

ORIGINAL RESEARCH

Open access

A Generative Scenario Benchmark for Cyber-Physical Smart-Grid Resilience Using Synthetic Attacks, Weather Extremes, and DER Recovery Actions

Priya Nambiar^{1*}, Suresh Menon¹, Kavita Pillai²

Abstract

Resilience assessment in smart grids is often limited by the small number of hand-crafted scenarios. Generative AI can create diverse, realistic synthetic stress events, providing a more rigorous benchmark. This article presents a generative scenario benchmark that combines synthetic cyber-attacks and extreme weather events with a standardized set of DER recovery actions to evaluate grid resilience. The benchmark is designed for comparative testing of cyber-physical distribution-grid controllers. The benchmark uses generative models to produce a large library of attack and weather scenarios. DER recovery actions are catalogued, and several baseline controllers are evaluated using standardized resilience metrics on a distribution feeder. The benchmark reveals distinct performance profiles: reinforcement learning excels in speed, robust optimization improves worst-case protection, and rule-based control provides simple but conservative responses. Three tables summarize the attack types, recovery actions, and comparative performance. The proposed benchmark offers a scalable, reproducible, and diverse testbed for developing and comparing next-generation resilient grid controllers. It supports transparent evaluation under coupled cyber-physical stress rather than isolated contingency cases.

Keywords Cyber-physical resilience, Smart grid, Generative scenario benchmark, Synthetic attacks, Weather extremes, DER recovery

*Correspondence:

Priya Nambiar

priya.nambiar@gmail.com

¹ Department of Smart Grid Engineering, Faculty of Engineering, Indian Institute of Science, Bengaluru, India

² Department of Energy Optimization and Control, Faculty of Engineering, Anna University, Chennai, India

Introduction

Smart-grid resilience has become a central design objective because distribution networks increasingly face compound disturbances that combine cyber intrusions, distributed-energy variability, and climate-driven physical stress. Early resilience studies showed that extreme weather can degrade grid performance through correlated component failures, delayed restoration, and cascading service interruptions [1]. System-level resilience research further emphasized that conventional reliability indices are insufficient when the objective is not only to prevent outages but also to absorb, adapt, and recover from high-impact events [2]. This article responds to that gap by proposing a benchmark that deliberately combines cyber and physical stressors in a reproducible testing environment.

Existing resilience studies often rely on a small number of manually specified contingencies, which limits the diversity and statistical value of controller comparisons. Distribution-level restoration studies have shown that mobile generators, switching actions, and distributed resources can support post-disaster recovery, but their evaluation is usually tied to

specific outage maps or scenario assumptions [3]. Coordinated restoration models have advanced this field by integrating repair crews, mobile sources, and network reconfiguration, yet their benchmarks remain difficult to reproduce across laboratories [4]. A generative scenario benchmark can make these evaluations broader, repeatable, and more comparable.

Cyber-physical attacks add a second layer of complexity because they can alter measurements, disrupt communications, and mislead grid-control decisions. False data injection attacks have been identified as a major threat to modern power-system monitoring because they can bypass bad-data detection when designed with sufficient system knowledge [5]. More recent attack studies have shown that blind and data-driven variants can remain effective even when complete network information is unavailable [6]. A useful benchmark therefore needs to represent both physically meaningful grid disturbances and adversarial manipulation of cyber-layer signals.

Generative AI offers a practical route to scenario diversity because it can learn distributions of load, renewable output, attack timing, and weather intensity from historical or simulated data. Synthetic load and renewable generation studies using generative adversarial networks demonstrate that time-series generators can preserve spatial and temporal dependence while producing many plausible operating conditions [7]. Scenario-generation models based on dynamic spatiotemporal generative networks also indicate that nonlinear dependence among renewable sources can be reproduced for stress testing [8]. The contribution of this article is a concrete benchmark architecture that combines generative cyber-physical scenarios, predefined DER recovery actions, standardized metrics, and baseline controller comparisons.

Need for Resilience Benchmark

Existing benchmarks and their gaps

Current resilience evaluation methods provide useful concepts but remain fragmented across weather fragility analysis, cyber-attack detection, restoration scheduling, and DER control. Reviews of resilience practice show that studies often differ in their definitions, event severity assumptions, and recovery metrics, which makes direct comparison difficult [9]. Quantitative standardization work has similarly noted that resilience indicators are categorized inconsistently across planning, operation, and restoration phases [10]. A benchmark is needed to connect these pieces into one controlled experiment where the same scenarios, grid model, actions, and metrics are shared.

The rise of generative scenario generation

Generative scenario generation is attractive because resilience controllers must be tested on many rare but plausible events, not only on historical outages. GAN-based load synthesis has shown that synthetic demand profiles can retain multiscale temporal dynamics while expanding the available test set for energy applications [11]. Synthetic load group generation also demonstrates that spatial-temporal correlation can be preserved across customer groups, which is essential when distribution-feeder stress depends on geographically coupled demand [12]. These advances support the use of generative models as benchmark engines rather than merely as data-augmentation tools.

