DEVELOPMENT OF A FRAMEWORK FOR POWER SYSTEM OPERATIONAL RISK ASSESSMENT USING PROBABILISTIC POWER FLOW WITH SCADA DATA

Luong Phu Nhan*, Hoang Minh Ha, Le Thi Da Thao

National Power System and Market Operator Company – Central Regional Power System Operator, Vietnam

*Corresponding author: luongphunhan93@gmail.com

(Received: October 29, 2025; Revised: December 04, 2025; Accepted: December 16, 2025)

DOI: 10.31130/ud-jst.2025.23(12).608

Abstract - High penetration of renewable energy poses significant challenges to power system operational security. This paper presents the development of a comprehensive framework for operational risk assessment using a data-driven Probabilistic Power Flow (PPF) approach. The proposed framework automatically identifies the best-fit probability distributions for loads and generation sources, constructs a cluster-based correlation model, and quantifies risks via an enhanced Monte Carlo simulation using Latin Hypercube Sampling (LHS). The framework's effectiveness is validated through a case study on the Vietnamese Central Power System. Results demonstrate that the methodology accurately quantifies operational risks, highlighting the critical importance of correlation modeling. This framework provides a powerful and widely applicable tool to support secure grid operation and planning.

Key words - Probabilistic Power Flow; SCADA; PSS/E; Latin Hypercube Sampling

1. Introduction

The global energy transition is progressing vigorously, oriented towards carbon emission reduction and prioritizing the development of clean energy sources. In line with this trend, Vietnam has been implementing breakthrough policies to promote renewable energy, concretized in Resolution 55-NQ/TW of the Politburo, which outlines the strategic direction for national energy development. Central Vietnam is a region with tremendous potential for solar and wind energy, having become a renewable energy hub of the country, with thousands of megawatts of installed capacity from these sources in recent years.

However, integrating a large proportion of renewable energy sources, which are inherently variable and weather-dependent, has posed unprecedented challenges to power system operation. The uncertainty and randomness of these generation sources, combined with the natural variability of loads, make grid operating states increasingly difficult to forecast. The deterministic power flow (DPF) method, which has traditionally been used, can only assess the system under a few individual scenarios and is no longer sufficient to comprehensively quantify operational risks.

To address these limitations, the Probabilistic Power Flow (PPF) method has been introduced and widely recognized [1-3]. Fundamentally, a PPF problem consists of two main components: (1) probabilistic modeling of uncertain input variables, and (2) computational methods to propagate these uncertainties throughout the system. For the first component, a common practice in previous studies is to pre-assume probability distributions for uncertain factors (e.g., load following a Normal distribution, wind power following a Weibull distribution) [1-2]. This approach may not accurately reflect the actual behavior of each element. For the second component, computational methods are typically divided into two main categories: analytical methods and simulation methods.

To overcome the aforementioned limitations and enhance the accuracy of the analytical model, this study proposes a PPF methodology with the following key differences and contributions:

Development of data-driven probabilistic models: Unlike assuming probability distributions, this study employs an automated process to analyze SCADA operational data, thereby fitting and validating the most appropriate statistical model for each individual load and renewable energy plant.

Development of an automated and efficient framework: The study constructs a complete PPF algorithm using the Python programming language to automate the entire analysis process. In this framework, PSS/E software is used as the core power flow engine, combined with an improved Monte Carlo simulation [5] utilizing high-efficiency Latin Hypercube Sampling (LHS) [6] to enhance convergence speed compared to traditional Monte Carlo simulation.

Integration of realistic correlation modeling: The study develops and integrates a cluster-based correlation model to capture the interdependencies among renewable energy sources, an often overlooked factor that significantly impacts risk assessment results.

2. Proposed PPF analytical framework

2.1. Overview of the PPF analytical framework

The methodology proposed in this study is a comprehensive PPF analytical process, constructed with a modular and automated architecture. The objective of this framework is to transform raw operational data from the SCADA system into meaningful operational risk indices, providing quantitative information for system operators. The entire process is divided into three main stages, illustrated in Figure 1 and detailed in the following sections.

