

A Frequency-Aware Representative-Day Framework for Energy Storage Capacity Planning Under High Penetration of Renewable Energy

Ling Yunxiao

*Department of Electrical and Electronic Engineering
The Hong Kong Polytechnic University
Hong Kong SAR
yunxiao.ling@connect.polyu.hk*

Qin Wang

*Department of Electrical and Electronic Engineering
The Hong Kong Polytechnic University
Hong Kong SAR
qin-ee.wang@polyu.edu.hk*

Abstract—High renewable penetration reduces inertia and increases variability, raising the risk of unacceptable post-disturbance frequency nadirs. This paper develops an integrated planning pipeline that couples representative-day reduction with tractable nadir-security constraints for battery energy storage system (BESS) sizing and dispatch. The representative-day module combines clustering with stress-based screening and an optimization-based reduction that retains frequency-critical tail conditions. The frequency module derives a linearizable nadir constraint from a reduced-order frequency response approximation, enabling direct embedding into a mixed-integer linear programming (MILP) expansion model. Case studies show that, compared with standard representative-day selection and frequency-blind planning baselines, the proposed pipeline significantly reduces nadir violations with only a marginal increase in total cost.

Index Terms—Battery energy storage system, representative day, capacity expansion planning, frequency nadir, low inertia.

I. INTRODUCTION

High penetrations of inverter-based renewable generation reduce system inertia and make grid frequency more sensitive to imbalances. After a disturbance, the frequency nadir (the lowest frequency reached) must remain above emergency limits to avoid load shedding or generator trips. BESS can respond almost instantaneously to frequency deviations, providing fast primary regulation and thereby improving frequency nadir. However, conventional expansion models often treat storage only as an energy resource and ignore its dynamic support role. As a result, plans that appear cost-optimal can fail to meet stability requirements when a severe event occurs.

Beyond clustering-based representative days used in planning, optimization-driven selection that preserves both typical profiles and extremes has proven effective for multi-energy design but is largely frequency-blind [1]. In parallel, a growing body of work embeds dynamic frequency-security into operation and planning via MSFR/ASFR-based nadir, rate of change of frequency, and quasi-steady-state constraints, with linearization or decomposition to retain tractability [2]. Recent advances include identify-then-eliminate bottlenecks with linearized post-disturbance nadir limits in MILP [3], scalable SAVLR-based stochastic planning with new linearizable nadir constraints and rolling horizon [4], and data-driven decision-tree surrogates that encode frequency-quality tar-

gets into planning under low inertia [5]. System-level formulations have also moved toward risk-aware coordination: chance-constrained bi-level regulation with BESSs in distribution networks [6] and two-stage robust unit-commitment for integrated transmission and distribution that mobilizes distributed energy resources (DERs) inertia-based reserves [7]. Complementary resilience-oriented multi-energy system planning leverages generalized storage (pipe/storage/thermal inertia) but often emphasizes extreme-event resilience rather than routine frequency-quality constraints [8]. Meanwhile, representative-day selection and frequency-security-aware planning have both been extensively studied. However, existing studies commonly address these elements in a loosely coupled manner—for example, scenario reduction that is not explicitly tailored to preserve frequency-critical nadir conditions, or frequency-security constraints evaluated on a small set of operating snapshots that may not be consistent with the reduction procedure. This can lead to systematic underestimation of fast-response needs when planning under low inertia and high renewable variability.

Motivated by this coupling issue, this paper develops a frequency-aware expansion-planning pipeline that integrates (i) representative-day reduction with explicit retention of stress conditions relevant to frequency security, and (ii) a tractable nadir-security constraint that can be embedded into a MILP capacity expansion model. Specifically, the representative-day module uses a two-stage selection: Stage 1 forms a candidate pool via clustering-based typical patterns and stress-based screening; Stage 2 performs optimization-based reduction to preserve annual variability while retaining nadir stress conditions. The frequency-security module employs a modal-equivalent approximation to obtain a closed-form nadir expression, which is then linearized to enforce minimum-frequency limits for each contingency within the MILP. The resulting model co-optimizes BESS capacity and operation while maintaining consistency between reduced operating scenarios and frequency-security requirements.