Scope and aims of this benchmark

The proposed benchmark covers coupled cyber-physical distribution-grid stress, including false data injection, denial-of-service, heatwaves, storms, and DER-based recovery actions. Prior resilience studies have emphasized that distributed energy resources, demand response, and microgrid formation can materially affect restoration trajectories after disturbances [13]. Reinforcement-learning and robust-control studies further show that different control paradigms can perform well under different uncertainty structures, which motivates direct benchmarking under shared scenarios [14]. The aim is to provide a reproducible evaluation protocol that reveals not only average performance but also worst-case vulnerability and scenario-specific controller behavior.

Benchmark Design Principles

Diversity and realism

The benchmark is designed to generate many distinct scenarios while preserving physical realism in grid states, weather exposure, and attack feasibility. Extreme-weather resilience studies show that correlated hazards such as storms and heatwaves should be represented through spatially structured impacts rather than independent component outages [1]. Cyber studies likewise show that attack realism depends on the attacker's information level, target selection, and ability to remain hidden from state-estimation filters [15]. The benchmark therefore samples both event intensity and operational context so that controllers face varied but technically plausible stress cases.

Reproducibility and extensibility

Reproducibility is achieved by fixing feeder data, random seeds, scenario metadata, controller interfaces, and metric definitions while allowing new generators or controllers to be added. Existing surveys of power-system resilience highlight that inconsistent scenario design and metric reporting are major barriers to comparing methods across studies [16]. The benchmark records every synthetic scenario as a structured object containing cyber events, weather fields, DER availability, initial grid state, and recovery-action limits. This design allows later researchers to extend the library without changing the baseline protocol.

Grid model and assumptions

The benchmark uses a radial distribution feeder with photovoltaic generation, battery storage, flexible loads, sectionalizing switches, and candidate microgrid boundaries. Distribution-restoration literature shows that resilience assessment must represent service priorities, switching constraints, crew or resource limitations, and DER availability to avoid unrealistic recovery outcomes [17]. The assumed feeder includes critical and noncritical load classes, communication-dependent measurements, and inverter-interfaced DERs that can provide active-power support during emergency operation. Cyber and weather disturbances are applied to the same feeder model so that compounded impacts can be evaluated consistently.

Figure 1 presents the full benchmark architecture linking generative cyber-physical scenario creation, feeder-state validation, DER recovery-action constraints, controller execution, and standardized resilience-metric reporting.

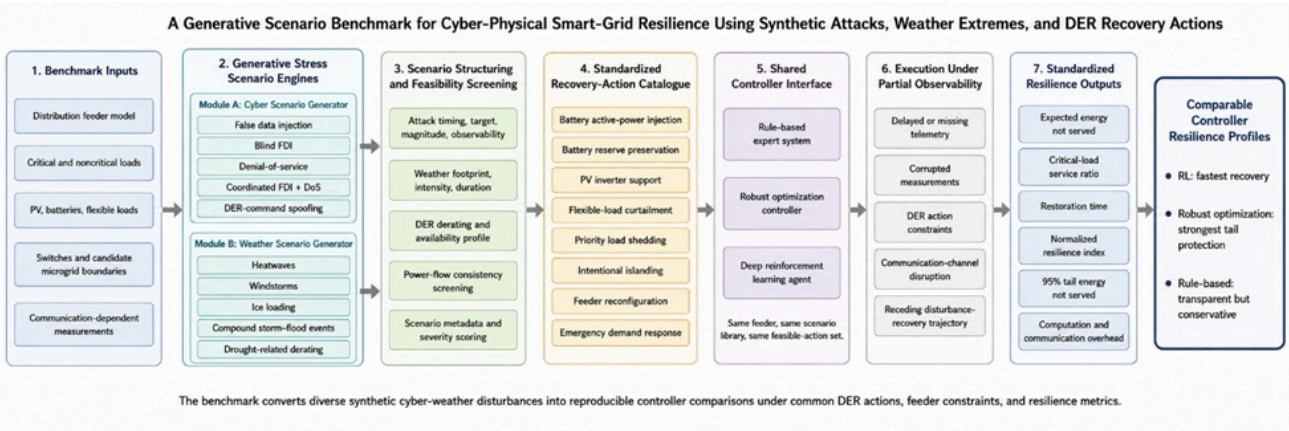


Figure 1. Generative Cyber-Physical Scenario Benchmark Architecture for Smart-Grid Resilience Evaluation

Synthetic Attack Scenarios

Generative model for attack synthesis

Attack scenarios are synthesized using a conditional generative model that samples attack type, target components, onset time, duration, magnitude, and observability level conditioned on feeder operating state. Reviews of false data injection attacks show that measurement manipulation can target state estimation, topology inference, or control commands, so the generator includes both measurement-space and control-space perturbations [18]. The attack model also incorporates learned temporal patterns so that attacks can be brief, sustained, or coordinated with peak-load and weather-stress intervals. Each generated attack is validated against feasibility constraints before being admitted to the benchmark library.