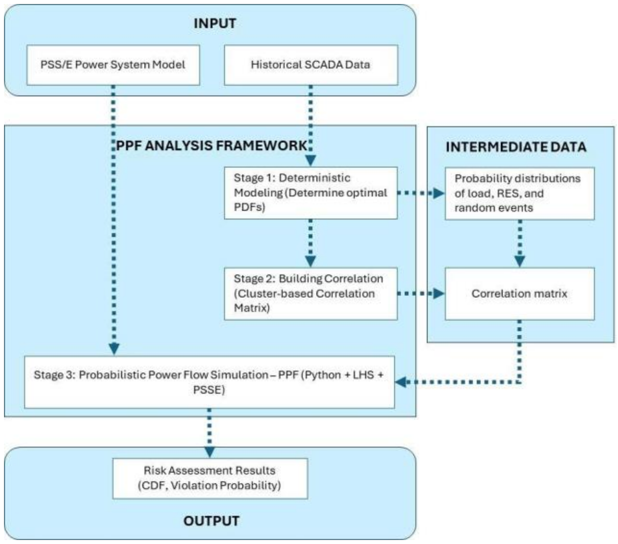

Figure 1. *PPF analytical process*

Stage 1 – Probabilistic Input Modeling: This stage focuses on constructing probabilistic models for uncertain input factors. Instead of traditional assumptions, a dedicated algorithm is developed to automatically process, clean, filter, and analyze historical SCADA data. The most suitable probability density function (PDF) is then identified and validated for each individual load, renewable energy source, and random event. This stage ensures that input models have high fidelity and accurately reflect actual operating characteristics.

Stage 2 – Correlation Structure Modeling: This stage addresses the interdependencies among renewable energy sources. Based on a cluster-based analysis method, a correlation matrix is constructed to describe the spatiotemporal relationships between generation sources within the same geographical area or sharing similar weather characteristics. This correlation structure is a crucial input to ensure the realism of simulation scenarios.

Stage 3 – PPF Simulation and Risk Assessment: This is the core computational stage. An algorithm written in Python orchestrates the entire simulation loop. This algorithm employs the Monte Carlo Simulation method, enhanced with the Latin Hypercube Sampling (LHS) technique to generate correlated operational scenarios. For each scenario, the PSS/E software is called to perform power flow calculations. Results from thousands of iterations are aggregated to construct probability distributions for system state variables (voltage, power flow) and to quantify risk indices.

2.2. Probabilistic modeling from SCADA data

The initial stage of the framework focuses on constructing high-fidelity probabilistic models for uncertain factors, based on historical operational data. This approach replaces pre-assumed distributions, ensuring that the PPF model inputs accurately reflect the characteristics of the actual power system.

2.2.1. Data sources and storage structure

The input data source consists of time-series datasets collected from the SCADA system of the Central Vietnam power grid over a three-year period. To ensure efficient retrieval and processing of large volumes of data - up to millions of records - all data are stored in Parquet file format. This columnar storage format allows for significantly faster querying and loading compared to traditional text formats such as CSV, which is essential for large-scale analysis.

2.2.2. Preprocessing and filtering by "Time Slices"

Raw SCADA data first undergo a preprocessing stage for cleaning, including handling missing values and removing invalid data points (e.g., negative power values for renewable energy plants).

To ensure reliable model inputs, the data processing workflow is fully automated using Python algorithms and the Pandas library for high-performance processing of large datasets. Data cleaning is performed using a multi-layer filter to remove invalid samples:

Layer 1: Handling Missing Values: Remove data points with missing values (NaN/Null) due to communication losses.

Layer 2: Handling Physical Limit Violations: Remove values outside the feasible operating range of equipment, such as negative generation values for power plants.

Layer 3: Outlier Detection: To address outliers caused by measurement errors or transmission noise, the Interquartile Range (IQR) method is applied. For each dataset, the boundaries for filtering outliers are determined as:

$$[Limit_{lower}, Limit_{upper}] = [Q1 - 1.5*IQR, Q3 + 1.5*IQR] \quad (1)$$

Where:

- $Limit_{lower}$: Lower boundary.
- $Limit_{uper}$: Upper boundary.
- Q1: First quartile, the value below which 25% of the data fall.
- Q3: Third quartile, the value below which 75% of the data fall.
- IQR = Q3 – Q1: Interquartile range.

Any data point outside this range is considered an outlier and removed from the dataset. Sensitivity analysis with other methods (such as Z-score or Modified Z-score) shows that the IQR method (with a threshold of 1.5) offers the best balance: it effectively removes measurement noise without discarding valid extreme operational points, which methods based on the normal distribution often miss or wrongly exclude. Quantitative results for processing an operational parameter (active power through bay 233 at the 220kV Hoa Khanh substation) are detailed in Table 1.

