# A Dataset Generation Toolbox for Dynamic Security Assessment: On the Role of the Security Boundary

Bastien Giraud*, Lola Charles, Agnes Marjorie Nakiganda, Johanna Vorwerk, and Spyros Chatzivasileiadis
*Department of Wind and Energy Systems*
*Technical University of Denmark*
Lyngby, Denmark
*bagir@dtu.dk

***Abstract*—This paper introduces an efficient dataset generation toolbox, aiming to substantially improve data-driven dynamic security assessment methods (DSA). As conventional DSA methods are becoming intractable due to the increasing power system complexity, data-driven DSA methods are gaining significant interest. However, computationally efficient and high-quality dataset generation, which is the cornerstone for these methods' performance, remains a major challenge. Despite efforts with generic or importance sampling techniques to focus on operating points near the system's security boundary, systematic methods for sampling in this region remain scarce. This paper presents two main contributions. First, it introduces an open-source dataset generation toolbox to efficiently generate balanced datasets along the security boundary of the system. Second, considering the lack of systematic evidence on the impact of sampling near the security boundary, this paper performs a comprehensive assessment of how accurately capturing the security boundary affects the performance of data-driven DSA methods. In this paper, we consider AC steady-state feasibility and small-signal stability. Our case studies on the PGLib-OPF 39-bus and 162-bus systems demonstrate the importance of (i) including boundary-adjacent operating points in training datasets, and (ii) maintaining a balanced distribution of secure and insecure points.**

***Index Terms*—Security boundary, data generation, dynamic security assessment, machine learning, power system operation**

## 1 Introduction

The green energy transition and the broad deployment of converter-interfaced resources introduce faster and more complex dynamics in power systems worldwide that challenge power system reliability, security, and stability. To address the growing uncertainty in power generation due to the integration of intermittent renewable energy sources [1], one potential solution is to evaluate all possible operating points (OPs) and severe contingencies offline. However, assessing these OPs against multiple contingencies through time-consuming simulations and high-fidelity system modeling [2] becomes intractable, as future power systems incorporate more distributed generation and variable prosumer behavior [3].

Data-driven approaches for dynamic security assessment (DSA), introduced in early works such as [4], are emerging as an alternative to conventional methods and have been gaining traction in recent years [5, 6]. DSA generally measures the ability of a power system to withstand a pre-defined set of contingencies by reaching a stable post-fault state. Data-driven techniques aim to address the challenges of conventional DSA by leveraging offline simulations and machine learning (ML) to create efficient proxies for both offline and online decision-making [5]. For example, the authors in [7] train a decision tree (DT) to derive operating rules for voltage collapse issues on the French transmission grid, classifying OPs based on whether they meet the required voltage margin. Similarly, [8] trains a DT for transient stability analysis, differentiating between OPs where the rotor angle difference remains below $180°$ (stable) or exceeds $180°$ (unstable) 10 seconds after fault clearance. In general, these methods classify OPs as *secure* or *insecure*, depending on the specific stability criterion under consideration. However, the performance of such ML tools heavily relies on the quality of the training dataset [8].

One way to acquire datasets is by utilizing historical data. The proliferation of phasor measurement units and advanced monitoring tools in control centers has significantly increased data availability [9]. However, since power systems predominantly operate in stable conditions, most recorded OPs reflect stable states. As a result, such datasets are often biased and lack representation of rare but critical OPs, creating gaps in classification accuracy for a security assessment [5]. To overcome this limitation, synthetic data generation is vital in constructing balanced datasets that include both secure and insecure OPs.

State-of-the-art methods for generating synthetic datasets for DSA are broadly categorized into three approaches: (1) historical sampling, (2) importance sampling, and (3) generic sampling. Historical sampling methods, such as the vine-copula-based approach in [10] and the composite modeling

The work was supported by the ERC Starting Grant VeriPhIED, funded by the European Research Council, Grant Agreement 949899 and the AI-EFFECT project, funded by the European Union's Horizon Europe Research and Innovation Program, Grant Agreement 101172952.

approach in [11], leverage historical data to account for the dependent structure of variables. However, they are limited in their ability to generalize to OPs outside the historical records. In contrast, importance sampling methods [7, 12, 13] aim to enhance the representation of rare events by introducing a bias in the sampling process, thereby concentrating more OPs near the security boundary. However, the input space is typically unbalanced, since the insecure region is significantly larger than the secure region. Thus, despite the advantages that importance sampling entails, obtaining a sufficient number of secure samples close to the boundary remains challenging as highlighted in [14]. To obtain more secure samples, generic sampling methods, such as those in [15, 16], aim to maximize coverage of the operating space defined by AC power flow equations. But the computational complexity of sampling in high-dimensional spaces poses significant challenges even for high performance computation methods. Although [7, 8, 12, 13] emphasize the importance of sampling near the security boundary to consider information-rich areas, there is still no scientific evidence in the literature confirming that these boundary-adjacent samples actually enhance data-driven DSA methods.

In this work, we address the research gap related to the security boundary by investigating how sampling near it affects the accuracy and efficiency of ML classifiers for DSA. While DSA encompasses various aspects of system stability—such as voltage, transient, and small-signal stability—this work focuses specifically on small-signal stability. Future work will focus on transient and voltage stability. In this paper, we demonstrate for the first time that an ML model trained with a higher proportion of samples near the security boundary can accurately classify OPs far from this boundary, highlighting the importance of precisely characterizing this region. To obtain OPs near the security boundary, we extend the sampling approach from previous work [14, 17, 18]. Specifically, our method addresses the challenge of sampling in high-dimensional spaces by incorporating optimization-based bound tightening (OBBT) techniques [19], which systematically reduce the search space. Further significant reduction is achieved through convex relaxations and separating hyperplanes, as discussed in [17]. Each OP is evaluated for AC feasibility and small-signal stability, while the directed walk algorithm from [14] is used to identify samples closer to the security boundary. The resulting dataset is constructed to contain samples within a user-specified margin around the security boundary, balanced between secure and insecure OPs, and validated for both AC feasibility and small-signal stability. Unlike traditional approaches, which often struggle to efficiently capture samples near the security boundary in high dimensions, our method strategically integrates these techniques to generate a dataset with a high share of boundary-adjacent samples. We then compare our method against naive and importance sampling benchmarks, evaluating performance on a decision tree (DT). Finally, we analyze the misclassified points and assess the need for a balanced dataset. In short, the contributions of the presented work are as follows:

- Building up on our previous work, we make for the first time publicly available an efficient dataset generation toolbox (see Github link in [20]), combining the separating hyperplane [17] and directed walk [14] algorithms. We show that our proposed method effectively finds many samples around the security boundary. Our goal with this toolbox is to enable researchers and industry to generate high quality datasets that can accelerate the development and testing of new data-driven methods. It is built in a modular way with the hope that fellow researchers can further develop it and extend it.
- We show that a DT trained on samples around the security boundary can also accurately classify points lying far from the boundary.
- We demonstrate the importance of accurately capturing the security boundary by analyzing the misclassified samples, and by comprehensively investigating the distribution of samples of our proposed method and two benchmark approaches.

The remainder of this paper is structured as follows. Section 2 provides an introduction to DSA and establishes crucial definitions. Then, Section 3 details our proposed method for finding and sampling in the vicinity of the security boundary. Section 4 analyses the obtained datasets and benchmarks the performance of a DT trained on the dataset obtained from the proposed method against generic and importance sampling methods for two representative test systems. Finally, Section 5 outlines the limitations of the proposed method and explores directions for future work. Section 6 concludes.

## 2 Dynamic Security Assessment

This Section provides an introduction to DSA of power systems. A secure system is one that respects operational limits under pre-contingency, during-contingency, and post-contingency conditions. DSA assesses if the system remains stable after a contingency and transitions to a stable post-contingency steady-state [21]. As such, all aspects of security and stability, including thermal limits, voltage and frequency constraints, and various types of system stability, such as voltage stability, transient stability, and small-signal stability, must be considered. The computations required to assess all these aspects of security for a single operating condition are thus computationally highly demanding. In this work, we focus on one specific stability aspect: small-signal stability assessment. Additionally, we consider the steady-state feasibility of the system, meaning that a stable OP exists for which a power flow solution satisfies the system's operational constraints. Hence, for the remainder of this paper, the term DSA refers to the assessment of steady-state feasibility and small-signal stability of an OP.

To clearly differentiate the terms used throughout the paper, we classify an OP as follows:

- an OP is *feasible* when it does not violate any operational constraints in steady-state, e.g. all voltage magnitudes remain within bounds and thermal limits are met,
- an OP is considered *stable* when it returns back to a stable state following a small-signal disturbance, and finally
- an OP is considered *secure* when it is both steady-state feasible and small-signal stable.

The remainder of this Section outlines how DSA methods assess feasibility and small-signal stability, before formally defining the security boundary.

### 2.1 Feasibility

For an OP to be feasible, a power flow solution must exist that fulfills all physical system constraints, i.e. line flow and voltage magnitude limitations. Finding every single feasible OP for a system is still an open research question. A systematic method, however, that can help us determine feasible OPs, and which we will use in this paper, is through optimization, and specifically the AC Optimal Power Flow (AC-OPF). The AC-OPF tries to determine the most economical operating point that meets all physical system constraints. Hence, all points that satisfy the AC-OPF constraints, i.e. the AC-OPF feasible space, correspond to the entire collection of feasible OPs that we are looking for.

AC-OPF is, however, a non-linear and non-convex optimization problem. To enhance computational efficiency and generate numerous feasible samples, our methodology incorporates a convex relaxation of the AC-OPF, specifically employing the quadratic convex (QC) relaxation. We choose the QC relaxation over other alternatives, such as Semi-Definite Programming and Second-Order Cone relaxations, as it provides a good trade-off between computational complexity and the tightness of the relaxation [22]. However, other convex relaxations could also be applied. A detailed description of the QC and other convex relaxations is provided in [23]. Here, we provide a brief overview of the applied AC-OPF and QC relaxation.