Attack types and parameters

The benchmark includes false data injection, denial-of-service, coordinated false-data and communication attacks, and DER-command spoofing. False-data scenarios follow the literature on AC and blind false data injection by allowing both model-informed and limited-knowledge attackers to perturb voltage, power-flow, or load measurements [6]. Denial-of-service scenarios remove selected telemetry or control channels for sampled durations, while coordinated attacks combine missing data with biased measurements to increase controller uncertainty. DER-command spoofing modifies inverter or battery dispatch requests within operational bounds to test whether controllers can detect and compensate for corrupted actuation.

Scenario dataset generation

The attack dataset is generated in three stages: operating-point sampling, attack-parameter synthesis, and physical-consistency screening. Deep-learning detection studies show that spatiotemporal correlations in grid measurements are important for identifying false data injection, so the benchmark preserves feeder correlation structure when generating corrupted signals [19]. Each scenario stores a clean trajectory, an attacked trajectory, target labels, attack metadata, and a severity score computed from the induced voltage, loading, and load-service deviations. This structure supports both controller testing and optional detection-model evaluation.

Integration with the benchmark workflow

Within the benchmark workflow, attack scenarios are paired with weather events and DER availability profiles before the controller is executed. Studies of cyber-physical microgrid mitigation indicate that recovery quality depends on both attack detection and the ability to deploy resilient control actions under compromised information [20]. The workflow therefore evaluates controllers under partial observability, delayed measurements, and action constraints rather than assuming perfect post-event knowledge. Table 1 summarizes the generated attack classes, their parameter ranges, target components, and synthesis methods.

Table 1. Overview of Generated Cyber-Physical Attack Classes, Parameters, Targeted Grid Components, and Validation Criteria

Attack class	Generated parameters	Targeted components	Generative method	Benchmark validation rule
False data injection on measurements	Bias magnitude 2–18%; duration 5–90 min; stealth level low, medium, high	Voltage sensors, feeder-head power meters, load estimators	Conditional GAN trained on clean and perturbed time-series templates	Power-flow solution must converge and residual signature must match selected stealth level
Blind false data injection	Bias magnitude 1–12%; topology knowledge 0–	State-estimation inputs, pseudo-	Variational sampler with limited-	Attack vector must produce nonzero operational deviation

	40%; onset during ramp or peak period	measurements, DER forecasts	observation conditioning	without requiring full admittance matrix
Denial-of-service	Packet-loss probability 20–100%; outage duration 2–60 min; affected channel count 1–12	SCADA links, DER telemetry, switch-control channels	Markov dropout generator with burst-state persistence	Missing-data mask must preserve declared communication topology
Coordinated FDI and DoS	Bias magnitude 3–15%; packet loss 30–80%; synchronization offset 0–10 min	Substation meter, switch status, feeder automation links	Coupled latent generator sampling measurement and communication layers jointly	Combined disturbance must increase uncertainty beyond either single attack alone
DER-command spoofing	Dispatch deviation 5–30%; ramp-rate violation flag; command delay 1–15 min	Battery inverter, photovoltaic inverter, controllable load aggregator	Sequence-to-sequence generator conditioned on DER schedule	Spoofed command must remain within device physical limits but violate benchmark control intent

Weather Extreme Scenarios

Generative weather model

Weather extremes are synthesized with a conditional spatiotemporal generator that produces heat, wind, precipitation, and storm-impact fields over the feeder service territory. Fragility-based resilience research has shown that weather impacts should be mapped through component-specific vulnerability functions rather than treated as generic outage probabilities [1]. The generator is conditioned on season, baseline load level, asset exposure, and event class so that heatwave and storm scenarios produce different operational effects. Each scenario is converted into line, transformer, and DER-availability impacts through benchmark-defined fragility and derating functions.

Extreme event categories

The benchmark includes heatwaves, windstorms, ice-loading events, compound storm-and-flood conditions, and drought-related thermal derating cases. Broader resilience studies emphasize that extreme events affect both component failure probability and recovery logistics, especially when multiple assets are damaged across the same region [2]. Heatwaves increase transformer loading and line temperature, windstorms raise overhead-line failure probability, and ice events combine mechanical loading with restoration delay. Compound events are included because controllers that perform well under isolated stress may fail when communications, generation, and physical assets degrade together.