Overall, the proposed framework ensures that both typical daily variations and extreme high-stress events are represented in the planning, and that the resulting expansion respects dynamic frequency limits. Case studies on a benchmark system

demonstrate that our approach significantly raises the post-contingency nadir under high renewable penetration, compared to frequency-blind planning with only a modest increase in cost. These results confirm that explicitly accounting for frequency stability via the proposed representative-day framework is essential for reliable expansion planning in low-inertia grids.

II. REPRESENTATIVE DAY SELECTION

This section presents a two-stage method to select representative operating days that not only approximate typical system behavior but also preserve critical extreme scenarios. The method ensures that reduced operating scenarios retain both economic and resilience information for planning.

A. Candidate Pool Construction

Each day $d' \in \mathcal{D}^{\text{all}}$ is transformed into a feature vector:

$$\mathbf{z}^{d'} = [P_{\text{load}}^{d'}(1), \dots, P_{\text{load}}^{d'}(24), P_{\text{RES}}^{d'}(1), \dots, P_{\text{RES}}^{d'}(24), T^{d'}(1), \dots, T^{d'}(24), \dots]^\top \quad (1)$$

Before clustering, all features are typically normalized (e.g., load scaled by the annual system peak, PV output scaled by installed capacity, temperature standardized) so that no single channel dominates the distance metric purely due to magnitude. The set $\mathcal{D}^{\text{all}}$ is partitioned via clustering (e.g., k -medoids or hierarchical clustering). Let $G_1, \dots, G_{K_{\text{norm}}}$ denote the resulting clusters, where $G_c \subseteq \mathcal{D}^{\text{all}}$. For each cluster c , we define its medoid:

$$d_c^* = \arg \min_{d \in G_c} \sum_{d' \in G_c} \text{dist}(d', d), \quad (2)$$

and collect these medoids as:

$$C_{\text{norm}} = \{d_c^* : c = 1, \dots, K_{\text{norm}}\}. \quad (3)$$

Standardized feature vectors are clustered using methods such as k -medoids to extract normal operating patterns. Medoids from each cluster form C_{norm} . Separately, stress-based filtering (e.g., peak demand, low renewables, extreme weather) identifies C_{ext} . The full candidate pool is:

$$C = C_{\text{norm}} \cup C_{\text{ext}} \quad (4)$$

Because a medoid is an actual historical day rather than an artificial average profile, each d_c^* can directly be simulated in unit commitment (UC) constraints.

Independent of clustering, we identify days that create system stress but may occur rarely. Typical triggers include:

- Annual peak demand day (thermal loading stress);
- Simultaneous high demand and low renewable availability (resource adequacy stress);
- Abnormal weather (heatwave, cold surge) or post-contingency/post-disaster restoration days (resilience stress);
- High distributed generation with low native load (reverse power flow, over-voltage stress).

All days meeting any such stress criterion are collected into:

$$C_{\text{ext}} \subseteq \mathcal{D}^{\text{all}}. \quad (5)$$

These extreme candidates may include both observed historical days and synthetically constructed stress scenarios.

The Stage-1 output is the unified candidate pool:

$$C = C_{\text{norm}} \cup C_{\text{ext}}. \quad (6)$$

This pool typically contains on the order of 10-30 days, substantially fewer than $|\mathcal{D}^{\text{all}}|$ but still too large to embed directly in a capacity expansion MILP. Stage 2 performs optimization-based reduction on C .

B. Optimization-Based Representative Day Selection

Stage 2 formulates a MILP that (i) selects a small subset of C and (ii) assigns each original day $d' \in \mathcal{D}^{\text{all}}$ to exactly one selected representative. The formulation jointly minimizes annual representation error and penalizes the exclusion of extreme days.

Decision variables and notation: The Stage-2 MILP selects representative days from the candidate pool C and assigns each original day $d' \in \mathcal{D}^{\text{all}}$ to exactly one selected representative. The binary variable $x_d \in \{0, 1\}$ indicates whether candidate day $d \in C$ is selected. The assignment variable $y_{d', d} \geq 0$ indicates whether original day d' is represented by candidate day d . The nonnegative weight w_d counts how many original days are assigned to representative day d , i.e., $w_d = \sum_{d' \in \mathcal{D}^{\text{all}}} y_{d', d}$. Parameters include the distance $\text{dist}_{d', d}$ between feature vectors, the extreme-day flag $\text{riskflag}_d \in \{0, 1\}$, the maximum number of representatives K , the penalty coefficient λ , and the big- M constant M .