#### 2.1.1 AC Optimal Power Flow

To efficiently generate a dataset of feasible OPs, the AC-OPF is employed to adjust infeasible OPs and identify the closest feasible solutions. Generally, an AC-OPF minimizes a user-defined objective, typically the fuel cost in a power system context, while ensuring compliance with operational constraints. Let us consider a power system with a set of $\mathcal{N}$ buses. While the subset $\mathcal{G}$ contains generator buses, the subset $\mathcal{D}$ contains buses serving a load. The subset $(i,j) \in \mathcal{L}$ contains power lines that connect bus $i$ to bus $j$. The goal of the optimization is to find the complex bus voltage magnitudes $V_k$ for each bus $k \in \mathcal{N}$ and the complex power dispatch $S_G = P_g + jQ_g$ for every generator $k \in \mathcal{G}$ that satisfy the following operational constraints:

$$(V_k^{\min})^2 \leq V_k(V_k)^\star \leq (V_k^{\max})^2 \quad \forall k \in \mathcal{N} \tag{1a}$$
$$S_{G_k}^{\min} \leq S_{G_k} \leq S_{G_k}^{\max} \quad \forall k \in \mathcal{G} \tag{1b}$$
$$|S_{ij}| \leq S_{ij}^{\max} \quad \forall (i,j) \in \mathcal{L} \tag{1c}$$
$$S_{G_k} - S_{D_k} = \sum_{k,j \in \mathcal{L}} S_{kj} \quad \forall k \in \mathcal{N} \tag{1d}$$
$$S_{ij} = (Y_{ij})^\star V_i(V_i)^\star - (Y_{ij})^\star V_i(V_j)^\star \quad \forall (i,j) \in \mathcal{L} \tag{1e}$$
$$\theta_{ij}^{min} \leq (V_i(V_j)^\star) \leq \theta_{ij}^{max} \quad \forall (i,j) \in \mathcal{L} \tag{1f}$$

where (1a) ensures that the bus voltage magnitudes are constrained by their upper and lower limits, (1b) bounds the generator's complex power outputs, (1c) limits the line flows across each transmission line and (1e) formulates the line flow. Furthermore, (1d) enforces that generation and demand are balanced at every node and (1f) constrains the angle difference between the two line ends for each transmission line.

#### 2.1.2 Quadratic Convex (QC) Relaxation

The AC-OPF constraints described in (1) include non-convex terms arising from the product of the voltages (i.e. $V_k(V_k)^\star$), which complicates the optimization problem and motivates the use of convex relaxations to reduce computational complexity. QC relaxations, as proposed in [22], employ convex envelopes in the polar form of the AC-OPF to relax the dependencies among voltage variables. Additionally, an auxiliary matrix variable $W$ is introduced to represent the product of the complex bus voltages, where:

$$W_{ij} = V_i(V_j)^\star. \tag{2}$$

This allows for reformulation of (1a), (1e) and (1f) as follows:

$$(V_k^{\min})^2 \leq (W_{kk}) \leq (V_k^{\max})^2 \quad \forall k \in \mathcal{N} \tag{3a}$$
$$S_{ij} = (Y_{ij})^\star W_{ii} - (Y_{ij})^\star W_{ij} \quad \forall (i,j) \in \mathcal{L} \tag{3b}$$
$$S_{ji} = (Y_{ij})^\star W_{jj} - (Y_{ij})^\star (W_{ij})^\star \quad \forall (i,j) \in \mathcal{L} \tag{3c}$$
$$\tan(\theta_{ij}^{\min}) \leq \frac{\Im(W_{ij})}{\Re(W_{ij})} \leq \tan(\theta_{ij}^{\max}) \quad \forall (i,j) \in \mathcal{L} \tag{3d}$$

The non-convexity is captured by the voltage product in (2). To obtain a convex relaxation, (2) is removed from the optimization, and new variables are introduced for the voltages $v_i \angle \theta_i \quad \forall i \in \mathcal{N}$ and squared current flows $l_{ij} \quad \forall (i,j) \in \mathcal{L}$ are added. Additionally, the following convex constraints and envelopes are incorporated:

$$W_{kk} = \langle v_k^2 \rangle^T \quad \forall k \in \mathcal{N} \tag{4a}$$
$$\Re(W_{kk}) = \langle \langle v_i v_j \rangle^M \langle \cos(\theta_i - \theta_j) \rangle^C \rangle^M \quad \forall (i,j) \in \mathcal{L} \tag{4b}$$
$$\Im(W_{kk}) = \langle \langle v_i v_j \rangle^M \langle \sin(\theta_i - \theta_j) \rangle^S \rangle^M \quad \forall (i,j) \in \mathcal{L} \tag{4c}$$
$$S_{ij} + S_{ji} = Z_{ij} l_{ij} \quad \forall (i,j) \in \mathcal{L} \tag{4d}$$
$$|S_{ij}|^2 \leq W_{ii} l_{ij} \quad \forall (i,j) \in \mathcal{L} \tag{4e}$$

The superscripts $T$, $M$, $C$ and $S$ represent convex envelopes for the square, bilinear product, cosine, and sine functions, respectively. $Z_{ij}$ denotes the line impedance. The resulting relaxation of the AC-OPF problem is formulated as a second-order cone program (SOCP) that minimizes a user-defined objective function.

### *2.2 Stability*

OPs that are feasible are not necessarily small-signal stable and vice versa. While feasibility ensures that an OP does not violate operational limits, it does not account for the system's ability to remain stable following a disturbance. As such, stability extends beyond feasibility by evaluating the system's ability to recover and settle at a stable OP after a disturbance. These disturbances can manifest as changes in load, generation, or faults, and may affect different types of stability. A critical subset that receives more attention due to the increase of inverter-dominated generators is small signal stability. Generally, a disturbance in this context is considered small if the dynamic response of the system is accurately captured when linearizing the differential equations that govern the system response. Typically, in power systems, this includes load and generation fluctuations or generator outages of up to around 0.1 p.u. power with respect to the base power of the entire system.

Small signal stability can be assessed by analyzing the linearized differential equations of a system around a steady-state operating condition, i.e. the state-space representation of the system. The eigenvalues $\lambda_n = \sigma_n + j\omega_n$ of the linearized system state space matrix $\mathbf{A}$, provide insights into system stability. Generally, a system is stable if the real part of all eigenvalues of the state-space matrix is negative. Besides the distinction of stability and instability, the real and imaginary parts of the eigenvalue permit an analysis on the oscillation magnitude and duration of an eigenmode after a disturbance: The smaller the real part $\sigma_n$, the faster a post-disturbance oscillation decays. In contrast, the imaginary part $\omega_n$ indicates the oscillation frequency.

The damping ratio $\zeta$ characterizes the rate at which the magnitude of post-disturbance oscillations diminishes and is calculated as:

$$\zeta = \frac{-\sigma_n}{\sqrt{\sigma_n^2 + \omega_n^2}}. \tag{5}$$

Consequently, the damping ratio provides valuable insights into the post-disturbance dynamic behavior of that mode:

- Positive damping $\zeta > 0$ indicates a decaying oscillation with rate $\zeta$. As such, the smaller the damping, the slower the oscillation decays.
- Zero damping, $\zeta = 0$, indicates a stable, sinusoidal oscillation of the mode in the post-disturbance dynamics.
- Negative damping, $\zeta < 0$, indicates an increasing oscillation magnitude over time, rendering the mode, and thus the entire system, unstable.

Note that one unstable mode can cause the system to collapse, and the system is considered small-signal unstable if a single mode is subject to negative damping. Thus, small-signal stability assessment typically aims to identify the least damped mode. In addition, power systems are typically operated with a stability margin, e.g. the damping of the mode closest to the instability boundary should be higher than a pre-defined margin.

The integration of renewable energy sources, such as wind and solar, further challenges small-signal stability due to their intermittency and lower predictability, increasing generation variability. Additionally, renewable generators do not contribute to system inertia, thus lowering overall inertia levels. As such, power systems with high integration of renewable generation are more susceptible to small disturbances because they inherit less inertia to dampen frequency changes. Consequently, small-signal stability analysis is becoming more crucial for early detection of stability issues, particularly in low-carbon grids, to ensure secure, reliable grid operation.

For our proposed sampling technique, a stability index that measures the distance to the stability boundary is required. For example, if the focus was on voltage stability, the voltage margin could be used as a measure. However, determining the voltage margin at every bus requires extensive analysis, thus it is not considered here and should be considered in future work. In contrast, the damping ratio is a computationally efficient, system-wide measure for small-signal stability, motivating its use as the assessment criterion.

### *2.3 Security Boundary*

The security boundary $\gamma$ divides the secure from the insecure region. It can correspond to a specific stability boundary, e.g. small-signal or voltage stability. It can also represent a certain stability margin, i.e. OPs not satisfying the margin belonging to the insecure region. It can also be a combination of security indices, e.g. the intersection of AC feasibility and small-signal stability, as it is defined in this work.

The OPs around the stability boundary are potentially extremely valuable for data-driven DSA. As such, the high information content (HIC) region is defined as the set of OPs within a margin $\beta$ of the security boundary. Formally, the HIC region includes all points satisfying:

$$\Omega = \{\mathrm{OP}_k \in \Psi | \gamma - \beta < \gamma_k < \gamma + \beta\}, \tag{6}$$

where $\Omega$ is the set of OPs belonging to the HIC region, $\Psi$ is the total set of OPs and $\gamma_k$ is the value of the chosen stability margin for the $k$-th OP.

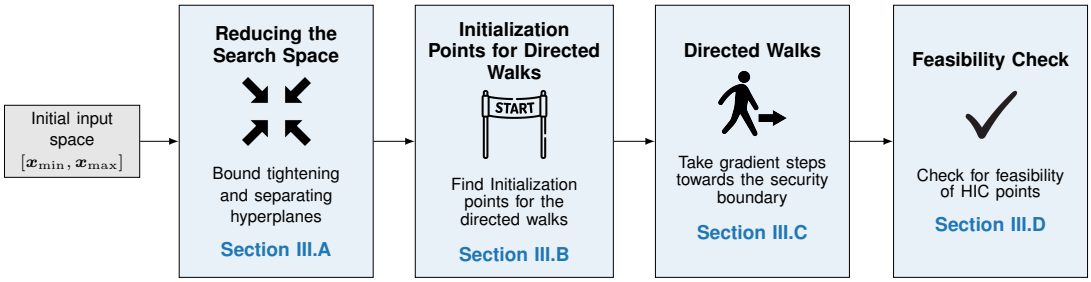


Fig. 1. Flowchart of the proposed methodology: In 3.1, we begin by reducing the initial search space using OBBT and separating hyperplanes. Following this, 3.2 outlines how sampling is performed within the reduced space to initialize the directed walk. The directed walk algorithm itself is explained in 3.3, while 3.4 concludes with a final feasibility check of the identified operating points.

For the remainder of this paper, the security boundary $\gamma$ is represented by the combination of feasibility and a pre-defined level of minimum damping ratio, which corresponds to our stability margin.

