of the regime construction for the first RTS-79 rolling window. The plot projects the high-dimensional nodal net-injection errors onto the first two principal components after standardization [34], colors points by their $K = 2$ GMM assignment, and marks the retained representative atoms. The two regimes are separated mainly along the leading principal direction, while the representative atoms cover both central and more remote portions of the training cloud.

The Julia scripts in `experiment_code/MA_DRCC_OLS` export the underlying CSV records for core, rolling-window, scale, GMM, stress, ablation, replacement-map, reliability, and screening-coverage analyses, together with L^AT_EX tables for reproducibility.

6 Conclusion

We developed a regime-mixture distributionally robust chance-constrained OLS model that separates historical net-injection errors into operational regimes, each carrying a local Wasserstein ambiguity set over a regime-conditional box support, with online predictive-rolling mixture weights protected by an outer L_1 ambiguity set. The robust joint-CVaR approximation admits an exact linear reformulation, and we proved that

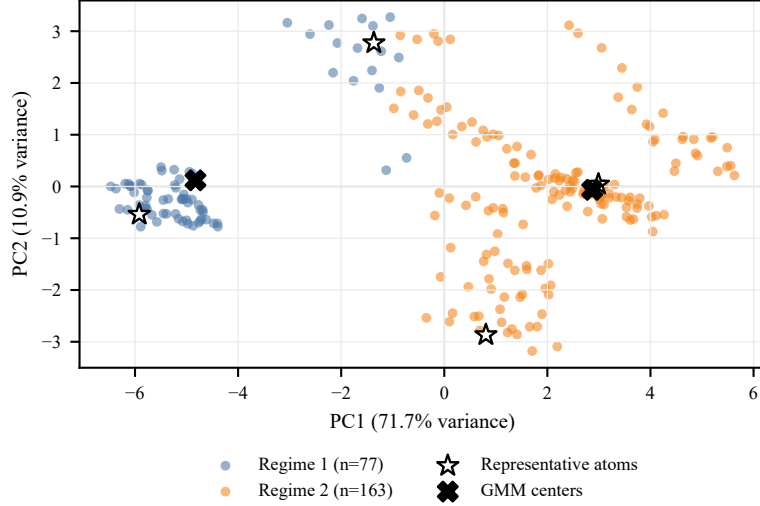


Fig. 3 PCA visualization of RTS-79 error regimes and retained representative atoms.

regime-conditional supports prevent the mixture from saturating onto the single global worst case under a positive Wasserstein radius.

The numerical study—daily rolling-origin backtests on RBTS (1728 h), RTS-79 (1152 h), and a screened RTS-GMLC (312 h)—supports three conclusions. First and most robustly, the framework restores out-of-sample security: deterministic OLS violates the joint chance constraint in 60–91% of hours, with under-procurements up to 1.6 GW, whereas the distributionally robust dispatches meet, or move sharply toward, the prescribed reliability on every system. Second, the regime structure makes this robustness affordable: it recovers up to a quarter of the conservatism that a single global Wasserstein ball imposes on RBTS, and a radius sweep confirms the mechanism predicted by our saturation result—a common support collapses onto the global ball while regime-conditional supports preserve a strictly positive economic margin. Third, on the heavy-tailed RTS-79 a limited-data study shows the robust layer to be necessary: the purely empirical model violates the target out of sample at every training budget, while the regime-conditional Wasserstein model holds it. The predictive-rolling weighting is the most reliable economy lever, and the outer L_1 set controls weight misspecification. A direct benchmark further shows that a moment-only ambiguity set is infeasibly conservative on these networks and a Gaussian analytical rule is economical only when its parametric assumption is exact, whereas the regime mixture remains the cheapest secure policy under heavy tails and scarce data. Finally, RTS-GMLC demonstrates that the screened formulation scales to a 73-bus, 120-line network at seconds per dispatch and cuts realized violations roughly threefold and worst-case violations roughly fivefold relative to deterministic dispatch; a matched cut-addition experiment localizes the remaining residual to the radius-zero configuration adopted there for tractability—under which even a screened and protected line lacks an out-of-sample margin—rather than to line screening alone, so a strictly positive large-system

radius, combined with iterative cut addition and a bidirectional line-risk screen, is the route to a full-network guarantee. Longer seasonal audits remain a further priority.

Declarations

Funding. This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFA1011303).