1) Objective Function

The MILP objective is:

$$\min \sum_{d' \in \mathcal{D}^{\text{all}}} \sum_{d \in C} \text{dist}_{d', d} y_{d', d} + \lambda \sum_{d \in C} \text{riskflag}_d (1 - x_d). \quad (7)$$

$\text{dist}_{d', d} \geq 0$: distance between $\mathbf{z}^{d'}$ and $\mathbf{z}^d$. A weighted norm is typically used,

$$\text{dist}_{d', d} = \|\mathbf{W}(\mathbf{z}^{d'} - \mathbf{z}^d)\|_2,$$

where $\mathbf{W}$ is a diagonal weighting matrix emphasizing critical hours such as evening peaks, low-renewable intervals, or voltage-stress periods.

$\text{riskflag}_d \in 0, 1$: indicator of whether candidate day $d \in C$ is classified as “extreme,” i.e.,

$$\text{riskflag}_d = \begin{cases} 1, & d \in C_{\text{ext}}, \\ 0, & d \notin C_{\text{ext}}. \end{cases}$$

- The first part in (7) enforces fidelity to the full-year operational landscape: if day d' is “far” (in $\text{dist}_{d', d}$) from all selected representatives, the cost increases.
- The second part in (7) enforces resilience: excluding an extreme candidate day ($\text{riskflag}_d = 1$) incurs a penalty λ . A sufficiently large λ forces the MILP to keep most or all extreme days.

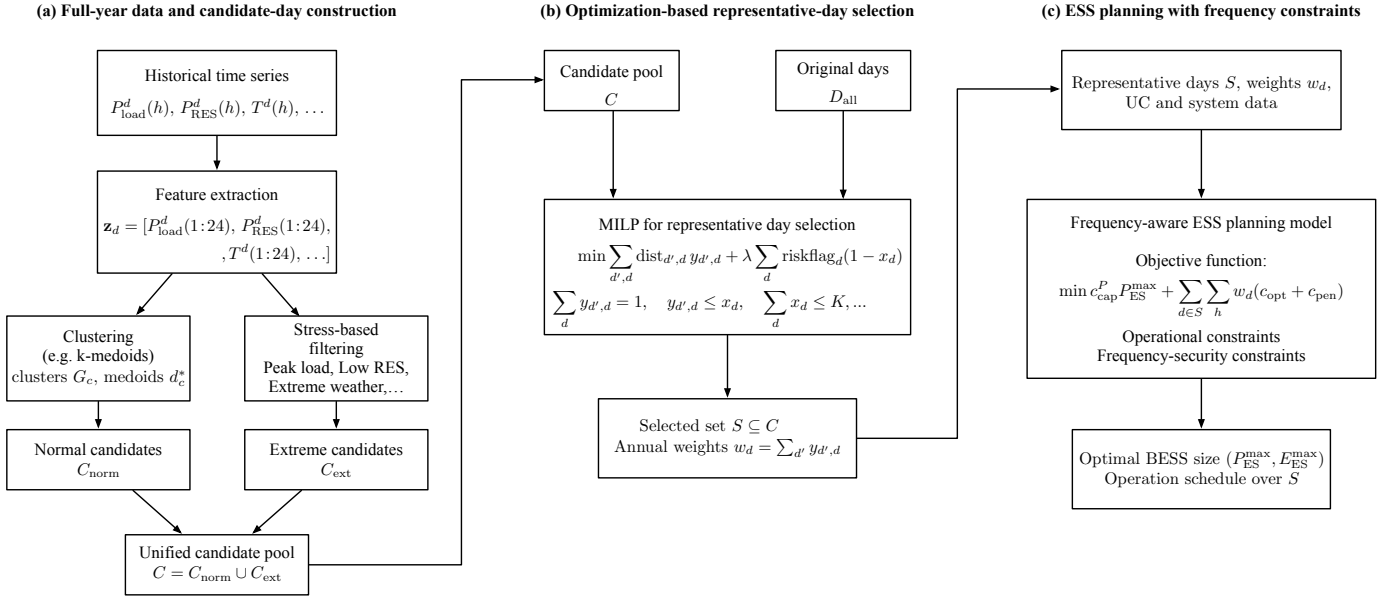


Fig. 1. The flowchart of the proposed method.