Table 1. *Comparison of outlier detection methods*

Method	Threshold	Samples removed	Removal rate (%)
Z-score	3.0	274	0.13%
Modified Z-score	3.5	5522	2.62%
IQR	1.5	3519	1.67%

With the proposed three-layer filter, the proportion of data removed is less than 2% of the total samples. Since power system parameters typically do not fluctuate much over short intervals (with a 5-minute sampling period), removed values are replaced by the previous valid value.

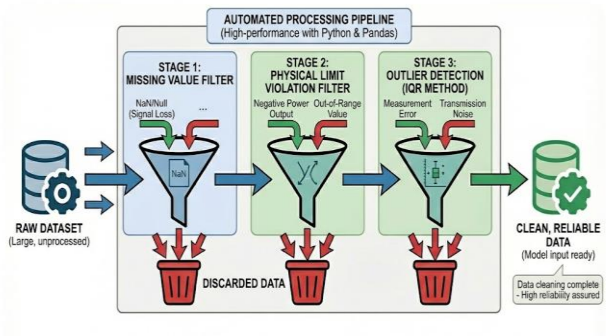

Figure 2. *Multi-layer data cleaning filter*

A core aspect of the proposed methodology is transforming raw time-series data into meaningful datasets representing characteristic operational scenarios of the system. To achieve this, a technique called "time slicing" is applied.

This technique involves filtering and partitioning the SCADA dataset based on specific time frames, such as hours of the day and months of the year. Combining these filters allows for precise extraction of data samples corresponding to critical operational states of the power system.

Analyzing each "time slice" separately enables targeted risk assessment. For example, the "midday off-peak in dry season" scenario is crucial for evaluating overvoltage risk due to low load and high solar generation, while the "evening peak in rainy season" scenario is central to analyzing line overload risk.

2.2.3. Automated fitting and selection of probability distributions

For each dataset corresponding to a characteristic operational scenario, the next stage is to identify the mathematical model (Probability Density Function - PDF) that best describes the data distribution. This study employs a two-step automated model selection process to ensure statistical accuracy and reliability.

Step 1: Goodness-of-Fit Test. Each candidate distribution (e.g., Student, Normal, Weibull, Beta, Gamma, etc.) is fitted to the actual data. The Kolmogorov-Smirnov (K-S) test [7] is then applied to assess whether the data follow the theoretical distribution. The K-S test is non-parametric, with the following hypotheses:

• H_0 (Null hypothesis): The data follow the specified distribution.

• H_1 (Alternative hypothesis): The data do not follow the specified distribution.

A p-value is calculated from the test. With a pre-selected significance level α (typically $\alpha = 0.05$), the decision rule is:

• If p-value < 0.05, H_0 is rejected. The distribution is considered unsuitable and excluded.

• If p-value > 0.05, there is insufficient evidence to reject H_0. The distribution is considered statistically suitable and proceeds to the next selection step.

Step 2: Model Selection by AIC Criterion. After Step 1 filters out unsuitable distributions, several candidate models may remain. To select the best among them, the Akaike Information Criterion (AIC) [8] is used. AIC is a statistical indicator that compares the relative quality of models by balancing goodness of fit and model complexity. The general formula for AIC is:

$$AIC = 2k - 2\ln(L) \tag{2}$$

Where:

• k: Number of estimated parameters in the model, penalizing model complexity. For example, the Normal distribution includes mean and standard deviation.

• L: Maximum likelihood of the model. The log-likelihood value represents the goodness of fit; the higher the value, the better the fit.

The final result of this stage is a set of characteristic parameters for the selected distribution for each load and renewable energy source, corresponding to each operational scenario. These parameters are used as inputs for the PPF simulation stage, ensuring accuracy and practicality throughout the analysis process.

Figures 3 and 4 illustrate a system parameter (reactive power at the 220kV Tuy Hoa substation) extracted from SCADA, represented as the empirical probability distribution and the fitted theoretical Student-t distribution (with p-value > 0.05 and lowest AIC in Table 2). The estimated parameters for the Student-t distribution are: df: 6.3606, loc: 1.1389, scale: 4.2107.

