D3.2. Risk-based optimisation of preventive and corrective actions using data-driven optimisation

This deliverable is based on the publication:

Khaloie, H., Yurtseven, K., Çalık, H., Cao, J., Toubeau, J.-F., Vallée, F., & Ergun, H. (2026). Stage-Wise Dataset Generation for Two-Stage Security-Constrained AC Optimal Power Flow, 26th International Conference on Environment and Electrical Engineering, 29 June – 02 July, 2026, Lisbon, Accepted for publication

Table of Contents

1. Purpose within the project

This paper contributes a dataset-generation layer for learning-assisted security-constrained operation. The starting point is that two-stage security-constrained AC optimal power flow represents a realistic operational sequence: day-ahead decisions define preventive dispatch and reserve positions, while real-time decisions apply corrective redispatch, curtailment, load shedding, and post-contingency operating points after renewable deviations and N−1 outages are realised. Solving this full two-stage stochastic AC-SCOPF repeatedly is computationally expensive because the nonlinear AC constraints and the non-anticipativity of day-ahead decisions couple all scenarios in one large nonconvex optimisation problem.

The motivation is therefore not to replace SCOPF by a simplified deterministic approximation, but to create high-fidelity synthetic datasets that preserve the sequential structure of the actual decision problem. Such datasets are essential for downstream machine-learning models that aim to support or accelerate operator decision-making. Common dataset-generation practices based on repeated deterministic OPF or SCOPF solves can generate many samples, but they do not capture the inter-scenario coupling that defines a two-stage stochastic problem. This can lead to first-stage decisions that are not representative of the preventive decisions an operator would make when anticipating a broad set of real-time uncertainty and contingency outcomes.

Aspect	Summary
Main method	Stage-wise decomposition for tractable two-stage stochastic AC-SCOPF dataset generation.
Operational structure	Day-ahead preventive scheduling followed by real-time corrective redispatch under wind deviations and N−1 line outages.
Compared algorithms	Two-Phase Sampling and Bundle-and-Fix Sampling.
Main result	Bundle-and-Fix preserves the two-stage cost behaviour with much higher fidelity while enabling parallel generation of independent real-time recourse samples.

2. Methodological contribution

The paper formulates a single-period corrective SCOPF as a two-stage stochastic program. The first-stage decision vector contains day-ahead variables that are shared by all scenarios, including scheduled conventional generation, storage operation, reserve procurement, base-case voltages, and base-case line flows. The second-stage decision vector contains scenario-dependent real-time variables, including generator and storage regulation, wind curtailment, load shedding, post-contingency voltages, and post-contingency line flows. Each scenario combines a wind-power deviation with a post-outage network topology. The objective minimises day-ahead operating and reserve costs plus the expected cost of real-time balancing, curtailment, and load shedding.
The key methodological contribution is the stage-wise decomposition of dataset generation. Instead of repeatedly solving the complete scenario-coupled two-stage model for every desired training sample, the framework separates the generation process into a first-stage decision acquisition phase and a second-stage recourse sampling phase. Once a representative day-ahead decision has been obtained, the real-time recourse problems can be solved independently for many uncertainty and contingency realisations. This makes the dataset-generation task naturally parallelisable and avoids the memory and computational burden of repeatedly enlarging the monolithic two-stage SCOPF.

Two algorithms are compared. The first, Two-Phase Sampling, repeatedly solves deterministic single-scenario realisations and stores the resulting day-ahead decisions. These decisions are then paired with out-of-sample real-time scenarios to generate recourse samples. This approach is computationally light and resembles common deterministic sampling practice, but it does not preserve the full stochastic coupling of first-stage decisions. The second, Bundle-and-Fix Sampling, solves the original two-stage stochastic model once over a scenario bundle to obtain a single non-anticipative day-ahead decision. This decision is then fixed while many independent real-time recourse problems are solved. The bundle size becomes the main lever controlling the trade-off between fidelity and computational effort.

3. Case study results

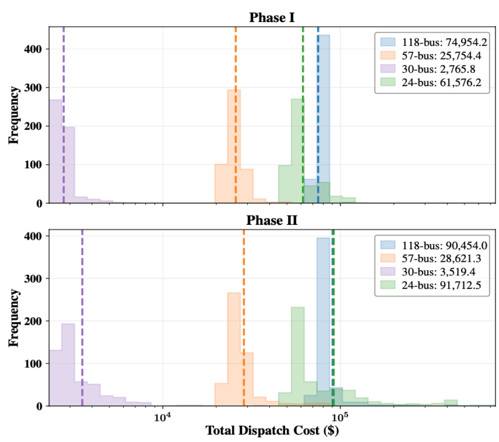
The numerical study evaluates the two algorithms on four PGLib benchmark systems: 24-bus, 30-bus, 57-bus, and 118-bus networks. Each system is augmented with storage and wind resources to create a two-stage corrective SCOPF test bed. Real-time uncertainty is represented by wind deviations and N−1 transmission line outages. For a fair comparison, both algorithms are evaluated using 500 out-of-sample real-time scenarios in the second phase.

The main trend observed for Two-Phase Sampling is a clear deterioration when moving from the deterministic Phase-I solutions to the out-of-sample Phase-II evaluation. Although the method is computationally light, the cost distributions shift upward and develop heavier right tails once the recorded day-ahead decisions are exposed to new wind deviations and contingency realisations. This shows that first-stage decisions obtained from isolated deterministic scenarios are not sufficiently robust or representative of the underlying two-stage stochastic problem. In practical terms, the method can generate diverse samples, but it risks producing preventive schedules that look attractive for the scenario used to create them and perform poorly when evaluated under different real-time conditions.