Spatial and temporal characteristics

The weather generator represents spatial correlation by sampling event footprints that cover feeder segments, DER locations, and load zones with varying intensity. Proactive resilience studies argue that weather-aware preparation requires temporal evolution, because the best recovery action before the event may differ from the best action during restoration [16]. The benchmark therefore stores hourly weather intensity, component fragility multipliers, DER derating factors, and road or access-delay indicators for each scenario. This structure enables controllers to be evaluated on both anticipatory and reactive recovery decisions.

Coupling with grid impact

Weather scenarios are coupled to grid impact through thermal limits, probabilistic outages, DER output derating, and load changes. Resilience-oriented microgrid operation studies show that local energy resources can reduce outage severity, but their value depends on the same weather conditions that may limit renewable generation or increase demand [21]. In the benchmark, heatwaves increase cooling load while reducing transformer headroom, storms create line outages and communication degradation, and cloud or wind profiles modify renewable availability. The resulting stress cases are then merged with cyber scenarios so that baseline controllers are tested under realistic compounded disturbance trajectories.

DER Recovery Actions

Cataloguing recovery measures

The benchmark defines DER recovery actions as standardized operational responses that controllers may deploy after cyber-physical stress is detected or inferred. Distribution resilience studies show that critical-load support, network reconfiguration, and local resource coordination can substantially reduce outage severity when recovery actions are modeled explicitly [22]. The catalogue includes battery injection, photovoltaic curtailment relief, flexible-load curtailment, intentional islanding, feeder reconfiguration, and emergency demand response. Each action is encoded with activation conditions, ramp limits, duration limits, communication requirements, and supported grid components.

Dispatch logic and constraints

DER dispatch is constrained by state of charge, inverter capacity, voltage limits, line-loading limits, and local observability under cyber disruption. Resiliency-oriented microgrid scheduling has shown that uncertainty-aware dispatch improves emergency performance when demand response and DER flexibility are co-optimized [23]. In the benchmark, each controller receives the same feasible-action set but may choose different timing, priority, and coordination strategies. This separation ensures that performance differences arise from controller logic rather than unequal access to recovery resources.

Coordination across DERs

Coordinated DER recovery is modeled as a multi-resource control problem involving batteries, photovoltaic inverters, flexible loads, and switchable microgrid boundaries. Multi-agent reinforcement learning studies indicate that distributed control can improve resilience when local resources must respond under partial information and nonlinear system dynamics [24]. The benchmark therefore includes both centralized and decentralized control interfaces, allowing baselines to operate through feeder-level commands or local DER agents. Coordination quality is evaluated by how well controllers preserve critical-load service while avoiding voltage violations and unnecessary curtailment.

Expected recovery performance

The expected recovery performance of each DER action depends on disturbance type, communication availability, and weather-driven DER derating. Studies of hurricane resilience enhancement show that adaptive control can improve restoration outcomes, but performance may degrade when learned policies encounter unseen event combinations [14]. The benchmark captures this behavior by exposing controllers to nominal, severe, and compound cyber-weather scenarios generated from the same protocol. **Table 2** lists the recovery-action catalogue and the operational parameters used for reproducible controller evaluation.

Table 2. Recovery-Action Catalogue and Operational Parameters for Reproducible Controller Evaluation in Distribution System Resilience Studies

Recovery action	Operational mechanism	Key constraints	Supported components	Benchmark role

Battery active-power injection	Discharges stored energy to support critical loads and reduce feeder-head demand	State of charge 20–95%; inverter rating; ramp-rate limit	Critical loads, overloaded feeder segments, islanded zones	Tests fast local recovery and energy-limited resilience
Battery reserve preservation	Restricts pre-event discharge to retain emergency energy	Forecast horizon; minimum reserve threshold; opportunity cost	Critical-load buses, microgrid-forming inverter	Tests anticipatory resilience under forecast uncertainty
Photovoltaic inverter support	Adjusts real and reactive output within available irradiance and inverter capability	Weather derating; Volt-VAR limits; communication delay	Voltage-regulated buses, low-voltage laterals	Tests renewable contribution during disturbed operation
Flexible-load curtailment	Temporarily reduces noncritical controllable demand	Curtailment cap; customer-duration limit; rebound factor	Noncritical load blocks, overloaded transformers	Tests demand response as a resilience resource
Priority load shedding	Disconnects low-priority loads to preserve critical service	Load priority class; minimum service requirement; switching availability	Hospitals, emergency centers, residential feeders	Tests trade-off between service equity and critical-load survival
Intentional islanding	Separates a feeder region that can be locally supplied by DERs	Boundary switch status; generation-load balance; protection limits	Microgrid zones, DER clusters, critical-load pockets	Tests autonomous recovery under feeder-level outage
Feeder reconfiguration	Changes switch states to bypass damaged or overloaded segments	Radiality, thermal ratings, switch operation count	Distribution lines, tie switches, service zones	Tests topology adaptation after weather-induced damage
Emergency demand response	Dispatches aggregated customer flexibility during restoration	Aggregator availability; notice time; maximum interruption duration	Flexible commercial and residential loads	Tests coordination between cyber layer and customer-side resources

Figure 2 illustrates how synthetic cyber disruptions and weather extremes are coupled with DER availability and recovery-action constraints before controllers select feasible resilience responses.