2) Constraints

Each original day must be assigned to exactly one representative:

$$\sum_{d \in \mathcal{C}} y_{d',d} = 1, \quad \forall d' \in \mathcal{D}^{\text{all}}. \quad (8)$$

Assignment is only allowed to selected candidates:

$$y_{d',d} \leq x_d, \quad \forall d' \in \mathcal{D}^{\text{all}}, \forall d \in \mathcal{C}. \quad (9)$$

The number of selected representative days is capped:

$$\sum_{d \in \mathcal{C}} x_d \leq K. \quad (10)$$

Define the annual weight of each candidate $d \in \mathcal{C}$ as the number of calendar days assigned to it:

$$w_d = \sum_{d' \in \mathcal{D}^{\text{all}}} y_{d',d}, \quad \forall d \in \mathcal{C}. \quad (11)$$

The weights must sum to the total number of days in the year:

$$\sum_{d \in \mathcal{C}} w_d = D_{\text{year}}. \quad (12)$$

Weights are only allowed to be positive if the day is selected:

$$w_d \leq M x_d, \quad \forall d \in \mathcal{C}. \quad (13)$$

Finally, impose variable domains:

$$x_d \in \{0, 1\}, \quad y_{d',d} \geq 0, \quad w_d \geq 0. \quad (14)$$

The MILP (1)-(8) yields:

- a selected subset $S = \{d \in \mathcal{C} \mid x_d^* = 1\}$,
- corresponding weights $w_{d \in S}^*$, which quantify how many real days each representative stands for in annualization.

In some applications, certain extreme scenarios (e.g., the annual peak load day, the mandated restoration day) must always be enforced in planning. This requirement can be added as

$$x_d = 1, \quad \forall d \in C_{\text{ext,req}} \subseteq C_{\text{ext}}, \quad (15)$$

where $C_{\text{ext,req}}$ is a subset of extreme days that are deemed non-negotiable. Under (15), those days are guaranteed to enter S , regardless of λ . The penalty term in (7) then primarily governs “soft” extreme days outside $C_{\text{ext,req}}$.

III. ESS PLANNING MODEL

After solving (7)-(15), the resulting S and $\{w_d^*\}_{d \in S}$ are embedded in the lower-level ess planning model.

Decision variables and indices: The planning MILP is solved over representative days $d \in S$ and hours $h \in \mathcal{H} = \{1, \dots, 24\}$ with annual weights w_d . The investment (planning) variables are the BESS power rating $P_{\text{ES}}^{\text{max}}$ (MW) and energy rating $E_{\text{ES}}^{\text{max}}$ (MWh). The operational variables include BESS charging/discharging power $P_{d,h}^{\text{ch}} \geq 0, P_{d,h}^{\text{dis}} \geq 0$, and state of charge $E_{d,h}$. For conventional generators $i \in \mathcal{G}$, unit-commitment variables are $u_{i,d,h} \in \{0, 1\}$ and dispatch power $P_{i,d,h} \geq 0$. Renewable curtailment is represented by $P_{d,h}^{\text{curt}} \geq 0$, applied to the available renewable injections $P_{d,h}^{\text{wd}}$ and $P_{d,h}^{\text{pv}}$. Demand response is represented by $P_{d,h}^{\text{dr}} \geq 0$ subject to program limits. Other exogenous inputs include hourly load $L_{d,h}$, generator capacity limits $(P_{i,\min}, P_{i,\max})$, ramp limits (R_i^u, R_i^d) , minimum up/down times $(T_i^{\text{on}}, T_i^{\text{off}})$, and BESS efficiencies (η_c, η_d) .