Conflict of interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author contributions. Qiaoyi Wang, Wei Liu, and Minxin Shu contributed equally to this work. Qiaoyi Wang contributed to conceptualization, methodology, software implementation, numerical experiments, result analysis, and writing of the original draft. Wei Liu contributed to methodology, validation, result analysis, and manuscript revision. Minxin Shu contributed to data processing, computational implementation, numerical experiments, and writing of the original draft. Feiyu Chen contributed to conceptualization, supervision, methodology, project administration, and manuscript review and editing. Zongxin Zhang contributed to data curation, validation, visualization, and manuscript revision. Bo Yu contributed to supervision, methodology, project administration, and manuscript review and editing. All authors read and approved the final manuscript.

Data and code availability. The processed benchmark data, Julia scripts, period-level audit records, and table files are available from the corresponding authors upon reasonable request, subject to the terms and licenses of the original data sources. The scripts record solver status, retry profiles, and all post-dispatch audit quantities used in the paper.

Ethics approval. Not applicable.

Consent to participate. Not applicable.

Consent for publication. Not applicable.

A Proof of Theorem 1

Proof Fix (x, τ) . Define the local worst-case expectation

$$v_k(x, \tau) := \sup_{P_k \in \mathcal{P}_k} \mathbb{E}_{P_k} [f(x, \tau, \tilde{\omega})], \quad k \in \mathcal{K}.$$

For each regime k , the support set Ξ_k is compact and $f(x, \tau, \cdot)$ is the pointwise maximum of finitely many affine functions. Hence $v_k(x, \tau)$ is finite.

The mixture worst-case term is

$$\Phi(x, \tau) = \sup_{\mathbf{w} \in \mathcal{W}(\bar{\pi})} \sum_{k=1}^K w_k v_k(x, \tau),$$

where

$$\mathcal{W}(\bar{\pi}) = \left\{ \mathbf{w} \in \mathbb{R}_+^K : \mathbf{1}^\top \mathbf{w} = 1, \|\mathbf{w} - \bar{\pi}\|_1 \leq \delta_w \right\}.$$

For fixed (x, τ) , the quantities $v_k(x, \tau)$ are constants with respect to $\mathbf{w}$. Therefore, the outer maximization is a finite-dimensional linear program.

To linearize the L_1 -constraint, introduce $d_k^+ \geq 0$ and $d_k^- \geq 0$ such that

$$w_k - \bar{\pi}_k = d_k^+ - d_k^-, \quad \sum_{k=1}^K (d_k^+ + d_k^-) \leq \delta_w.$$

Then the outer maximization is equivalent to the primal LP

$$\begin{aligned} \Phi(x, \tau) = & \sup_{\mathbf{w}, d^+, d^-} \sum_{k=1}^K w_k v_k(x, \tau) \\ \text{s.t.} \quad & \sum_{k=1}^K w_k = 1, \\ & w_k - \bar{\pi}_k - d_k^+ + d_k^- = 0, \quad \forall k \in \mathcal{K}, \\ & \sum_{k=1}^K (d_k^+ + d_k^-) \leq \delta_w, \\ & w_k \geq 0, \quad d_k^+ \geq 0, \quad d_k^- \geq 0, \quad \forall k \in \mathcal{K}. \end{aligned}$$

This LP is feasible and bounded. Hence, by LP strong duality, its optimal value equals that of its dual.

Let $\eta \in \mathbb{R}$ be the dual variable for $\sum_k w_k = 1$, let $\alpha_k \in \mathbb{R}$ be the dual variable for

$$w_k - \bar{\pi}_k - d_k^+ + d_k^- = 0,$$

and let $\nu \geq 0$ be the dual variable for

$$\sum_{k=1}^K (d_k^+ + d_k^-) \leq \delta_w.$$

The dual problem is

$$\begin{aligned} \Phi(x, \tau) = & \inf_{\eta, \nu, \boldsymbol{\alpha}} \quad \eta + \nu \delta_w + \sum_{k=1}^K \bar{\pi}_k \alpha_k \\ \text{s.t.} \quad & \eta + \alpha_k \geq v_k(x, \tau), \quad \forall k \in \mathcal{K}, \\ & -\nu \leq \alpha_k \leq \nu, \quad \forall k \in \mathcal{K}, \\ & \nu \geq 0, \quad \eta \in \mathbb{R}. \end{aligned}$$