## 3 Dataset Generation Considering the Security Boundary

This Section outlines the methodology for efficiently sampling within the HIC region. The goal is to generate a dataset with a high number of samples in the HIC region while maintaining a balanced distribution of secure and insecure samples. Fig. 1 presents an overview of the proposed method. The process begins by reducing the input search space using optimization-based bound tightening and a separating hyperplanes algorithm. The next step involves generating initialization points for the directed walk algorithm, which is subsequently used to identify OPs near the security boundary. Finally, a feasibility check is applied to the HIC-region samples. The remainder of this Section provides a detailed mathematical formulation of each step in the methodology.

To simplify the analysis, the methodology focuses on feasibility and small-signal stability. However, it can be easily extended to account for contingencies, generation or load uncertainty, and other types of stability. The relevant degrees of freedom are defined by the control variables described by the input vector $\mathbf{x}$ as follows:

$$\mathbf{x} = \begin{bmatrix} P_{G_i} \\ |V_j| \\ S_{D_k} \end{bmatrix} \quad \forall i \in \mathcal{G} \setminus \{\text{slack}\}, \quad \forall j \in \mathcal{G}, \quad \forall k \in \mathcal{D}, \quad (7)$$

where $P_{G_i}$ represents the active power generated by each generator $i \in \mathcal{G}$, excluding the generator at the slack bus, $|V_j|$ denotes the voltage magnitude of each generator $j \in \mathcal{G}$, and $S_{D_k}$ corresponds to the complex power demand at each bus $k \in \mathcal{D}$. All other states in the AC-OPF are determined by solving the non-linear AC power flow equations with the input vector $\mathbf{x}$. The bounds on $\mathbf{x}$ are specified in (1a) and (1b). It is important to note that, when using the QC relaxation, the variable $|V_j|$ is substituted with $v_j$.

### 3.1 *Reducing the Search Space*

The main challenge in creating a balanced dataset arises from the large number of control variables. As the system size increases, the dimensionality of $\mathbf{x}$ grows exponentially, making naive sampling of the entire input space computationally intractable for large systems. Moreover, as demonstrated in [17], a significant portion of the input space, $\mathbf{x} \in \left[\mathbf{x}^{\min}, \mathbf{x}^{\max}\right]$, is infeasible. To address this, we employ two methods to reduce the initial input space. First, we apply the OBBT algorithm proposed in [19] to achieve a tighter relaxation of the AC-OPF. Second, we use the separating hyperplanes method from [17] to pre-classify large portions of the input region as insecure.

#### 3.1.1 *Optimization-Based Bound Tightening*

The OBBT algorithm has two main objectives. First, it directly tightens the bounds on the input vector $\mathbf{x}$, and second, it improves the QC relaxation, which in turn makes the subsequent separating hyperplane algorithm more efficient. The bounds on the voltage magnitudes and angles significantly affect the tightness of the envelopes in (4) and thus influence the quality of the QC relaxation. Through iterative bound tightening, we adjust the voltage magnitude and angle bounds to obtain a tighter relaxation, which also reduces the initial input space.

To achieve this, we employ an OBBT technique as described in [19], which systematically refines the bounds for critical system variables. Specifically, we tighten the bounds on voltage magnitudes ($V^{max}$, $V^{min}$) at each bus and angle differences ($\theta_{ij}^{min}$, $\theta_{ij}^{max}$) for each transmission line. This process involves iteratively solving convex optimization problems, where at each step, the minimum and maximum achievable values for the optimization variables—i.e. voltage magnitude and angle difference—are determined within the relaxed problem. The procedure is repeated for a predefined number of iterations, as tightening one variable's bounds may allow for further tightening of previously adjusted bounds, gradually narrowing the feasible region. As a result, we obtain a more precise characterization of the operating space, with tightened bounds for the input vector $\mathbf{x}$, denoted as $\left[\mathbf{x}^{\mathrm{BT,min}}, \mathbf{x}^{\mathrm{BT,max}}\right]$. Any input $\mathbf{x}$ falling outside this tightened region is guaranteed to be infeasible in the full non-convex AC-OPF.

### 3.1.2 *Infeasibility Certificates*

After the bound tightening process, the input space is further reduced by generating infeasibility certificates using separating hyperplanes, as proposed in [17]. This method eliminates infeasible regions of the input space in a computationally efficient manner, significantly narrowing the search space. Consider an OP $\hat{\mathbf{x}}$ that is infeasible with respect to the non-convex AC-OPF. To find the closest feasible dispatch $\mathbf{x}^\star$ within the convex QC relaxation, we solve an optimization problem that minimizes the distance between $\hat{\mathbf{x}}$ and $\mathbf{x}$, subject to the QC relaxation constraints. Specifically, we minimize the radius $R$ which quantifies the distance from $\hat{\mathbf{x}}$ to the feasible space of the QC relaxation:

$$\min_{\mathbf{x},\mathbf{S},\mathbf{S_G},\mathbf{v},\theta,\mathbf{l},\mathbf{W},R} R \tag{8a}$$

$$\text{s.t.} (1b), (1c), (1d), (3), (4), (7) \tag{8b}$$

Here, the QC relaxation constraints include generator limits, line limits, power balance, and other relevant variables, as shown in equation (8b). The radius $R$ is determined by the Euclidean distance between the infeasible point $\hat{\mathbf{x}}$ and the closest feasible point $\mathbf{x}$, and is mathematically represented as:

$$\sqrt{\sum_{k \forall \mathcal{X}} (x_k - \hat{x}_k)^2} \leq R \tag{8c}$$

If the radius $R^*$ obtained from solving the optimization problem is greater than zero - meaning that the OP $\hat{\mathbf{x}}$ is infeasible with respect to the convex relaxation - then no feasible OP $\mathbf{x}$ exists that is closer to $\hat{\mathbf{x}}$ than the obtained OP $\mathbf{x}^\star$. Using this property, we can construct an infeasibility certificate in the form of a hypersphere [14] or hyperplane to classify a large volume as infeasible. If a non-zero radius $R^*$ and optimal solution $\mathbf{x}^\star$ are obtained by the optimization in (8), all vectors $\mathbf{x}$ which comply with the following half-space inequality constraint are infeasible with respect to the AC-OPF constraints (1):

$$\vec{\mathbf{n}}^{\mathtt{T}}(\mathbf{x} - \mathbf{x}^\star) < 0 \tag{9}$$

The normal vector of the hyperplane is defined as $\vec{\mathbf{n}} := \mathbf{x}^\star - \hat{\mathbf{x}}$, where $\mathtt{T}$ denotes the transpose operator. Infeasibility with respect to the QC relaxation constraints, given in (8b), guarantees infeasibility with respect to the non-convex AC-OPF constraints in (1). For a detailed explanation and mathematical proof, the reader is referred to our previous work in [17]. Fig. 2 presents an illustration of how the hyperplane is constructed using $\hat{\mathbf{x}}$ and $\mathbf{x}^\star$. This Figure is reproduced from [17].

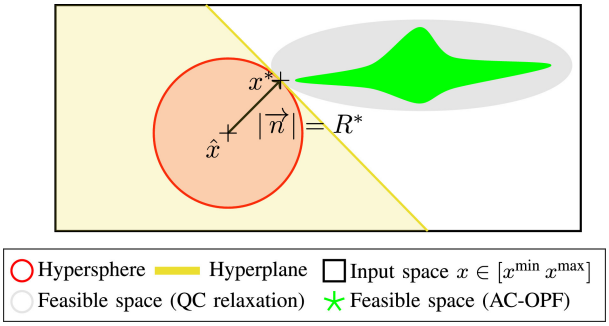


Fig. 2. Illustrative example of an infeasibility certificate using a hyperplane. Given an infeasible OP $\hat{\mathbf{x}}$, we compute the closest feasible OP $\mathbf{x}^\star$ under the QC relaxation. The normal vector $\vec{\mathbf{n}}$ is perpendicular to the feasible space of the QC relaxation, which encompasses the feasible space of the non-convex AC-OPF. All points on the left side of the hyperplane are guaranteed to be infeasible to both the QC relaxation and the AC-OPF problem. This Figure is reproduced from our previous work in [17].

### 3.1.3 *Convex Polytope Construction*

Using the method of separating hyperplanes, we employ the algorithm proposed in [17] to classify large portions of the input space as infeasible. The initial input region defined by $\mathbf{x}$, combined with subsequently constructed hyperplanes, defines a convex polytope represented by the inequaltiy $\mathbf{Ax} \leq \mathbf{b}$. We keep track of the remaining unclassified input space volume of the polytope, denoted as $V$. The row of $\mathbf{A}$ and the corresponding entry in $\mathbf{b}$ for the hyperplane derived in (8) can be written as $\mathbf{A_k} := \vec{\mathbf{n}}^{\mathtt{T}}$ and $b_k := \vec{\mathbf{n}}^{\mathtt{T}}\mathbf{x}^\star$. The addition of this hyperplane reduces the volume of the polytope to $V^{HP}$. "Hit-and-Run" sampling [24] provides an efficient method for uniformly sampling points within a convex polytope. This approach enables the iterative construction of hyperplanes by restricting sampling to the currently unclassified region.

The full algorithm from [17] is presented in Algorithm 1 for completeness. In Lines 1 to 5, the algorithm begins by defining a convex polytope constrained by the tightened input bounds, $\mathbf{x}^{\mathrm{BT,min}}$ and $\mathbf{x}^{\mathrm{BT,max}}$, obtained through bound tightening. In Line 6, for a fixed number of iterations, denoted as $N_1$, the algorithm iteratively samples points within the polytope defined by $\mathbf{A}^{(k)}\mathbf{x} \leq \mathbf{b}^{(k)}$. The parameter $N_1$ represents the maximum number of iterations, corresponding to the maximum number of hyperplanes that will be constructed. In each iteration, a random point, $\mathbf{x}^{(k)}$, is drawn from the current feasible region. Line 8 checks the feasibility of the OP. If the OP is infeasible, indicated by $R > 0$, a new hyperplane, perpendicular to $\vec{\mathbf{n}}$, is constructed and added to the polytope in Lines 11 to 12, reducing the unclassified region. The volume of the unclassified space is tracked by computing the volume $V^{HP(k)}$ of the convex polytope at each iteration $k$. In Lines 14 to 17, a stopping criterion is applied: the algorithm halts if the polytope's volume does not decrease by more than a user-specified percentage, $\tau$, over a defined number of consecutive iterations, $\eta$. This ensures that the algorithm stops once further refinement does not significantly reduce the polytope volume.