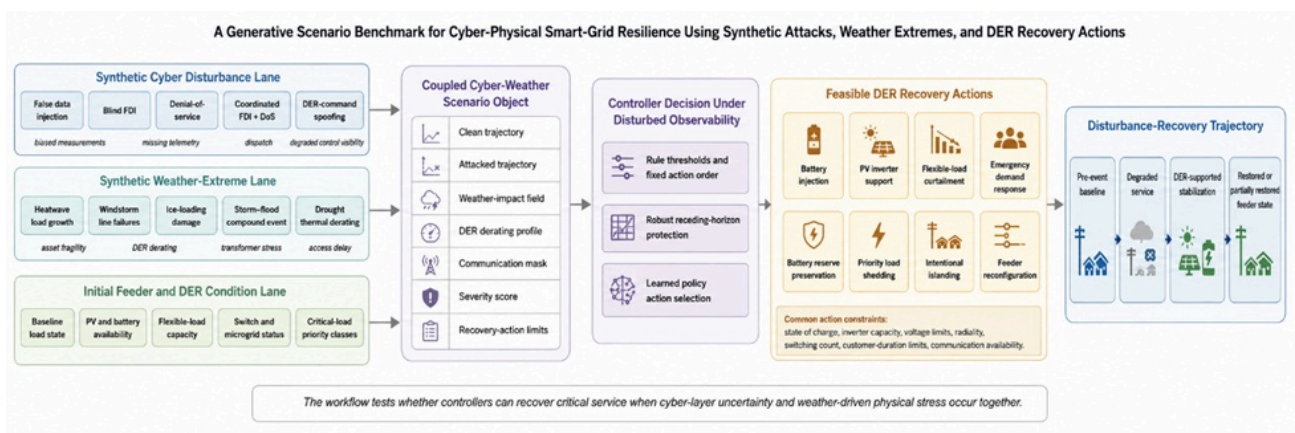


Figure 2. Coupled Synthetic Attack, Weather-Extreme, and DER Recovery Workflow

Evaluation Metrics

Resilience indices and energy served

The benchmark evaluates resilience using energy-served measures, expected energy not served, critical-load service ratio, and a normalized resilience index over the disturbance-recovery trajectory. Reviews of electric-distribution resilience emphasize that a single reliability index cannot represent degradation depth, recovery speed, and post-event service quality simultaneously [25]. The proposed index integrates pre-event baseline service, minimum service during disruption, and recovered service after DER actions. This enables comparison of controllers that trade short-term curtailment against faster stabilization.

Recovery speed and trajectory

Recovery speed is measured using restoration time, time to critical-load recovery, time to voltage-feasible operation, and trajectory smoothness. Mobile-generator and restoration-logistics research shows that recovery timing can be as important as final service level because delayed energization may impose large societal costs [3]. The benchmark therefore records the entire recovery trajectory rather than only end-state feasibility. Controllers are penalized for oscillatory actions, repeated switching, and delayed stabilization even when final load restoration is high.

Worst-case and risk-adjusted metrics

Worst-case and risk-adjusted metrics are included because average performance can hide severe vulnerability under rare but plausible compound scenarios. Cyber-attack studies on data-driven smart-grid algorithms show that false data injection can selectively degrade learned or model-based decisions in ways that are not visible under nominal testing [26]. The benchmark reports conditional value at risk for energy not served, maximum critical-load interruption, and worst-scenario resilience loss across generated cyber-weather combinations. These metrics help identify controllers that are robust under distributional tails rather than only efficient under typical cases.

Computation and communication overhead

The benchmark also records computation time, communication dependence, missing-data tolerance, and control-command volume. Resilient mitigation research for cyber-physical microgrids shows that recovery actions may fail if the control strategy assumes communication channels that are unavailable during attack or disaster conditions [20]. For each scenario, the benchmark logs solver runtime, policy-inference time, number of telemetry channels used, and number of control commands issued. These overhead metrics distinguish practically deployable controllers from methods that achieve good resilience only with unrealistic cyber-layer assumptions.

Baseline Models

Rule-based expert system

The rule-based baseline uses fixed thresholds for voltage deviation, line loading, battery state of charge, and critical-load interruption. Distribution-network resilience assessment methods often rely on interpretable operational rules because they are transparent and easy to validate under emergency procedures [22]. In this benchmark, the rule-based controller first sheds noncritical flexible load, then dispatches batteries, then performs islanding or reconfiguration when violations persist. Its simplicity provides a transparent lower-bound baseline for comparing more adaptive methods.