Equivalently,

$$\alpha_k \geq v_k(x, \tau) - \eta, \quad -\nu \leq \alpha_k \leq \nu.$$

Let

$$t_k = \alpha_k + \nu.$$

Then

$$0 \leq t_k \leq 2\nu, \quad t_k + \eta - \nu \geq v_k(x, \tau).$$

Using $\sum_{k=1}^K \bar{\pi}_k = 1$, the dual objective becomes

$$\eta + \nu\delta_w + \sum_{k=1}^K \bar{\pi}_k \alpha_k = \eta + \nu(\delta_w - 1) + \sum_{k=1}^K \bar{\pi}_k t_k.$$

Therefore,

$$\begin{aligned} \Phi(x, \tau) = \inf_{\eta, \nu, \mathbf{t}} \quad & \eta + \nu(\delta_w - 1) + \sum_{k=1}^K \bar{\pi}_k t_k \\ \text{s.t.} \quad & t_k + \eta - \nu \geq v_k(x, \tau), \quad \forall k \in \mathcal{K}, \\ & 0 \leq t_k \leq 2\nu, \quad \forall k \in \mathcal{K}, \\ & \nu \geq 0, \quad \eta \in \mathbb{R}. \end{aligned}$$

It remains to reformulate $v_k(x, \tau)$. The hard GMM assignment rule is

$$\kappa(i) = \arg \max_{\ell \in \mathcal{K}} \gamma_{i\ell}, \quad \mathcal{I}_k = \{i : \kappa(i) = k\}.$$

The representative set satisfies $\mathcal{I}_k^{\text{rep}} \subseteq \mathcal{I}_k$. Hence each retained atom $i \in \mathcal{I}_k^{\text{rep}}$ is associated with regime k and is evaluated under the regime-conditional support set Ξ_k .

By definition,

$$\mathcal{P}_k = \left\{ P_k \in \mathcal{P}(\Xi_k) : W_1(P_k, \hat{P}_k^{\text{rep}}) \leq \varrho_k \right\}, \quad \hat{P}_k^{\text{rep}} = \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \delta_{\hat{\omega}_i}.$$

Since $f(x, \tau, \cdot)$ is upper semicontinuous and Ξ_k is compact, strong duality holds with no duality gap, and the Kantorovich dual representation of the Wasserstein robust expectation [20, 25] gives

$$v_k(x, \tau) = \inf_{\lambda_k \geq 0} \left\{ \lambda_k \varrho_k + \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} \sup_{\boldsymbol{\omega} \in \Xi_k} [f(x, \tau, \boldsymbol{\omega}) - \lambda_k \|\boldsymbol{\omega} - \hat{\omega}_i\|] \right\}.$$

Since

$$f(x, \tau, \boldsymbol{\omega}) = \max_{m \in \mathcal{M}_+} \left\{ \tilde{\mathbf{a}}_m(x)^\top \boldsymbol{\omega} + \tilde{b}_m(x, \tau) \right\},$$

the inner supremum is bounded above by s_{ik} if and only if, for all $m \in \mathcal{M}_+$,

$$s_{ik} \geq \sup_{\boldsymbol{\omega} \in \Xi_k} \left[\tilde{\mathbf{a}}_m(x)^\top \boldsymbol{\omega} + \tilde{b}_m(x, \tau) - \lambda_k \|\boldsymbol{\omega} - \hat{\omega}_i\| \right].$$

Using the dual-norm identity

$$-\lambda_k \|\boldsymbol{\omega} - \hat{\omega}_i\| = \inf_{\|\boldsymbol{\zeta}_{ikm}\|_* \leq \lambda_k} \boldsymbol{\zeta}_{ikm}^\top (\boldsymbol{\omega} - \hat{\omega}_i),$$

and applying Sion's minimax theorem [35] to the continuous function—affine in $\boldsymbol{\omega}$ and in $\boldsymbol{\zeta}_{ikm}$ —over the compact convex sets Ξ_k and $\{\boldsymbol{\zeta} : \|\boldsymbol{\zeta}\|_* \leq \lambda_k\}$, we obtain

$$\begin{aligned} & \sup_{\boldsymbol{\omega} \in \Xi_k} \left[\tilde{\mathbf{a}}_m(x)^\top \boldsymbol{\omega} - \lambda_k \|\boldsymbol{\omega} - \hat{\omega}_i\| \right] \\ &= \inf_{\|\boldsymbol{\zeta}_{ikm}\|_* \leq \lambda_k} \left[S_{\Xi_k}(\tilde{\mathbf{a}}_m(x) + \boldsymbol{\zeta}_{ikm}) - \boldsymbol{\zeta}_{ikm}^\top \hat{\omega}_i \right], \end{aligned}$$

where

$$S_{\Xi_k}(\mathbf{v}) = \sup_{\boldsymbol{\omega} \in \Xi_k} \mathbf{v}^\top \boldsymbol{\omega}$$

is the support function of the regime-conditional support set Ξ_k .