**Algorithm 1** Separating hyperplanes algorithm

Run bound tightening and obtain $\mathbf{x}^{\mathrm{BT,min}}$ and $\mathbf{x}^{\mathrm{BT,max}}$
Set iteration count: $k \leftarrow 0$
Initialize unclassified region $\mathbf{A}^{(0)}x \leq \mathbf{b}^{(0)}$:
  $\mathbf{A}^{(0)} := \left[\mathbf{I}^{|\mathbf{x}|\times|\mathbf{x}|} - \mathbf{I}^{|\mathbf{x}|\times|\mathbf{x}|}\right]^{\mathsf{T}}$
  $\mathbf{b}^{(0)} := \left[(\mathbf{x}^{\mathrm{BT,max}})^{\mathsf{T}}\,(\mathbf{x}^{\mathrm{BT,min}})^{\mathsf{T}}\right]^{\mathsf{T}}$
**while** $k \leq N_1$ **do**
  Draw random $\mathbf{x}^{(k)}$ from inside $\mathbf{A}^{(k)}\mathbf{x} \leq \mathbf{b}^{(k)}$
  Solve (8) with $\hat{\mathbf{x}} := \mathbf{x}^{(k)}$ and obtain $\mathbf{x}^{\star}$
  **if** $R > 0$ **then**
    Add hyperplane to reduce unclassified region:
      $\mathbf{A}^{(k+1)} = \left[(\mathbf{A}^{(k)})^{\mathsf{T}}\vec{\mathbf{n}}\right]^{\mathsf{T}}$
      $\mathbf{b}^{(k+1)} = \left[(\mathbf{b}^{(k)})^{\mathsf{T}}\vec{\mathbf{n}}^{\mathsf{T}}\mathbf{x}^{\star}\right]^{\mathsf{T}}$
  **end if**
  Compute polytope volume $V^{HP(k)}$
  **if** $V^{HP(k)} > (1-\tau)V^{HP(k-1)}$ for $\eta$ consecutive iterations **then**
    **break**
  **end if**
  $k \leftarrow k+1$
**end while**

### 3.2 Initialization Points for Directed Walks

Once the separating hyperplanes algorithm is completed, we will have classified a large portion of the input space as insecure. The next step involves continuing the sampling process within the constructed convex polytope which includes the complete feasible region for the steady-state operation, as well as several infeasible points. This phase aims to identify initialization points for directed walks towards the security boundary. In this paper, the security boundary is defined as the intersection of the AC feasible and small-signal stable operating points. For this purpose, a predefined number of samples, denoted as $N_2$, are drawn from the convex polytope defined by the final set of hyperplanes $\mathbf{A}^{(N_1)}\mathbf{x} \leq \mathbf{b}^{(N_1)}$, where $N_1$ refers to the number of iterations of the separating hyperplane algorithm executed in the previous step. Then, for each sample the AC power flow equations are solved to check for feasibility. If the sample is feasible, it is added to the feasible set, otherwise to the infeasible set. If the sample is infeasible, another AC power flow enforcing the reactive power limits of the generators is conducted. This step is added because preliminary tests have shown that generator reactive power limits are often the only violation. If the sample is feasible, adjusted setpoints are added to the feasible set. If the sample is still infeasible, the non-convex optimization in (10) is solved to find the closest AC-OPF feasible dispatch to this sample. Compared to (8), where we solved this optimization with respect to the convex QC relaxation, we now solve the full non-convex AC-OPF.

$$\min_{\mathbf{x},\mathbf{V},\mathbf{S},\mathbf{S_G},R} R \tag{10a}$$

$$\text{s. t.} (1), (7), (8c) \tag{10b}$$

The dispatch found in (10) is added to the feasible set. This procedure is repeated for all $N_2$ samples. Next, small-signal stability for all the samples in both the feasible and infeasible set is assessed by computing the damping ratio of the least damped mode.

### 3.3 Directed Walks

The feasible samples obtained from the convex polytope serve as initialization points for the directed walk algorithm. Section 3.3.1 outlines the directed walk algorithm, Section 3.3.2 discusses its use of the damping ratio as a sensitivity measure, and Section 3.3.3 describes the approach for obtaining a high number of samples in the HIC region.

#### 3.3.1 Directed Walk Algorithm

The directed walk algorithm, first applied in [14], is a gradient descent algorithm used to find OPs closer to a user-defined stability index. The algorithm employs a variable step size $\alpha$, which depends on the distance $d$ of the OP from the security boundary $\gamma$. The distance $d(\mathrm{OP}_k)$ of an OP to the security boundary is defined as:

$$d(\mathrm{OP}_k) = |\gamma_k - \gamma|, \tag{11}$$

where $\gamma_k$ represents the value of the stability index for OP $\mathrm{OP}_k$. The variable step size $\alpha_k$ is defined as:

$$\alpha_k = \begin{cases} \epsilon_1 \cdot P^{max}, & \text{if} d(\mathrm{OP}_k) > d_1, \\ \epsilon_2 \cdot P^{max}, & \text{if}, d_1 \geq d(\mathrm{OP}_k) > d_2, \\ \epsilon_3 \cdot P^{max}, & \text{if}, d_2 \geq d(\mathrm{OP}_k) > d_3, \\ \epsilon_4 \cdot P^{max}, & \text{otherwise} \end{cases} \tag{12}$$

The step size $\alpha$ is a function of the generator's maximum capacities $P^{max}$ and scalars $\epsilon_{1-4}$, which depend on the distance $d_{1-3}$ to the security boundary. As the OP approaches the boundary, the step size is gradually reduced to prevent overshooting. Both $\epsilon$ and $d$ are user-defined and are adjustable. The direction of the step is determined by the steepest descent of the metric $d(\mathrm{OP}_k)$. The next OP is found using the step size $\alpha_k$ and the gradient $\nabla d(\mathrm{OP}_k)$:

$$\mathrm{OP}_{k+1} = \mathrm{OP}_k - \alpha_k \cdot \nabla d(\mathrm{OP}_k) \tag{13}$$

This method is compatible with any sensitivity measure that can quantify the distance to a selected stability index. In this work, we use the damping ratio to evaluate small-signal stability.

#### 3.3.2 Sensitivity Measure

For small-signal stability, the direction of the directed walk step is determined by the sensitivity of the damping ratio $\zeta$. More specifically, we use the damping ratio of the least

damped mode of the system. The sensitivity of the damping ratio $\zeta$ to parameter $\rho_i$ can be found analytically by taking the partial derivative of the damping with respect to system parameter $\rho_i$. However, because the computation of this partial derivative is computationally demanding, the damping ratio sensitivity of $\zeta$ to $\rho_i$ is determined by a small perturbation of $\rho_i$. In this work, the parameters $\rho_i$ we perturb are the active power setpoints of the generators.

#### *3.3.3 Sampling the HIC Region*

The directed walk algorithm is applied to all $N_2$ samples, with a user-specified number of $\kappa_{\max}$ directed walks performed for each initialization point. When a directed walk enters the HIC region, the following steps are executed:

i) The current OP is stored.
ii) All OPs around this point are analyzed to determine if they also belong to the HIC region. Any additional samples within the HIC region are also stored.
iii) The directed walk continues along one single dimension with the minimal step size until either an OP already in the dataset is encountered, the directed walk exits the HIC region or the directed walk algorithm reaches its limit of $\kappa_{\text{HIC}}$ steps. A discretization interval between generator setpoints of 1MW is used.

All unique OPs found within the HIC region are stored for feasibility evaluation. The DW algorithm is highly parallelizable, enabling simultaneous processing of multiple OPs.

### *3.4 Final Feasibility Check*

After applying the directed walk algorithm to all initialization points, a feasibility check is performed on the samples in the HIC region. Since only generator active power setpoints are modified during the directed walk, many resulting OPs exhibit minor violations, such as voltage magnitude, reactive power limits, or active power limits of the slack bus generator. For all infeasible samples in the HIC region, the optimization problem in (10) is solved to determine the closest AC-feasible dispatch. The damping ratio of the lowest damped mode is then recomputed. After this final step, the dataset contains a diverse set of unique OPs near the security boundary, balanced between secure and insecure OPs.

## 4 Case Studies

This Section provides a comprehensive analysis of six datasets—three for each of two test systems—generated using our proposed method and two benchmark methods: naive sampling and importance sampling. The goal is to illustrate the impact of a high share of OPs in the HIC region. We also train DTs on these six datasets to assess the performance of a data-driven DSA tool, comparing datasets with a high share of HIC OPs to those derived from the benchmark methods. This Section first outlines the case study settings and benchmark sampling methods, followed by the implementation details of the sampling techniques. Next, we conduct a data analysis of the obtained datasets and present the results of the DT performance using the different datasets.

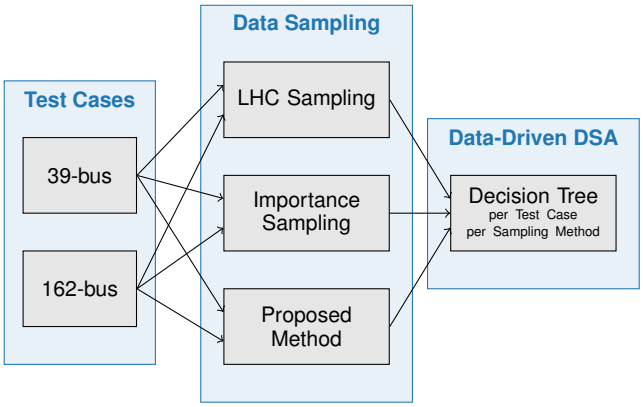


Fig. 3. Datasets are generated for evaluation and decision tree training. For each test system, one dataset is created using each of the three sampling methods. With two test systems, this results in six datasets, each corresponding to a decision tree.

### *4.1 Case Study Settings*

We evaluate the proposed method using two test systems from the PGLib-OPF networks (v23.07) [25]: the 39-bus and 162-bus systems. The 39-bus system comprises 10 generators and has a peak demand of 6.3 GW, while the 162-bus system includes 12 generators with a peak demand of 7.4 GW. Both systems operate without renewable energy resources, and their nominal load profiles are scaled to 80 % of the nominal load values for this analysis. For both systems, we employ a constant power load model, where both active and reactive power loads are scaled equally.