A. Objective function

The objective function of the planning model is to minimize total cost, including investment cost of BESS, renewable energy curtailment cost, demand response penalty and system operating cost:

$$\begin{aligned} \min \quad & c_P^{\text{cap}} P_{\text{ES}}^{\text{max}} + \sum_{d \in S} w_d \sum_{h \in \mathcal{H}} \left[c^{\text{run}} P_{d,h}^{\text{dis}} + c^{\text{gen}} \sum_{i \in \mathcal{G}} P_{i,d,h} \right. \\ & \left. + c^{\text{curt}} P_{d,h}^{\text{curt}} + c^{\text{dr}} P_{d,h}^{\text{dr}} \right] \end{aligned} \quad (16)$$

B. Operational Constraints

$$\sum_{i \in \mathcal{G}} P_{i,d,h} + P_{d,h}^{\text{wd}} + P_{d,h}^{\text{pv}} - P_{d,h}^{\text{curt}} + P_{d,h}^{\text{dr}} + P_{d,h}^{\text{dis}} - P_{d,h}^{\text{ch}} = L_{d,h} \quad \forall d \in \mathcal{S}, h \in \mathcal{H}. \quad (17)$$

$$u_{i,d,h} P_{i,\min} \leq P_{i,d,h} \leq u_{i,d,h} P_{i,\max}, \quad \forall i \in \mathcal{G}, d \in \mathcal{S}, h \in \mathcal{H}. \quad (18)$$

$$-R_i^d \leq P_{i,d,h} - P_{i,d,h-1} \leq R_i^u, \quad \forall i \in \mathcal{G}, d \in \mathcal{S}, h \in \mathcal{H} \setminus \{1\}. \quad (19)$$

$$\sum_{k=h}^{h+T_i^{\text{on}}-1} u_{i,d,k} \geq T_i^{\text{on}} (u_{i,d,h} - u_{i,d,h-1}), \quad \forall i \in \mathcal{G}, d \in \mathcal{S}, h \in \{2, \dots, 24 - T_i^{\text{on}} + 1\}. \quad (20)$$

$$\sum_{k=h}^{h+T_i^{\text{off}}-1} (1 - u_{i,d,k}) \geq T_i^{\text{off}} (u_{i,d,h-1} - u_{i,d,h}), \quad \forall i \in \mathcal{G}, d \in \mathcal{S}, 2 \leq h \leq 24 - T_i^{\text{off}} + 1. \quad (21)$$

$$0 \leq P_{d,h}^{\text{ch}} \leq P_{\text{ES}}^{\text{max}}, \quad \forall d, h \quad (22)$$

$$0 \leq P_{d,h}^{\text{dis}} \leq P_{\text{ES}}^{\text{max}}, \quad \forall d, h \quad (23)$$

$$0 \leq E_{d,h} \leq E_{\text{ES}}^{\text{max}}, \quad \forall d, h \quad (24)$$

$$E_{d,h} = E_{d,h-1} + \eta_c P_{d,h}^{\text{ch}} - \frac{1}{\eta_d} P_{d,h}^{\text{dis}}, \quad \forall d \in \mathcal{S}, h \in \mathcal{H} \setminus \{1\}. \quad (25)$$

$$E_{d,24} = E_{d,0}, \quad \forall d \in \mathcal{S} \quad (26)$$

Constraint (17) indicates the power balance of the system. The limit on output power of thermal unit is described by constraint (18) and the ramp-up and ramp-down limit is imposed by constraint (19). The minimum on/off time maintained for conventional units is limited by constraints (20) and (21), respectively. Constraints (22)-(24) indicate the BESS operating limit.

C. Frequency Constraints

To guarantee secure system frequency during operation, we embed an explicit frequency-nadir security constraint into the planning MILP. For each considered contingency with net power imbalance P_0 , the constraint requires the post-contingency frequency nadir to remain above a minimum threshold, i.e., $\omega_{\text{nadir}} \geq \omega_{\min}$.

$$G_{u_i}(s) = J_{u_i} s + D_{u_i} + \frac{1}{K_{u_i}}, \quad (27)$$

$$\Delta\omega_{\text{cm}}(s) = -\frac{P_0}{J_{\text{cm}} s^2 + D_{\text{cm}} s + 1/K_{\text{cm}}}, \quad (28)$$

$$J_{\text{cm}} = \sum_{i=1}^n w_i J_{u_i}, \quad D_{\text{cm}} = \sum_{i=1}^n w_i D_{u_i}, \quad \frac{1}{K_{\text{cm}}} = \sum_{i=1}^n \frac{w_i}{K_{u_i}}. \quad (29)$$

$$\Delta\omega_{\text{nadir}} = -P_0 \sqrt{\frac{K_{\text{cm}}}{J_{\text{cm}}}} \exp\left(\frac{\sigma}{\omega_d} \arctan\left(-\frac{\omega_d}{\sigma}\right)\right) \quad (30)$$

$$\approx -\frac{P_0}{\frac{e-1}{2} D_{\text{cm}} + \sqrt{\frac{J_{\text{cm}}}{K_{\text{cm}}}}} \quad f_{\text{nadir}} \geq f_{\min} \quad (31)$$