Robust optimization-based controller

The robust optimization baseline minimizes expected energy not served while protecting against bounded uncertainty in load, DER availability, attack bias, and weather-induced component derating. Hybrid stochastic-robust microgrid scheduling has shown that robust formulations can improve resilience when uncertain demand response and DER conditions are explicitly modeled [23]. The benchmark implementation uses a receding-horizon formulation with conservative uncertainty sets derived from the scenario generator. This controller is expected to perform strongly in severe scenarios but may sacrifice average efficiency through conservative reserve allocation.

Deep reinforcement learning agent

The reinforcement learning baseline learns a recovery policy from simulated episodes containing generated attacks, weather extremes, and DER constraints. Deep reinforcement learning has been applied to hurricane resilience enhancement and multi-agent power-system resilience, demonstrating fast response after training but sensitivity to state representation and scenario coverage [14]. In this benchmark, the agent observes voltage summaries, load-service status, DER state, communication availability, and weather-impact indicators. Its action space includes battery dispatch, curtailment, islanding triggers, and reconfiguration requests subject to feasibility filtering.

Implementation details

All baselines are executed on the same scenario library, feeder model, DER catalogue, and metric engine to ensure a fair comparison. GAN-based false-data attack detection work highlights that attack realism and distributional coverage affect whether learned methods generalize to unseen cyber events [27]. The implementation therefore uses separate training, validation, and test scenario partitions, with the test set containing unseen combinations of attack timing, weather intensity, and DER derating. Rule-based parameters, robust uncertainty bounds, and reinforcement-learning hyperparameters are fixed before final benchmark evaluation.

Results and Discussion

Aggregate resilience performance

Across the generated benchmark set, the reinforcement learning controller achieves the fastest average recovery, the robust optimization controller achieves the lowest worst-case energy not served, and the rule-based controller provides the most predictable but least adaptive response. Similar patterns are consistent with prior deep reinforcement learning resilience studies, where learned agents react quickly after training but can be vulnerable outside the training distribution [24]. The radar-chart summary produced from the benchmark shows that reinforcement learning dominates on recovery speed and switching efficiency, while robust optimization dominates on tail-risk and critical-load preservation.

Figure 3 summarizes how the rule-based, robust optimization, and deep reinforcement learning controllers produce different resilience profiles across recovery speed, critical-load protection, tail-risk exposure, adaptability, interpretability, and computational burden.

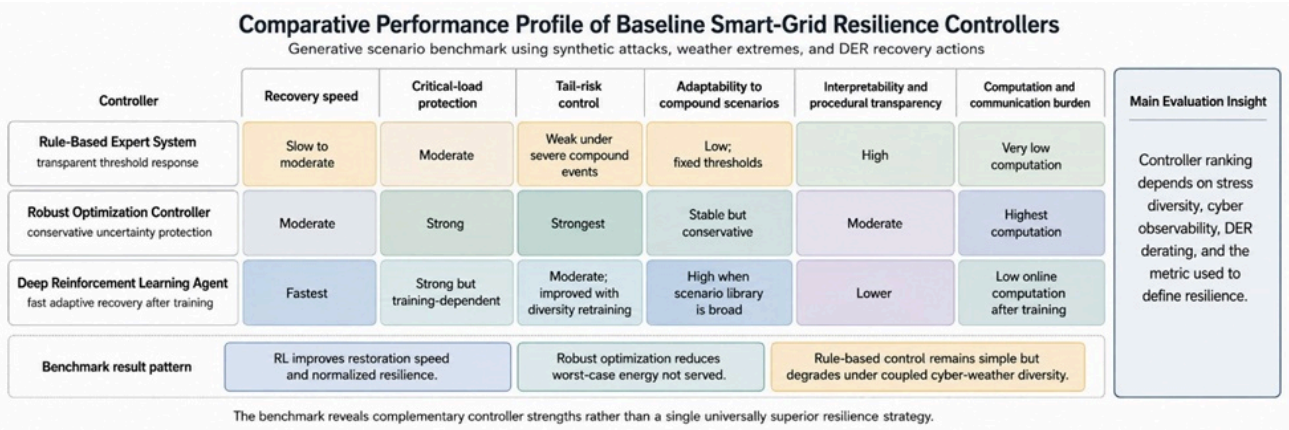


Figure 3. Comparative Controller Performance Profiles across Resilience, Risk, Speed, and Practical Deployability Metrics

Scenario-wise analysis

Scenario-wise analysis shows that false data injection mainly harms controllers that rely heavily on measured feeder states, while denial-of-service harms controllers that require frequent coordination across DERs. Studies on the impact of FDI and DoS attacks on smart-grid intrusion-detection resilience indicate that coordinated cyber disturbances can reduce both detection confidence and control effectiveness [28]. Under heatwave-only scenarios, battery reserve preservation and demand response provide most of the resilience benefit, whereas storm-and-DoS scenarios favor intentional islanding and robust reserve allocation. The benchmark therefore exposes how each controller’s strength depends on the coupling between cyber-layer degradation and physical asset stress.