For both test cases at this loading level, we solved an AC-OPF and found that the damping ratio of the most critical mode is slightly above 3%. This motivates the definition of the security boundary, determined by feasibility and a minimum damping ratio of $\zeta_{min} = 3\%$, a threshold we consider to provide a sufficiently large margin for stability. The boundary ensures that, at this loading level, the lowest dampened mode meets the minimum damping requirement, thereby maintaining a sufficient stability margin. Furthermore, we select a HIC margin value of $\beta = 0.25\%$, resulting in a HIC region defined as $2.75\% < \gamma_{min} < 3.25\%$. This margin is carefully selected during pre-tests to strike a balance between ensuring a sufficiently large HIC region, which allows for effective boundary characterization, and maintaining a narrow search space that facilitates the selective inclusion of OPs. This approach optimizes both the quality and quantity of the resulting dataset.

All generators in both test systems are modeled as fourth-order round-rotor generators using the RoundRotorQuadratic model from [26]. Each generator is equipped with an automatic voltage regulator (AVR) of type EXST1, a governor of type TGOV1, and a power system stabilizer (PSS) of type IEEEST, all sourced from the library in [26]. The network is modeled using static lines, and they are represented by algebraic equations. Detailed parameter values for the generator models and network data can be found in [20]. All simulations were performed on 20 cores of the DTU HPC cluster [27].

#### *4.1.1 Benchmark Methods*

To benchmark the proposed method, we utilize both a naive sampling approach and an importance sampling method. The naive sampling method employs Latin Hypercube Sampling (LHC) to uniformly sample the entire input space. LHC sampling is a technique that ensures a uniform distribution of samples across the parameter space by dividing each dimension into intervals (subsections) and selecting one sample from each. This approach aims to maximize the minimum distance between samples, ensuring that they are evenly distributed and not clustered together [28]. Such an even distribution is essential for capturing a comprehensive representation of the parameter space, including regions that may be infeasible or critical for the analysis. However, while LHC improves sampling efficiency compared to simple random sampling, it becomes computationally expensive as the dimensionality increases, since the number of required samples grows rapidly with the number of parameters.

For each sampled OP, AC power flow calculations are performed to assess feasibility. Infeasible samples are added to the infeasible set, while feasible samples are obtained by projecting the infeasible ones onto the AC feasible region, as outlined in (10), ensuring a balanced dataset. Following this, small-signal stability analysis is conducted on both feasible and infeasible samples, allowing for the complete classification of the dataset.

The importance sampling method, on the other hand, focuses on biasing the sampling process toward the HIC region. Initially, LHC sampling is used to identify OPs within the HIC region. A multivariate normal distribution, $\mathcal{M}(\mu, \Sigma)$, is then fitted to the feasible samples within this region, where $\mu$ represents the mean and $\Sigma$ the covariance matrix of the feasible points. To bias the sampling process towards points closer to the boundary, the covariance matrix $\Sigma$ is scaled by a factor $s < 1$, resulting in a reduced covariance matrix $\Sigma_{red} = s \cdot \Sigma$. In this study, we use $s = 0.25$. Samples are then drawn from the adjusted distribution $\mathcal{M}(\mu, \Sigma_{red})$. For these new samples, AC power flow calculations are performed to determine feasibility. Infeasible samples are added to the infeasible set, while feasible samples are obtained by projecting the infeasible ones onto the AC feasible region, as outlined in (10). Finally, small-signal stability analysis is performed on all sampled points.

#### *4.1.2 Dataset Construction Parameters*

For the 39-bus system, 25 000 samples are drawn uniformly across the entire input space for the LHC benchmark, while 10 000 samples are drawn for the 162-bus system. In contrast, the importance sampling benchmark starts with 4 000 samples drawn using LHC. These are used to construct a multivariate normal distribution that guides the subsequent biased sampling process. For the 39-bus system, an additional 25 000 samples are drawn, and for the 162-bus system, 10 000 samples are drawn, with a focus on the HIC region.

For the method proposed in this paper, the parameters used to construct the convex polytope differ for each test system. In the case of the 39-bus system, the polytope is constructed using either $N_1 = 100$ hyperplanes or until the volume reduction is less than $\tau = 5\%$ for $\eta = 30$ consecutive iterations. For the 162-bus system, the polytope is created with up to $N_1 = 400$ hyperplanes, or until the same volume reduction of less than $\tau = 5\%$ is achieved after $\eta = 100$ consecutive iterations. These parameter values are based on the findings in [17]. From the resulting polytopes, $N_2 = 4\ 000$ samples are drawn to initialize the DWs.

For comparison and analysis, a subset of 10 000 samples is randomly selected from the six datasets generated by the proposed method and the two benchmark methods. These subsets are then used for case studies on the two test systems.

### *4.2 Implementation*

The entire framework for the proposed method, as well as for both benchmarks, is developed in Julia. For both AC power flows and AC-OPFs, we rely on Powermodels.jl [29] and JuMP [30]. All optimization problems are solved with the IPOPT solver [31]. Power system modeling uses PowerSystems.jl [26], while small-signal stability analysis employs PowerSimulationsDynamics.jl [32]. The volume of the convex polytope is computed through the Volesti package [33] in R. For the OBBT algorithm we perform three iterations using the OBBT algorithm from [29]. Subsequently, the DW algorithm is executed, with a maximum of $\kappa_{max} = 30$ steps towards the HIC region, followed by up to $\kappa_{HIC} = 15$ steps within the HIC region. The values for $\epsilon_{1-4}$ are empirically tuned and set to $[4, 3, 2, 1]$ respectively.

We train a total of six DTs to assess the impact of the different sampling techniques on data-driven DSA. For each of the two test systems using datasets from three methods, the proposed method, the LHC sampling benchmark, and the importance sampling benchmark, an individual DT is obtained. Training employs the Classification and Regression Tree (CART) algorithm [34] in Scikit-learn [35], with default settings, Gini impurity as the splitting criterion, and a maximum tree depth of 5. In order to assess the performance of the DTs and to prevent underfitting or overfitting, we apply 10-fold cross-

validation while simultaneously monitoring both training and testing accuracies. Additionally, we apply tree pruning with a ccp-alpha score of 0.01 to further improve model generalization. The DTs are employed for binary classification, where a value of 1 indicates a secure OP and a value of 0 indicates an insecure OP. For all six datasets, a 75:25 train-test split is consistently applied for training and evaluating the DTs.

### *4.3 Dataset Comparison*

This Section evaluates the datasets generated by the proposed method and two benchmark methods. We begin by providing both quantitative and visual analyses of the datasets for the 39-bus and 162-bus systems.

Fig. 4 shows the statistics of the 10 000 randomly sampled points across the three datasets for both systems. The importance sampling method begins with LHC sampling to pinpoint samples in the HIC region, which are then used to construct the multivariate normal distribution. In the initial sampling, 110 OPs (2.8% of the total samples) were identified in the HIC region for the 39-bus system, and 138 OPs (3.5%) for the 162-bus system. This highlights the challenge of identifying samples close to the security boundary.

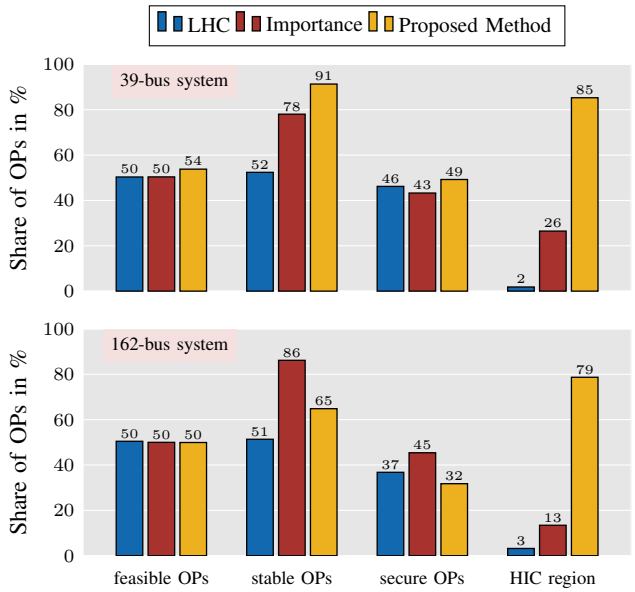


Fig. 4. Share of feasible, stable, secure, and HIC-region samples for the three datasets from the 39-bus system (top) and the three datasets from the 162-bus system (bottom). An OP is feasible, if it satisfies the AC-OPF constraints. An OP is stable, if it is small-signal stable. An OP is secure, if it is both feasible and stable.

All three methods ensure a balanced distribution of feasible and infeasible samples, as depicted in the left column of Fig. 4, where each infeasible sample is paired with a feasible counterpart. Notably, both the importance sampling and proposed methods generate a higher proportion of stable samples than unstable ones. For importance sampling, this is likely because the multivariate normal distribution is centered around secure and thus stable OPs, leading to the generation of many stable samples. However, these samples are not necessarily feasible and, therefore, not always secure. The proposed method, on the other hand, generates most of its samples during the directed walk phase, which starts from a stable OP. As a result, it predominantly produces stable OPs throughout this phase. For the 39-bus system, all methods maintain a balanced distribution between secure and insecure samples. However, for the 162-bus system, the LHC and proposed method show a slightly lower proportion of secure samples.

The most significant difference among the methods is the representation of samples in the HIC region. For the 39-bus system, 85% of the samples generated by the proposed method belong to the HIC region, while for the 162-bus system, this proportion is 79%. This result demonstrates the method's ability to concentrate on the critical region near the security boundary.

To further illustrate the distribution of OPs, we present several visualizations of the datasets in two dimensions, specifically the active power setpoints of two generators in each test system. Figures Fig. 5 and Fig. 6 show the spread of secure and insecure OPs for the 39-bus system. In Fig. 5, we observe that while the LHC method samples the entire input space, both the proposed and importance sampling methods identify a similar region of importance. However, the input space includes both secure and insecure OPs in close proximity, suggesting overlap between the categories. This overlap is further emphasized in Fig. 6, which shows a similar distribution across the input space.

For the 162-bus system, Figs. 7 to 9 show more distinct regions of importance, where secure samples are confined to a specific region, and the remaining space is occupied by insecure samples. The proposed method identifies a significant number of OPs within the HIC region, while the LHC method samples many secure and insecure OPs around this region, but not necessarily within it. The importance sampling method captures many OPs but slightly offsets the security boundary, highlighting the difficulty of accurately modeling the full security boundary in high-dimensional spaces.