In constraint (27), J_u^i , D_u^i and K_u^i represent the i -th devices' modal inertia, modal damping and model regulation coefficient, respectively. By summing up the modal inertia, modal damping and model regulation coefficient, common-mode inertia, damping and regulation coefficient can be obtained. These summed-up value can be used in constraint (28) to obtain the frequency variation. Constraint (30) describes the frequency nadir value in the system.

IV. CASE STUDY

A. System Description

In this section, the proposed representative day selection model and ESS planning model is implemented on a real-world province-level power grid in China. In 2024, renewable sources such as hydro, wind, and solar accounted for over 40% of the total generation capacity in the analyzed power system. All simulations are performed in the Julia environment on a PC equipped with an Apple M1 Pro CPU and 16 GB RAM.

In the following case studies, the three cases are considered to verify the effectiveness of the proposed method.

- Case 1: Use existing system devices only, as the baseline.
- Case 2: Planning ESS without considering frequency constraints with proposed representative day selection model.
- Case 3: Planning ESS considering frequency constraints with proposed representative day selection model.

B. Results and analysis

Throughout the 24-hour period, Case 3 consistently maintains higher nadir frequencies than Case 1 and Case 2, especially during peak demand hours. In particular, Case 1 and 2 exhibit several deep frequency dips that exceed the under-frequency threshold, whereas Case 3's frequency stays above those limits in all hours. These visual trends align with the quantitative metrics: Case 3 yields only a 0.4% frequency-violation rate versus 3.1% for Case 2. Figure 3 illustrates that at the worst hour further confirms that Case 3's response just meets the nadir constraint, while Case 1 and 2 violates it by a larger margin. In short, our proposed method effectively preserves security across the day, reflecting the framework's goal of maintaining frequency stability with only a modest cost increase.

Key planning outcomes for Cases 2 and 3 are summarized in Table I. In Case 2, the optimal BESS capacity is 3379 MW, compared to 3651 MW in Case 3. This higher reserve in Case 3 comes with a slightly higher total cost, but dramatically improves performance. The detailed comparison highlights the strengths of the frequency-aware planning approach. An

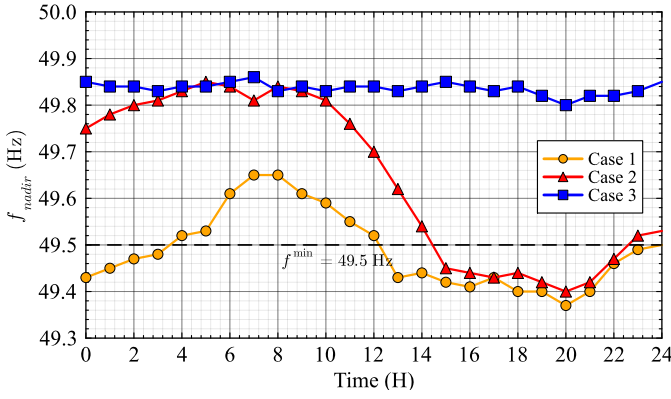


Fig. 2. Comparison of f_{nadir} on the selected day across three cases.

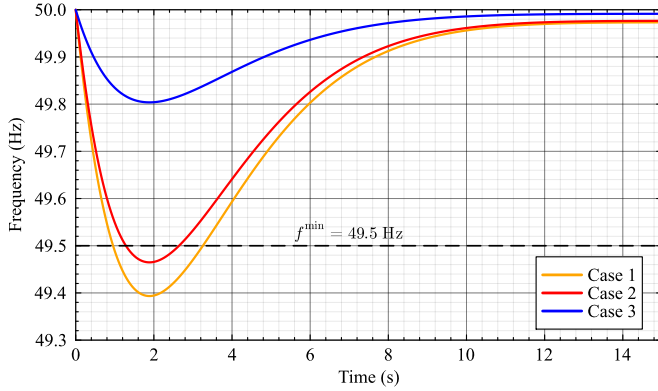


Fig. 3. Frequency trajectories across three case.