Bundle-and-Fix Sampling shows a substantially more stable pattern. The Phase-I and Phase-II cost distributions remain closely aligned, and the pronounced right-tail expansion seen with Two-Phase Sampling is largely avoided. This indicates that solving the two-stage model once over a scenario bundle produces a day-ahead decision that better internalises the range of possible real-time outcomes. When this decision is fixed and used to generate independent recourse samples, the resulting dataset preserves the dominant cost behaviour of the original stochastic SCOPF much more faithfully. The comparison with the ground-truth two-stage SCOPF confirms this trend: Bundle-and-Fix does not exactly reproduce every tail outcome, but it captures the main distributional behaviour with significantly higher fidelity than deterministic sampling.

The sensitivity analysis highlights the role of the Phase-I bundle size as the main fidelity–tractability tuning parameter. Increasing the number of bundled scenarios generally reduces the gap to the ground-truth two-stage SCOPF, because the preventive decision is exposed to a richer set of uncertainty and contingency outcomes during its construction. However, the required bundle size is system-dependent. Smaller and medium-sized systems reach good agreement with moderate bundle sizes, whereas the 118-bus system needs a larger bundle to achieve a comparably low gap. This trend suggests that larger networks, with more complex interactions between contingencies, reserves, storage, and wind deviations, require a broader scenario representation to obtain a sufficiently representative day-ahead decision.

The computational assessment further supports the practical value of the decomposition. The measured average solution times per generated sample remain suitable for large-scale dataset generation, and the independent recourse problems can be distributed across parallel workers. For a projected 10,000-sample dataset using 16 parallel workers, Algorithm 2 requires approximately 1.56 h, 1.36 h, 2.64 h, and 7.00 h for the 24-, 30-, 57-, and 118-bus systems, respectively. This compares favourably with repeatedly solving monolithic 500-scenario two-stage SCOPF benchmarks, especially for the larger systems where memory and scenario-coupling limitations become more severe.



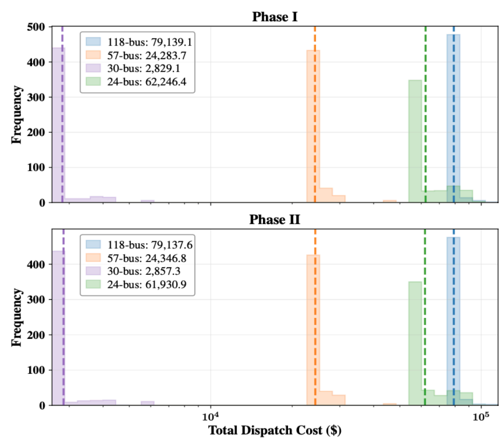


Figure 1 – Left: Total dispatch cost distributions from Algorithm 1 in Phase I and under Phase-II out-of-sample RT evaluation across the four test systems, Right: Total dispatch cost distributions from Algorithm 2 in Phase I and under Phase-II out-of-sample RT evaluation across the four test systems.

4. Expanded interpretation of the complete result set

The paper demonstrates that high-fidelity dataset generation for learning-assisted SCOPF requires more than simply sampling operating points and solving deterministic OPF or SCOPF instances. The key issue is the representativeness of the day-ahead decision. In a two-stage stochastic problem, the preventive schedule must be chosen before the uncertainty and contingency are known, and it must therefore reflect the expected consequences of many possible real-time states. When first-stage decisions are obtained from isolated deterministic realisations, as in Two-Phase Sampling, they may look feasible and economical for the sampled realisation but fail to generalise to out-of-sample recourse conditions. This explains the cost increase and heavier right tails observed in the Phase-II evaluation.

Bundle-and-Fix addresses this weakness by computing the day-ahead decision from a scenario bundle before generating independent real-time recourse samples. This preserves the central non-anticipativity structure of two-stage stochastic programming while keeping the large-scale dataset-generation phase tractable. The resulting dataset records contain the fixed day-ahead decision, the realised wind deviation and outage topology, the optimal real-time corrective actions, and feasibility or cost indicators. Such records are well suited for supervised learning tasks in which a model predicts recourse actions, security indicators, or cost quantities conditioned on a preventive operating point and realised system state.

The comparison with the ground-truth two-stage SCOPF is particularly important. It shows that the proposed decomposition does not merely accelerate data generation; it also preserves the relevant cost behaviour of the original stochastic model. The remaining deviations are mainly associated with the size and representativeness of the Phase-I scenario bundle. This provides a practical design rule: smaller systems or less diverse uncertainty sets may require only moderate bundle sizes, while larger systems with more complex network interactions require larger bundles to obtain a representative preventive decision.

For the broader project, the paper provides a complementary layer to optimisation-based decision-support work. Previous deliverables focus on solving stochastic and risk-aware AC/DC grid operation problems directly. This paper focuses on how to generate credible datasets for machine-learning models that could approximate or accelerate such sequential decisions. Its contribution is therefore methodological and enabling: it establishes how to construct datasets that respect the timing and coupling of preventive and corrective SCOPF decisions, while remaining computationally feasible. This is a necessary step if learning-based tools are to be trained on data that reflect the actual structure of secure power-system operation rather than on simplified deterministic samples.