Moreover, Fig. 8 shows a distinct secure region, where all three methods identified a region of importance at the top of the secure area. The importance sampling method captures many points in this region, while the proposed method captures part of it but does not fully encompass the entire boundary. Similarly, Fig. 9 reveals a well-defined secure region. The LHC and importance sampling methods primarily sampled along a straight line, indicating a potential boundary but not necessarily capturing the HIC region. In contrast, the proposed method sampled along a vertical line, with many points belonging to the HIC region and approaching it from within the secure region. Additionally, the importance sampling method

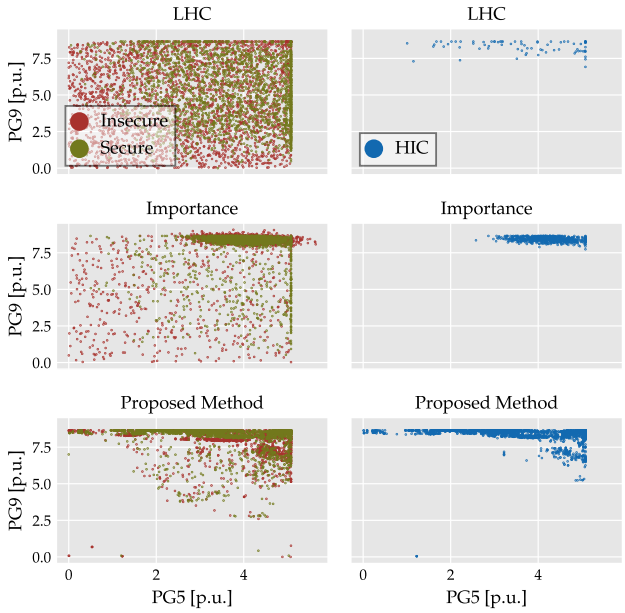


Fig. 5. Spread of secure and insecure OPs for the 39-bus system, plotted for generator 5 and 9.

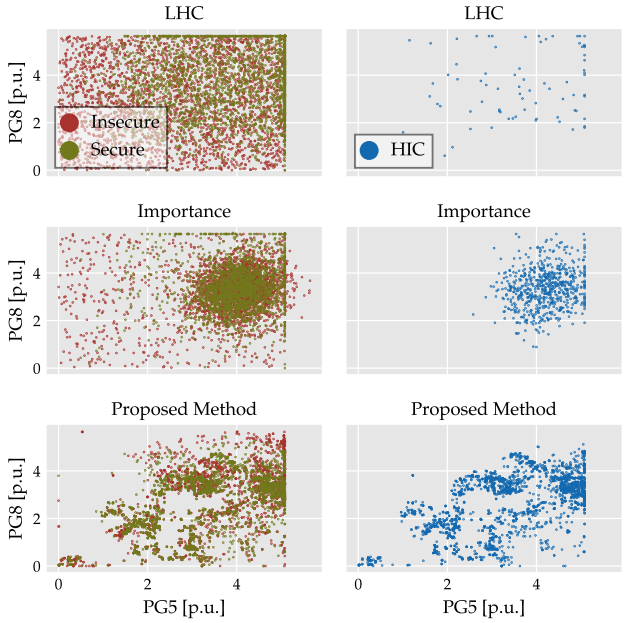


Fig. 6. Spread of secure and insecure OPs for the 39-bus system, plotted for generator 5 and 8.

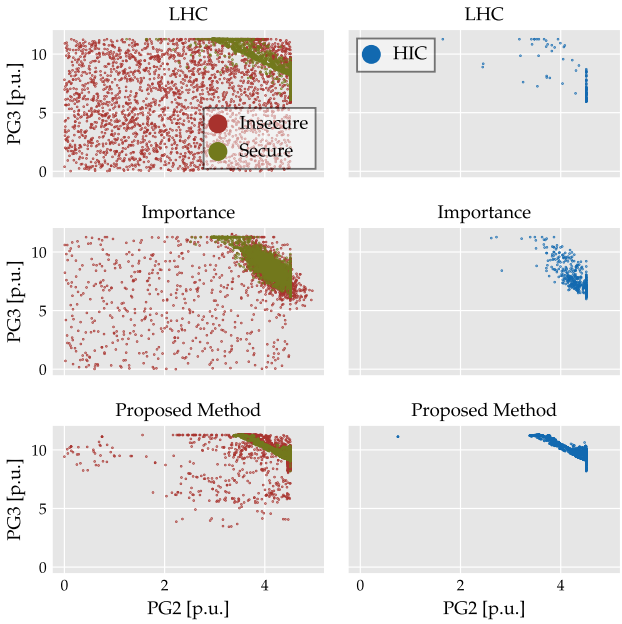


Fig. 7. Spread of secure and insecure OPs for the 162-bus system, plotted for generator 2 and 3.

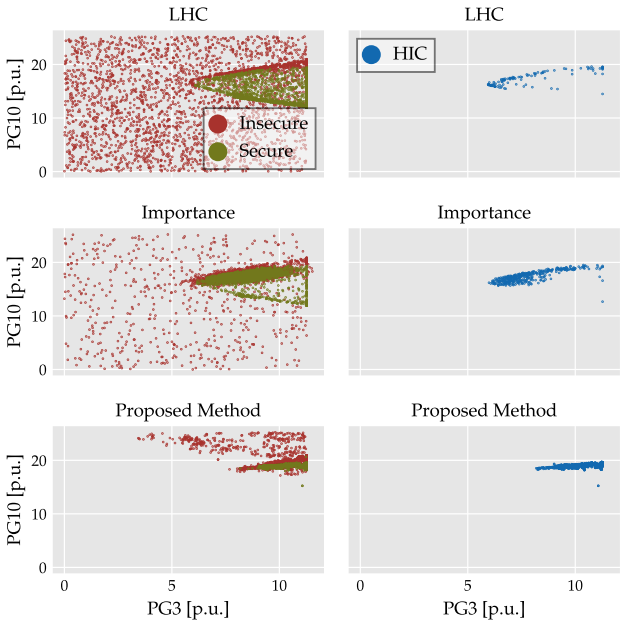


Fig. 8. Spread of secure and insecure OPs for the 162-bus system, plotted for generator 3 and 10.

concentrated most of its HIC samples around a single point. These examples highlight the challenge of accurately capturing the security boundary using a multivariate distribution.

In summary, the analysis in Figs. 7 to 9 demonstrates the efficacy of the proposed method, especially in reducing the input search space. Unlike LHC and importance sampling, which sample across the entire input space, the proposed method focuses on a more constrained region. This targeted approach enables more efficient sampling, as reflected in the substantial reduction of the feasible polytope's volume—by a factor of 111 for the 39-bus system (to 0.9% of its original size) and by a factor of 2 537 for the 162-bus system (to just 0.039% of its original size). This means that the proposed method already classified more than 99% of the initial search space as insecure, allowing it to focus on the truly interesting region, where (i) all secure OPs, (iii) many insecure OPs, and (iii) the security boundary lie.

In Table I and Table II, we compare (i) the total computational

cost of each dataset generation method for both test systems, and (ii) the time taken to generate 1 000 samples inside the HIC region for each method. For the 39-bus system, the LHC method generated 50 000 samples in about 10 hours, while the importance sampling method produced slightly more samples in less time. Our proposed method generated the most samples, taking 15 hours; from those 15 hours, 1.4 hours (9.3%) were spent on constructing the convex polytope and 6.5 hours (43.3%) were spent on the directed walk algorithm. For the 162-bus system, the proposed method performed more efficiently in terms of time per sample, with the convex polytope construction accounting for 2.9 hours (16.3%) and the directed walk algorithm taking 3.5 hours (19.6%). In all cases, the majority of the time was spent on solving the non-convex AC-OPFs (as defined in (10)) and conducting small-signal stability assessments. The directed walk phase was the primary contributor to sample generation. The convex polytope sampling, in turn, was mainly responsible for producing the initialization points for the walks near the security boundary. Let us highlight at this point that the directed walk algorithm is highly parallelizable, making it suitable for efficient large-scale sample generation; this can significantly reduce the computation time in larger cases. Examining the time required for each method to generate 1 000 samples inside the HIC region, shown in Table I and Table II, we find that our proposed method is approximately 50x-80x more efficient than LHC and 2x-16x more efficient than importance sampling in generating samples inside the HIC region.

### 4.4 Decision Tree Performance

After analyzing the constructed datasets, this section explores whether a higher number of OPs within the HIC region and a more comprehensive spread of OPs along the security boundary can enhance the performance of DTs for data-driven DSA. As such, this Section first describes the metric used to evaluate the DTs, and then presents the performance scores for the various DTs. In this Section, we also (i) examine the location of misclassified OPs, i.e. how close they are to the security boundary, and (ii) assess the impact of balanced vs unbalanced training datasets on the DT performance.

TABLE I
COMPUTATIONAL COST COMPARISON FOR THE 39-BUS SYSTEM.

| Method | # of Samples | Time [h] | Time/ 1k HIC OPs [h] |
|---|---|---|---|
| LHC | 50 000 | 9.9 | 9.9 |
| Importance | 57 884 | 6.4 | 0.4 |
| Proposed Method | 86 672 | 15 | 0.2 |

TABLE II
COMPUTATIONAL COST COMPARISON FOR THE 162-BUS SYSTEM.

| Method | # of Samples | Time [h] | Time/ 1k HIC OPs [h] |
|---|---|---|---|
| LHC | 19 498 | 21.4 | 36.9 |
| Importance | 25 663 | 26.7 | 8.0 |
| Proposed Method | 44 263 | 17.8 | 0.5 |

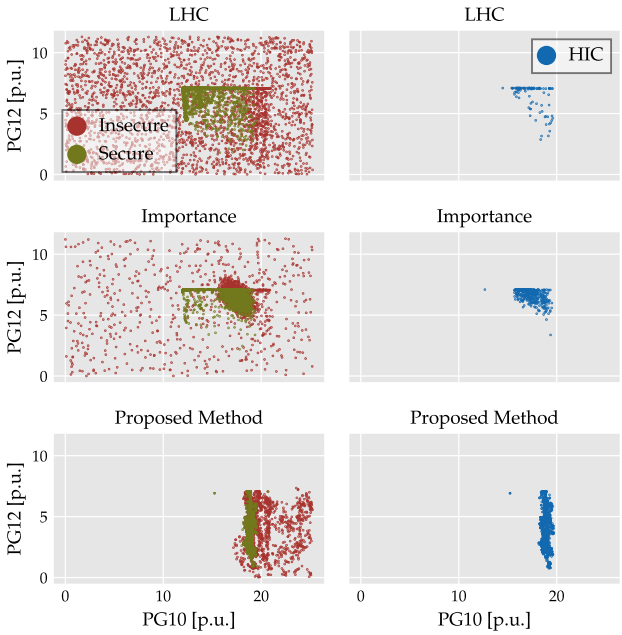


Fig. 9. Spread of secure and insecure OPs for the 162-bus system, plotted for generator 10 and 12.