Therefore, $v_k(x, \tau)$ is represented by variables $\lambda_k \geq 0$, s_{ik} , and $\boldsymbol{\zeta}_{ikm}$ satisfying

$$\begin{aligned} s_{ik} &\geq \tilde{b}_m(x, \tau) + S_{\Xi_k}(\tilde{\mathbf{a}}_m(x) + \boldsymbol{\zeta}_{ikm}) - \boldsymbol{\zeta}_{ikm}^\top \hat{\boldsymbol{\omega}}_i, \\ \|\boldsymbol{\zeta}_{ikm}\|_* &\leq \lambda_k, \quad \forall i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+. \end{aligned}$$

Combining the outer weight dual and the local Wasserstein dual yields

$$\begin{aligned} \Phi(x, \tau) &= \min_{\eta, \nu, \mathbf{t}, \boldsymbol{\lambda}, \mathbf{s}, \boldsymbol{\zeta}} \quad \eta + \nu(\delta_w - 1) + \sum_{k=1}^K \bar{\pi}_k t_k \\ \text{s.t.} \quad &t_k + \eta - \nu \geq \lambda_k \varrho_k + \frac{1}{q_k} \sum_{i \in \mathcal{I}_k^{\text{rep}}} s_{ik}, \quad \forall k \in \mathcal{K}, \\ &s_{ik} \geq \tilde{b}_m(x, \tau) + S_{\Xi_k}(\tilde{\mathbf{a}}_m(x) + \boldsymbol{\zeta}_{ikm}) - \boldsymbol{\zeta}_{ikm}^\top \hat{\boldsymbol{\omega}}_i, \\ &\forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \\ &\|\boldsymbol{\zeta}_{ikm}\|_* \leq \lambda_k, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k^{\text{rep}}, m \in \mathcal{M}_+, \\ &0 \leq t_k \leq 2\nu, \quad \forall k \in \mathcal{K}, \\ &\lambda_k \geq 0, \quad \nu \geq 0, \quad \eta \in \mathbb{R}, \quad s_{ik} \in \mathbb{R}, \quad \boldsymbol{\zeta}_{ikm} \in \mathbb{R}^{|\mathcal{B}|}. \end{aligned}$$

This is the deterministic reformulation stated in Theorem 1. $\square$

References

- [1] Billinton, R., Allan, R.N.: Reliability Evaluation of Power Systems, 2nd edn. Plenum Press, New York (1996)
- [2] Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research* **21**(5), 1154–1157 (1973)
- [3] Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming, 2nd edn. Springer, New York (2011)
- [4] Zhang, P., Li, F.: Chance constrained programming for optimal power flow under uncertainty. *IEEE Transactions on Power Systems* **26**(4), 2417–2424 (2011)
- [5] Vrakopoulou, M., Margellos, K., Lygeros, J., Andersson, G.: Chance constrained reserve scheduling using uncertain wind power forecasts. *IFAC Proceedings Volumes* **46**(27), 71–76 (2013)
- [6] Figueroa, J., Bubbar, K., Young-Morris, G.: An open-source tool for composite power system reliability assessment in Julia. *Energies* **17**(20), 5023 (2024) <https://doi.org/10.3390/en17205023>
- [7] Jia, Y., Hug, G., Shen, S.: Chance-constrained economic dispatch with joint constraints. *IEEE Transactions on Power Systems* **36**(2), 1206–1217 (2021)