TABLE I
COMPARISON RESULTS OF THREE METHODS

Case No.	BESS Cap. (MW)	Freq. Vio. Rate	RES Curt. Rate	Total Cost (M\$)	CPU Time (s)
1	2315	4.3%	7.2%	532.7	377
2	3379	3.1%	4.5%	793.2	1982
3	3651	0.4%	3.3%	813.4	3721

additional 272 MW of storage in Case 3 directly enables much better frequency performance and less curtailment. In other words, enforcing the nadir constraints leads to a more robust plan: nearly all contingency events now remain above the minimum frequency limit, at the expense of a small cost increase. By contrast, Case 2's cheaper plan with proposed representative-day selection fails to capture some extreme conditions. This difference is primarily driven by the scenario-reduction component: under the same nadir-security formulation, the standard representative-day selection used in Case 2 does not sufficiently retain some nadir high-stress operating conditions, which leads to an underestimation of required fast-response capacity. In contrast, Case 3's nadir-preserving representative-day selection maintains those stress conditions in the reduced set, yielding a more conservative but more frequency-secure plan with fewer nadir violations and lower renewable curtailment.

V. CONCLUSION

In conclusion, this paper has presented a frequency-aware energy storage planning framework for systems with high renewable penetration. The framework includes a two-stage representative-day selection method to capture both typical and extreme operating conditions, and a mixed-integer linear programming formulation for battery storage planning that explicitly enforces frequency-nadir constraints. By incorporating dynamic frequency stability into planning, the proposed approach improves upon planning baselines that do not enforce explicit nadir-security constraints or that use standard representative-day reduction without nadir preservation. It ensures adequate frequency response during severe events and enables higher renewable integration, thereby enhancing system security and operational flexibility. Simulation results demonstrate that retaining frequency-critical nadir conditions during representative-day reduction, together with explicit nadir-security constraints, yields improved frequency-security metrics with modest additional cost. Future research includes developing higher-fidelity, nonlinear models of frequency dynamics to better capture the complex behaviors of these systems, as well as incorporating stochastic representations of demand and renewable generation uncertainty (e.g., probabilistic forecasting and robust optimization). These enhancements aim to improve the generality and real-world applicability of the results by providing broader scenario coverage and empirical validation.

REFERENCES

- [1] C. Zhang, L. Liu, H. Cheng, D. Liu, J. Zhang, and G. Li, "Frequency-constrained co-planning of generation and energy storage with high-penetration renewable energy," *J. Mod. Power Syst. Clean Energy*, vol. 9, no. 4, pp. 760–775, Jun. 2021.
- [2] H. Saboori, R. Hemmati, and M. A. Jirdehi, "Reliability improvement in radial electrical distribution network by optimal planning of energy storage systems," *Energy*, vol. 93, pp. 2299–2312, Dec. 2015.
- [3] R. Li, W. Wang, and M. Xia, "Cooperative planning of active distribution system with renewable energy sources and energy storage systems," *IEEE Access*, vol. 6, pp. 5916–5926, Dec. 2017.
- [4] M. Nick, R. Cherkaoui, and M. Paolone, "Optimal planning of distributed energy storage systems in active distribution networks embedding grid reconfiguration," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1577–1590, Aug. 2017.
- [5] H. Alsharif, M. Jalili, and K. N. Hasan, "Power system frequency stability using optimal sizing and placement of battery energy storage system under uncertainty," *J. Energy Storage*, vol. 50, p. 104610, Jun. 2022.
- [6] Y. Sun, S. Bahrami, V. W. Wong, and L. Lampe, "Chance-constrained frequency regulation with energy storage systems in distribution networks," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 215–228, Jun. 2019.
- [7] M. Nazemi, M. Moeini-Agtaie, M. Fotuhi-Firuzabad, and P. Dehghanian, "Energy storage planning for enhanced resilience of power distribution networks against earthquakes," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 795–806, Mar. 2019.
- [8] A. Oshnoei, M. Kheradmandi, and S. M. Muyeen, "Robust control scheme for distributed battery energy storage systems in load frequency control," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4781–4791, May 2020.