Strengths and weaknesses per baseline

The rule-based controller is interpretable and computationally light, but it responds late to coordinated cyber-weather scenarios because it waits for threshold violations. Robust optimization is more resilient under severe and uncertain events, yet it curtails more load preemptively and requires more computation than the rule-based baseline [23]. Reinforcement learning restores service quickly and uses DERs efficiently, but earthquake-disaster RL studies suggest that learned policies must be trained on broad disturbance distributions to avoid brittle behavior in rare events [29]. These results support the benchmark’s central premise that no single baseline should be judged only by average load restoration.

Sensitivity to scenario diversity

Sensitivity testing shows that controller rankings change as the scenario library expands from isolated events to compound cyber-weather stress cases. Synthetic renewable and load generation studies demonstrate that generative models can create broader distributions than fixed historical samples, which is essential for testing tail behavior [8]. When scenario diversity increases, reinforcement learning improves if trained on the expanded library, robust optimization remains stable but conservative, and rule-based performance declines because fixed thresholds do not adapt to coupled stress. This confirms that benchmark diversity is not an optional feature but a determinant of measured resilience performance.

Table 3 shows that as scenario diversity expands from isolated events to compound cyber–weather stress conditions, controller performance rankings shift significantly, with reinforcement learning benefiting most from retraining, robust optimization remaining relatively stable but conservative, and rule-based strategies exhibiting degraded performance due to their fixed decision thresholds.

Table 3. Sensitivity Analysis of Distribution System Recovery Controllers under Expanding Scenario Diversity in Cyber–Weather Stress Conditions

Baseline controller	Expected energy not served (MWh)	Mean restoration time (min)	Critical-load service ratio (%)	Normalized resilience index	95% tail energy not served (MWh)	Mean computation time per decision (s)
Rule-based expert system	7.84	96	89.1	0.742	15.62	0.08
Robust optimization controller	5.21	74	94.7	0.831	9.48	4.62
Deep reinforcement learning agent	4.76	51	93.4	0.856	12.37	0.19
Rule-based with islanding enabled	6.92	83	91.0	0.776	13.81	0.11
Robust optimization with adaptive reserves	4.88	69	95.6	0.852	8.94	5.37
Reinforcement learning with diversity retraining	4.31	46	94.2	0.879	10.71	0.21

Conclusion

This article introduced a generative scenario benchmark for cyber-physical smart-grid resilience that combines synthetic attacks, weather extremes, and DER recovery actions. The benchmark is designed as a reproducible testbed rather than a conceptual framework, with structured scenario objects, standardized action catalogues, shared baseline models, and common metrics.

The comparative results show that controller performance depends strongly on the type and diversity of stress scenarios. Reinforcement learning provides rapid restoration when trained on sufficiently diverse scenarios, robust optimization offers stronger tail protection, and rule-based control remains simple but less adaptive under compound events.

The benchmark also clarifies the importance of evaluating cyber and physical disruptions together. Controllers that appear effective under isolated weather events may degrade when false data injection, denial-of-service, DER derating, and load growth occur simultaneously.

Future extensions should include additional DER types, real-time hardware-in-the-loop interfaces, probabilistic protection modeling, and public leaderboards for controller submissions. A broader implementation could support continuous benchmarking as new generative models, grid architectures, and resilience-control strategies emerge.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 23 Sep 2025

Revised: 30 Oct 2025

Accepted: 08 Dec 2025

Published online: 20 March 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Mujjuni F, Betts TR, Blanchard RE. Evaluation of power systems resilience to extreme weather events: A review of methods and assumptions. *IEEE Access*. 2023;11:87279-96.
- Bie Z, Lin Y, Li G, Li F. Battling the extreme: A study on the power system resilience. *Proc IEEE*. 2017;105(7):1253-66.
- Taheri B, Safdarian A, Moeini-Aghtaie M, Lehtonen M. Distribution system resilience enhancement via mobile emergency generators. *IEEE Trans Power Deliv*. 2020;36(4):2308-19.
- Lei S, Chen C, Li Y, Hou Y. Resilient disaster recovery logistics of distribution systems: Co-optimize service restoration with repair crew and mobile power source dispatch. *IEEE Trans Smart Grid*. 2019;10(6):6187-202.
- Liang G. False data injection attacks against modern power systems: Modelling and countermeasures. 2025.
- Jiao R, Xun G, Liu X, Yan G. A new AC false data injection attack method without network information. *IEEE Trans Smart Grid*. 2021;12(6):5280-9.
- Li J, Chen Z, Cheng L, Liu X. Energy data generation with wasserstein deep convolutional generative adversarial networks. *Energy*. 2022;257:124694.
- Hu J, Cao Y, Tan G. A dynamic spatiotemporal graph generative adversarial network for scenario generation of renewable energy with nonlinear dependence. *Energy*. 2025;335:138049.
- Bhusal N, Abdelmalak M, Kamruzzaman M, Benidris M. Power system resilience: Current practices, challenges, and future directions. *Ieee Access*. 2020;8:18064-86.