#### 4.4.1 Evaluation Metrics

To evaluate the performance of the DTs, the F1-Score is applied:

$$\text{F1-Score} = \frac{2T_p}{2T_p + F_p + F_n}, \tag{14}$$

where, $T_p$, $F_p$, and $F_n$ are true positives, false positives, and false negatives, respectively. The F1-score provides a balanced measure of accuracy by considering both false positives (insecure samples incorrectly classified as secure) and false negatives (secure samples incorrectly classified as insecure). This balance is especially important in DSA, where false positives — misclassifying an insecure operational point as secure — can have severe consequences, potentially causing security threats to the system.

The F1-score, being the harmonic mean of precision and recall, is particularly valuable when there is a need to balance the correct identification of positive cases (precision) with the ability to identify all actual positive cases (recall). By accounting for both types of misclassification, the F1-score ensures that the model performs well in distinguishing between secure and insecure operational points, providing a more insightful measure of performance than accuracy alone.

#### 4.4.2 DT Performance

Table III and Table IV present the F1-scores of the DTs trained using three different datasets. Each DT is evaluated on the test

sets generated by the other two sampling methods, as well as an additional test set named *boundary*. This test set contains OPs generated by the proposed method which lie near the security boundary ($2.9\% < \zeta < 3.1\%$).

For the 39-bus system, the DT trained on the dataset generated by the proposed method generalizes well to the test sets derived from both the LHC and importance sampling methods, i.e. signified by the gray box in Table III. Similarly, the DT trained using the importance sampling dataset generalizes well to the LHC test set. However, the reverse is not observed; the LHC DT does not perform well on the importance test set. Notably, the LHC DT outperforms the importance DT on both the proposed method test set and the boundary test set. Overall, the DT trained on the dataset generated by the proposed method outperformed the other DTs across all test sets on average.

The overall conclusions for the 162-bus system overlap with the previous results. The DT trained on the method test set provides the most stable performance when evaluated across the other datasets. In particular, the DT trained on the proposed method's dataset achieves strong performance on the LHC test set and outperforms the LHC DT on the importance sampling test set. In contrast, both the LHC and importance DTs show limited generalization capabilities, particularly on the boundary test set.

#### *4.4.3 Damping Ratio of Misclassified OPs*

Fig. 10 and Fig. 11 provide insights into the misclassification errors of the different DTs for both test systems. For this assessment a large test set, including the points from the LHC, importance sampling, and proposed method sampling, is formed. The diagrams are histograms, where the ordinate depicts the number of misclassified points for a given damping ratio bin. Additionally, the colored section of the bars indicate the cause of misclassification.

TABLE III
39-BUS SYSTEM F1-SCORES FOR THE TRAINED DTS.

| Training \ Testing | LHC | Importance | Proposed Method | Boundary |
|---|---|---|---|---|
| LHC | 0.98 | 0.62 | 0.78 | 0.67 |
| Importance | 0.95 | 0.92 | 0.67 | 0.64 |
| Proposed Method | 0.95 | 0.91 | 0.92 | 0.77 |

TABLE IV
162-BUS SYSTEM F1-SCORES FOR THE TRAINED DTS.

| Training \ Testing | LHC | Importance | Proposed Method | Boundary |
|---|---|---|---|---|
| LHC | 0.98 | 0.76 | 0.60 | 0.42 |
| Importance | 0.97 | 0.97 | 0.64 | 0.48 |
| Proposed Method | 0.99 | 0.82 | 0.90 | 0.72 |

Both figures indicate that misclassifications generally occurs near the security boundary, and within the HIC region that is highlighted in red on the figures. The DT trained on the LHC dataset show the highest proportion of misclassified OPs for both the 39- and 162-bus systems. In contrast, the DT trained on the proposed method dataset performs well, accurately classifying many points situated further from the security boundary. This is particularly notable given that $85\%$ and $79\%$ of the samples in the proposed method dataset fall within the HIC region for the 39- and 162-bus systems, respectively.

Fig. 10 and Fig. 11 further highlight that points with a critical damping ratio exceeding $3\%$ are often misclassified due to feasibility or security issues, and rarely solely due to stability. This observation is particularly interesting, as all datasets maintain a nearly balanced distribution between feasible and infeasible points.

#### *4.4.4 Balancing of Datasets*

Finally, we examine the importance of balancing secure and insecure samples in the dataset. A new dataset for the 162-bus system, generated by the proposed method, initially resulted in an unbalanced distribution between secure and insecure samples. To investigate the impact of this imbalance, we constructed two distinct training sets. The first was created by randomly sampling 10 000 OPs from the unbalanced dataset, preserving the original class proportions. The second was a balanced dataset, where secure and insecure points were sampled in equal proportions by applying stratified sampling based on the security labels. This ensured that both classes were represented equally, preventing potential bias in model training.

The dataset statistics for both the unbalanced and balanced datasets are presented in Fig. 12, while the LHC and importance sampling datasets used for testing are shown in Fig. 4. The results of the DTs trained on these datasets, as detailed in Table V, highlight that dataset balance is crucial for achieving high accuracy. Specifically, the F1-score on the proposed method's test set increased from 0.64 to 0.88, and for the boundary test set, it improved from 0.59 to 0.85. Generalization to other test sets, such as LHC and importance sampling, also improved with the balanced dataset, with F1-scores rising from 0.90 to 0.97 and from 0.69 to 0.87, respectively. These results suggest that a balanced dataset may enhance generalization capabilities and reduce classification bias, leading to potentially more reliable predictions across different test sets.

## 5 Summary of Findings and Discussion

Generally, the proposed methods effectively obtains many samples within the HIC region, as depicted in Fig. 4. Despite this high concentration of OPs centered around the security boundary, Table III and Table IV demonstrate that DTs trained

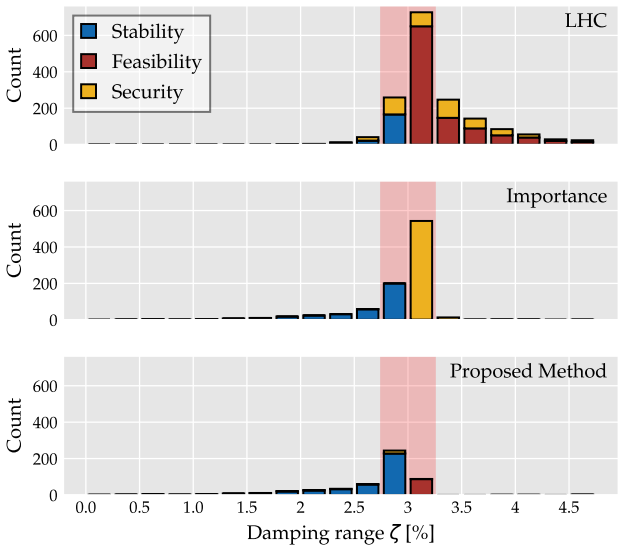


Fig. 10. 39-bus misclassified OPs and the cause for missclassificaton. The HIC region is highlighted in red.

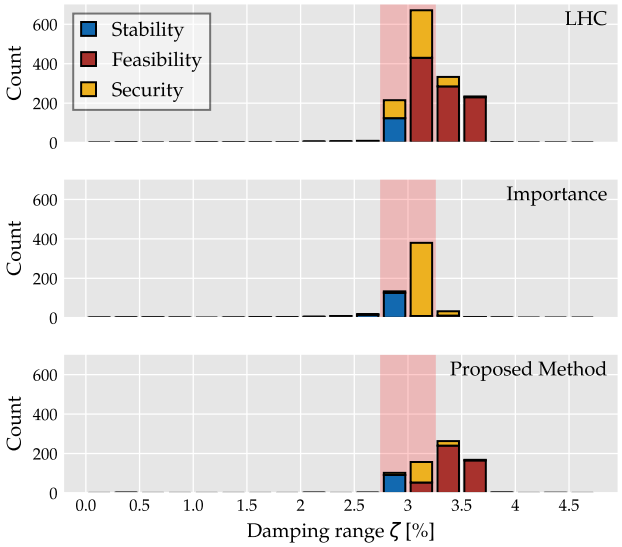


Fig. 11. 162-bus misclassified OPs and the cause for missclassificaton. The HIC region is highlighted in red.

TABLE V
162-BUS SYSTEM F1-SCORES WITH AN UNBALANCED AND A BALANCED DATASET FOR THE PROPOSED METHOD.

| Training \ Testing | LHC | Importance | Proposed Method | Boundary |
|---|---|---|---|---|
| Unbalanced | 0.90 | 0.69 | 0.64 | 0.59 |
| Balanced | 0.97 | 0.87 | 0.88 | 0.85 |

on this dataset exhibit good generalization to test sets generated by other methods, accurately classifying many points far from the security boundary. Moreover, as shown in Fig. 10 and Fig. 11, misclassification frequently occurs near the security boundary, even by a DT trained with a high density of sampled points in the HIC region. These findings underscore the critical

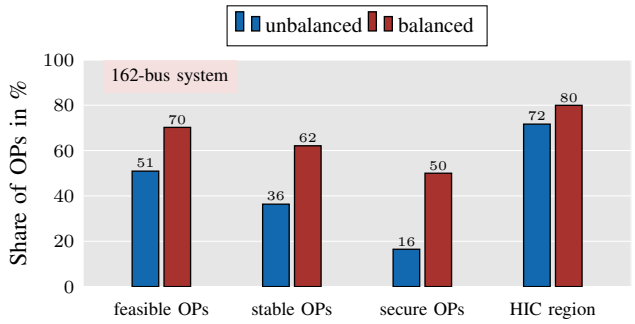


Fig. 12. Share of feasible, stable, secure and HIC region samples for a balanced and unbalanced dataset of the 162-Bus System.

role of the security boundary in data-driven DSA methods and highlight its importance when designing synthetic datasets for such analyses.

Furthermore, Fig. 5 and Fig. 6 reveal that the security boundary is not always sharply defined. This lack of clarity may cause datasets generated by the proposed method and the importance sampling method to exhibit biases toward specific regions, even when the security boundary is broadly distributed. Such bias could explain why the LHC DT outperformed the importance DT on the proposed method's test set, as shown in Table III. Therefore, it is important to consider the security boundary in data-driven DSA, but to accurately describe the whole input space, a comprehensive dataset also needs OPs lying further away from the security boundary.