- [8] Lubin, M., Dvorkin, Y., Backhaus, S.: Robust and chance-constrained optimal power flow. *IEEE Transactions on Power Systems* **31**(5), 3847–3856 (2016)
- [9] Peña-Ordieres, A., Molzahn, D.K., Roald, L., Wächter, A.: Improved formulations for joint chance-constrained AC optimal power flow. *Electric Power Systems Research* **186**, 106349 (2020)
- [10] Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3), 595–612 (2010)
- [11] Wang, Z., Glynn, P., Ye, Y.: Likelihood robust optimization for data-driven problems. *Computational Management Science* **13**(2), 241–261 (2016)
- [12] Bertsimas, D., Sim, M.: The price of robustness. *Operations Research* **52**(1), 35–53 (2004)
- [13] Rahimian, H., Mehrotra, S.: Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization* **3**, 1–85 (2022) <https://doi.org/10.5802/ojmo.15>
- [14] Kuhn, D., Shafiee, S., Wiesemann, W.: Distributionally robust optimization. *Acta Numerica* (2025)
- [15] Duan, C., Fang, W., Jiang, L., Yao, L., Liu, J.: Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric. *IEEE Transactions on Power Systems* **33**(5), 4924–4936 (2018) <https://doi.org/10.1109/TPWRS.2018.2799247>
- [16] Zhou, A., Yang, M., Wang, M., Zhang, Y.: A linear programming approximation of distributionally robust chance-constrained dispatch with Wasserstein distance. *IEEE Transactions on Power Systems* **35**(5), 3366–3377 (2020) <https://doi.org/10.1109/TPWRS.2020.2978934>
- [17] Arrigo, A., Ordoudis, C., Kazempour, J., Pinson, P., Morales, J.M.: Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation. *European Journal of Operational Research* **296**(1), 304–322 (2022) <https://doi.org/10.1016/j.ejor.2021.03.033>
- [18] Guo, Y., Baker, K., Dall’Anese, E., Hu, Z., Summers, T.H.: Data-based distributionally robust stochastic optimal power flow—part I: Methodologies. *IEEE Transactions on Power Systems* **34**(2), 1483–1492 (2019) <https://doi.org/10.1109/TPWRS.2018.2878385>
- [19] Villani, C.: *Optimal Transport: Old and New*. Springer, Berlin (2009)

- [20] Mohajerin Esfahani, P., Kuhn, D.: Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1–2), 115–166 (2018)
- [21] Blanchet, J., Murthy, K.: Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research* **44**(2), 565–600 (2019)
- [22] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22 (1977)
- [23] Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, New York (2006)
- [24] Esteban-Pérez, A., Morales, J.M.: Distributionally robust optimal power flow with contextual information. *European Journal of Operational Research* **306**(3), 1047–1058 (2023)
- [25] Gao, R., Kleywegt, A.J.: Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research* **48**(2), 603–655 (2023)
- [26] Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. *Journal of Risk* **2**(3), 21–42 (2000)
- [27] Lloyd, S.: Least squares quantization in PCM. *IEEE Transactions on Information Theory* **28**(2), 129–137 (1982)
- [28] Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications* **24**(2–3), 187–206 (2003) <https://doi.org/10.1023/A:1021805924152>
- [29] Billinton, R., Kumar, S., Chowdhury, N., Chu, K., Debnath, K., Goel, L., Khan, E., Kos, P., Nourbakhsh, G., Oteng-Adjei, J.: A reliability test system for educational purposes: RBTS basic data. *IEEE Transactions on Power Systems* **4**(3), 1238–1244 (1989) <https://doi.org/10.1109/59.325505>
- [30] IEEE Reliability Test System Task Force: IEEE reliability test system. *IEEE Transactions on Power Apparatus and Systems* **PAS-98**(6), 2047–2054 (1979)
- [31] Grigg, C., Wong, P., Albrecht, P., Allan, R., Bhavaraju, M., Billinton, R., Chen, Q., Fong, C., Haddad, S., Kuruganty, S., Li, W., Mukerji, R., Patton, D., Rau, N., Reppen, D., Schneider, A., Shahidehpour, M., Singh, C.: The IEEE reliability test system-1996. *IEEE Transactions on Power Systems* **14**(3), 1010–1020 (1999) <https://doi.org/10.1109/59.780914>
- [32] Barrows, C., Bloom, A., Ehlen, A., Ikaheimo, J., Jorgenson, J., Krishnamurthy, D., Lau, J., McBennett, B., O’Connell, M., Preston, E., Staid, A., Stephen, G.,

- Watson, J.-P.: The IEEE reliability test system: A proposed 2019 update. *IEEE Transactions on Power Systems* **35**(1), 119–127 (2020) <https://doi.org/10.1109/TPWRS.2019.2925557>
- [33] Chen, Y., Xu, J.: Solar and wind power data from the Chinese State Grid renewable energy generation forecasting competition. *Scientific Data* **9**(1), 577 (2022) <https://doi.org/10.1038/s41597-022-01696-6>
- [34] Jolliffe, I.T.: *Principal Component Analysis (PCA)*, 2nd edn. Springer, New York (2002)
- [35] Sion, M.: On general minimax theorems. *Pacific Journal of Mathematics* **8**(1), 171–176 (1958)