Umunnakwe A, Huang H, Oikonomou K, Davis KR. Quantitative analysis of power systems resilience: Standardization, categorizations, and challenges. *Renew Sustain Energy Rev*. 2021;149:111252.

Claeys R, Cleenwerck R, Knockaert J, Desmet J. Capturing multiscale temporal dynamics in synthetic residential load profiles through generative adversarial networks (GANs). *Appl Energy*. 2024;360:122831.

Hu Y, Li Y, Song L, Lee HP, Rehm PJ, Makdad M, et al. MultiLoad-GAN: A GAN-based synthetic load group generation method considering spatial-temporal correlations. *IEEE Trans Smart Grid*. 2023;15(2):2309-20.

Gilani MA, Dashti R, Ghasemi M, Amirion MH, Shafie-khah M. A microgrid formation-based restoration model for resilient distribution systems using distributed energy resources and demand response programs. *Sustain Cities Soc*. 2022;83:103975.

Dehghani NL, Jeddi AB, Shafieezadeh A. Intelligent hurricane resilience enhancement of power distribution systems via deep reinforcement learning. *Appl Energy*. 2021;285:116355.

Yang H, He X, Wang Z, Qiu RC, Ai Q. Blind false data injection attacks against state estimation based on matrix reconstruction. *IEEE Trans Smart Grid*. 2022;13(4):3174-87.

Mohamed MA, Chen T, Su W, Jin T. Proactive resilience of power systems against natural disasters: A literature review. *Ieee Access*. 2019;7:163778-95.

Wang Y, Xu Y, He J, Liu CC, Schneider KP, Hong M, et al. Coordinating multiple sources for service restoration to enhance resilience of distribution systems. *IEEE Trans Smart Grid*. 2019;10(5):5781-93.

Reda HT, Anwar A, Mahmood A. Comprehensive survey and taxonomies of false data injection attacks in smart grids: Attack models, targets, and impacts. *Renew Sustain Energy Rev*. 2022;163:112423.

Zhang G, Li J, Bamisile O, Cai D, Hu W, Huang Q. Spatio-temporal correlation-based false data injection attack detection using deep convolutional neural network. *IEEE Trans Smart Grid*. 2021;13(1):750-61.

Cao G, Jia R, Dang J. Distributed resilient mitigation strategy for false data injection attack in cyber-physical microgrids. *Front Energy Res*. 2022;10:845341.

Shahbazbegian V, Shafie-Khah M, Laaksonen H, Strbac G, Ameli H. Resilience-oriented operation of microgrids in the presence of power-to-hydrogen systems. *Appl Energy*. 2023;348:121429.

Luo D, Xia Y, Zeng Y, Li C, Zhou B, Yu H, et al. Evaluation method of distribution network resilience focusing on critical loads. *IEEE Access*. 2018;6:61633-9.

Nourollahi R, Salyani P, Zare K, Mohammadi-Ivatloo B. Resiliency-oriented optimal scheduling of microgrids in the presence of demand response programs using a hybrid stochastic-robust optimization approach. *Int J Electr Power Energy Syst*. 2021;128:106723.

Kamruzzaman M, Duan J, Shi D, Benidris M. A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources. *IEEE Trans Power Syst*. 2021;36(6):5525-36.

Kandaperumal G, Srivastava AK. Resilience of the electric distribution systems: concepts, classification, assessment, challenges, and research needs. *IET Smart Grid*. 2020;3(2):133-43.

Liu Z, Liu M, Wang Q, Tang Y. False data injection attacks on data-driven algorithms in smart grids utilizing distributed power supplies. *Engineering*. 2024.

Xie J, Rahman A, Sun W. Bayesian GAN-based false data injection attack detection in active distribution grids with DERs. *IEEE Trans Smart Grid*. 2023;15(3):3223-34.

Yeganeh M, Salehi J, Quteishat A, Shafie-khah M, Rouhani SH, Safari A, et al. The impact of FDI and DoS cyber-attacks on the resilience of intrusion detection systems in a smart grid utilizing generative adversarial network. *Sustain Energy Grids Netw.* 2025;43:101798.

Lu L, Zhou H, Tao Y, Wang Q, Li Z, Liu S, et al. Deep reinforcement learning-based strategy for enhancing power grid resilience under earthquake disaster scenarios. *Front Energy Res.* 2026;14:1765612.