Additionally, the importance sampling method employed a multivariate normal distribution to sample data. While effective to some extent, this distribution may be insufficient for accurately capturing the complex structure of the security boundary. These results highlight the importance of selecting appropriate probability distributions or copulas [10] to better model and distinguish the marginal distributions that characterize the security boundary.

Finally, the proposed method effectively sampled many points in the HIC region and across the security boundary. However, its performance is influenced by several hyperparameters, including $\kappa_{max}$, $\kappa_{HIC}$, and $\epsilon$, as well as the location of the initialization points for the DW algorithm. These initialization points are crucial for accurately capturing the security boundary and generating a balanced dataset. Their location is guided by an effective mapping of the feasible region, facilitated by the systematic framework involving optimization-based bound tightening and the separating hyperplanes algorithm. Proper hyperparameter tuning and a sufficient number of initialization points are essential for the success of the proposed method in capturing the security boundary. Furthermore, when an unbalanced dataset is generated, a resampling step to balance the dataset can substantially improve the performance of a DT trained on this dataset.

These findings suggest that the proposed method performs well in controlled settings and shows promise for real-world application. Data-driven machine learning proxies for DSA have the potential to enhance situational awareness in control rooms and enable the evaluation of large numbers of OPs, improving decision-making processes for system operators. The proposed method can further advance these techniques by generating training datasets with high information content for proxies, reducing dependence on extensive data collection. A key advantage of this method is its ability to target sampling near the security boundary—an information-rich yet often underrepresented region in datasets generated through traditional methods. To efficiently scale the method for real-world applications, parallelization of the directed walk algorithm will be required, ideally on HPC systems or servers that support parallel computing. This study focuses on synchronous machines, static loads, and static transmission lines. However, the proposed methodology is capable of modeling dynamic load and line models, as well as inverter-based resources. Future work will involve incorporating these dynamic models to capture the faster dynamics characteristic of inverter-dominated systems.

## 6 Conclusion

Data-driven methods for dynamic security assessment (DSA) are gaining traction due to their significant computational speed advantages over conventional time-domain simulations. However, the performance of these tools heavily depends on the quality of the datasets used. Since historical records often lack sufficient coverage of the operating space, synthetic data generation methods become essential.

This paper proposes a novel method to generate synthetic datasets with a high concentration of operating points (OPs) around the security boundary. We demonstrate that datasets with a substantial share of OPs near the security boundary significantly enhance the performance of data-driven DSA tools. Our approach combines optimization-based bound tightening and infeasibility certificates derived from separating hyperplanes [17] to reduce the initial search space. We then employ directed walks [14] to sample numerous OPs close to the security boundary.

We compare our proposed method against two benchmarks: a naive sampling benchmark and an importance sampling benchmark. Case studies on the PGLib-OPF 39- and 162-bus systems show that our method effectively samples a large number of points around the security boundary, capturing it more comprehensively than the benchmarks. Furthermore, decision trees (DTs) trained on the datasets generated by our method achieve a higher accuracy than those trained on benchmark-generated datasets. A closer examination of misclassified OPs reveals that misclassfication often occurs near the security boundary, emphasizing the need to accurately represent this region. Moreover, a comparison between DTs trained on balanced and unbalanced datasets highlights the importance of maintaining dataset balance while incorporating a substantial proportion of boundary samples, both of which are crucial for enhancing model performance.

Future work will focus on developing a unified sampling approach that combines the proposed method's ability to accurately describe the security boundary with copula-based importance sampling to efficiently generate large datasets. Additionally, we plan to evaluate the generated datasets for transient stability and extend the proposed method to other types of stability assessments, such as voltage stability.

## References


[1] P. Panciatici, G. Bareux, and L. Wehenkel, "Operating in the fog: Security management under uncertainty," *IEEE Power and Energy Magazine*, vol. 10, no. 5, pp. 40–49, 2012.

[2] I. Konstantelos, G. Jamgotchian, S. H. Tindemans, P. Duchesne, S. Cole, C. Merckx, G. Strbac, and P. Panciatici, "Implementation of a massively parallel dynamic security assessment platform for large-scale grids," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1417–1426, 2016.

[3] R. Zafar, A. Mahmood, S. Razzaq, W. Ali, U. Naeem, and K. Shehzad, "Prosumer based energy management and sharing in smart grid," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1675–1684, 2018.

[4] L. Wehenkel and M. Pavella, "Decision trees and transient stability of electric power systems," *Automatica*, vol. 27, no. 1, pp. 115–134, 1991.

[5] L. Duchesne, E. Karangelos, and L. Wehenkel, "Recent developments in machine learning for energy systems reliability management," *Proceedings of the IEEE*, vol. 108, no. 9, pp. 1656–1676, 2020.

[6] A. Marot, B. Donnot, K. Chaouache, A. Kelly, Q. Huang, R.-R. Hossain, and J. L. Cremer, "Learning to run a power network with trust," *Electric Power Systems Research*, vol. 212, p. 108487, 2022.

[7] V. Krishnan, J. D. McCalley, S. Henry, and S. Issad, "Efficient database generation for decision tree based power system security assessment," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2319–2327, 2011.

[8] A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, "Generating quality datasets for real-time security assessment: Balancing historically relevant and rare feasible operating conditions," *International Journal of Electrical Power & Energy Systems*, vol. 154, p. 109427, 2023.

[9] F. R. S. Sevilla, Y. Liu, E. Barocio, P. Korba, M. Andrade, F. Bellizio, J. Bos, B. Chaudhuri, H. Chavez, J. Cremer, *et al.*, "State-of-the-art of data collection, analytics, and future needs of transmission utilities worldwide to account for the continuous growth of sensing data," *International journal of electrical power & energy systems*, vol. 137, p. 107772, 2022.

[10] I. Konstantelos, M. Sun, S. H. Tindemans, S. Issad, P. Panciatici, and G. Strbac, "Using vine copulas to generate representative system states for machine learning," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 225–235, 2018.

[11] M. Sun, I. Konstantelos, S. Tindemans, and G. Strbac, "Evaluating composite approaches to modelling high-dimensional stochastic variables in power systems," in *2016 Power Systems Computation Conference (PSCC)*, pp. 1–8, IEEE, 2016.

[12] C. Liu, K. Sun, Z. H. Rather, Z. Chen, C. L. Bak, P. Thøgersen, and P. Lund, "A systematic approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 717–730, 2013.

[13] C. Liu, Z. H. Rather, Z. Chen, C. L. Bak, and P. Thøgersen, "Importance sampling based decision trees for security assessment and the corresponding preventive control schemes: The danish case study," in *2013 IEEE Grenoble Conference*, pp. 1–6, IEEE, 2013.

[14] F. Thams, A. Venzke, R. Eriksson, and S. Chatzivasileiadis, "Efficient database generation for data-driven security assessment of power systems," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 30–41, 2019.

[15] A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, "Split-based sequential sampling for realtime security assessment," *International Journal of Electrical Power & Energy Systems*, vol. 146, p. 108790, 2023.
[16] T. Joswig-Jones, K. Baker, and A. S. Zamzam, "Opf-learn: An open-source framework for creating representative ac optimal power flow datasets," in *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, IEEE, 2022.
[17] A. Venzke, D. K. Molzahn, and S. Chatzivasileiadis, "Efficient creation of datasets for data-driven power system applications," *Electric Power Systems Research*, vol. 190, p. 106614, 2021.
[18] L. Charles, "Final code thesis lola charles." https://github.com/lolachls/Final_Code_Thesis_Lola_Charles, 2024. Accessed: 2024-12-10.
[19] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "Strengthening convex relaxations with bound tightening for power network optimization," in *Principles and Practice of Constraint Programming*, pp. 39–57, Springer International Publishing, 2015.
[20] B. Giraud *et al.*, "Dsa-learn." https://github.com/bastiengiraud/DSA-learn, 2024. Accessed: 2024-11-30.
[21] C. W. G. C4.601, "Review of on-line dynamic security assessment tools and techniques," vol. 4, p. 601, 2007.
[22] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck, "The qc relaxation: A theoretical and computational study on optimal power flow," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3008–3018, 2015.
[23] D. K. Molzahn, I. A. Hiskens, *et al.*, "A survey of relaxations and approximations of the power flow equations," *Foundations and Trends® in Electric Energy Systems*, vol. 4, no. 1-2, pp. 1–221, 2019.
[24] D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of monte carlo methods*. John Wiley & Sons, 2013.
[25] S. Babaeinejadsarookolaee, A. Birchfield, R. D. Christie, C. Coffrin, C. DeMarco, R. Diao, M. Ferris, S. Fliscounakis, S. Greene, R. Huang, *et al.*, "The power grid library for benchmarking ac optimal power flow algorithms," *arXiv preprint arXiv:1908.02788*, 2019.
[26] J. D. Lara, C. Barrows, D. Thom, D. Krishnamurthy, and D. Callaway, "Powersystems. jl—a power system data management package for large scale modeling," *SoftwareX*, vol. 15, p. 100747, 2021.
[27] D. C. Center *et al.*, "DTU Computing Center resources," 2024.
[28] R. Preece and J. V. Milanović, "Efficient estimation of the probability of small-disturbance instability of large uncertain power systems," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1063–1072, 2015.
[29] C. Coffrin, R. Bent, K. Sundar, Y. Ng, and M. Lubin, "Powermodels. jl: An open-source framework for exploring power flow formulations," in *2018 Power Systems Computation Conference (PSCC)*, IEEE, 2018.
[30] I. Dunning, J. Huchette, and M. Lubin, "Jump: A modeling language for mathematical optimization," *SIAM review*, vol. 59, no. 2, pp. 295–320, 2017.
[31] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical programming*, vol. 106, pp. 25–57, 2006.
[32] J. D. Lara, R. Henriquez-Auba, M. Bossart, D. S. Callaway, and C. Barrows, "Powersimulationsdynamics. jl–an open source modeling package for modern power systems with inverter-based resources," *arXiv preprint arXiv:2308.02921*, 2023.
[33] A. Chalkis and V. Fisikopoulos, "volesti: Volume approximation and sampling for convex polytopes in r," *arXiv preprint arXiv:2007.01578*, 2020.
[34] L. Breiman *et al.*, "Classification and regression trees," *Monterey, CA: Wadsworth and Brooks/Cole*, 1984.
[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.