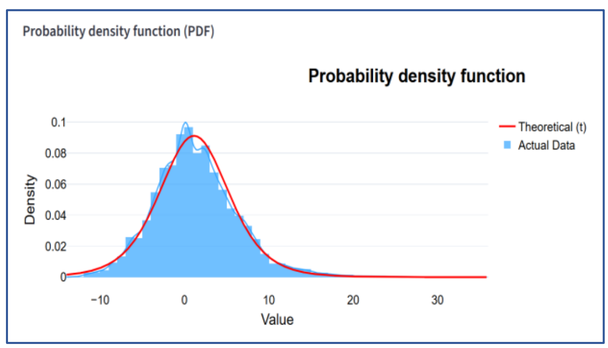

Figure 3. *PDF Plot of SCADA actual data and fitted distribution*

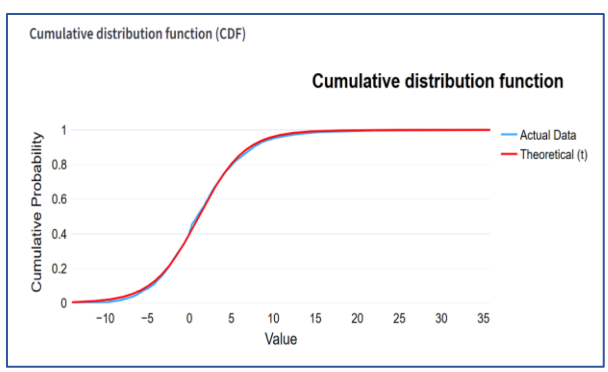

Figure 4. *Cumulative distribution function (CDF) plots of actual SCADA data and estimated*

***Table 2.** p-value and AIC for each distribution type*

Distribution	AIC	p-value
student	967	0.10
lognorm	969	0.48
gamma	976	0.38
beta	978	0.38
norm	1047	0.01

2.2.4. Modeling discrete events

In addition to continuous uncertainties such as load and renewable generation, a comprehensive operational risk assessment must consider low-probability but high-impact discrete events, typically random failures of system components. In this study, failures of critical elements such as transmission lines and generator units are modeled as discrete random variables.

Each component (generator i or line j) is modeled using a Bernoulli distribution with two possible states:

1. In-service (normal operation)
2. Outage (failure)

The probability of each state is determined by a key statistical parameter, the Forced Outage Rate (FOR). FOR is defined as the probability that a component is unavailable at any given time due to intrinsic failures. It is calculated from historical operation and maintenance data as follows:

$$FOR = \frac{\sum T_{Outage}}{\sum T_{Outage} + \sum T_{In-service}} \tag{3}$$

Where:

- $\sum T_{Outage}$: Total time the device is out of service.
- $\sum T_{In-service}$: Total time the device is in normal operation.

By integrating these models, the Monte Carlo simulation algorithm (Stage 3) is enhanced as follows: In each simulation iteration, in addition to random sampling for continuous variables (load, renewable generation), the states of lines and generators are also randomly sampled based on their FOR. For example, with FOR = 0.01, in each iteration, a random number between [0, 1] is generated; if the number is less than 0.01, the component is simulated in the outage state for that iteration.

2.3. Construction of correlation structure

Accurately modeling individual probability distributions (as described in Stage 1) is necessary but not sufficient for a realistic PPF problem. In actual operation, renewable energy sources, especially solar and wind power plants, do not operate entirely independently. Their output is often correlated due to shared exposure to large-scale meteorological factors such as cloud cover, wind direction, and solar irradiance within the same region [9].

Ignoring correlations and assuming independent uncertainties can dangerously underestimate operational risks. For example, an uncorrelated model may fail to capture scenarios where widespread cloud cover or calm winds simultaneously cause large power fluctuations across the system. Therefore, constructing a realistic correlation model is a core stage in the proposed analytical framework.

2.3.1. Cluster-based correlation analysis method

To capture the most significant correlations without making the problem overly complex, this study adopts a cluster-based correlation analysis method. The main idea is to group loads and renewable sources that are closely related geographically or meteorologically into the same "cluster", with reasonable assumptions:

- Loads and generation sources within a cluster are correlated.
- Loads and generation sources in different clusters are assumed independent.

For example, all wind farms in Quang Tri province can be grouped into one cluster, and all solar plants in Gia Lai region into another.

2.3.2. Calculation of correlation matrices

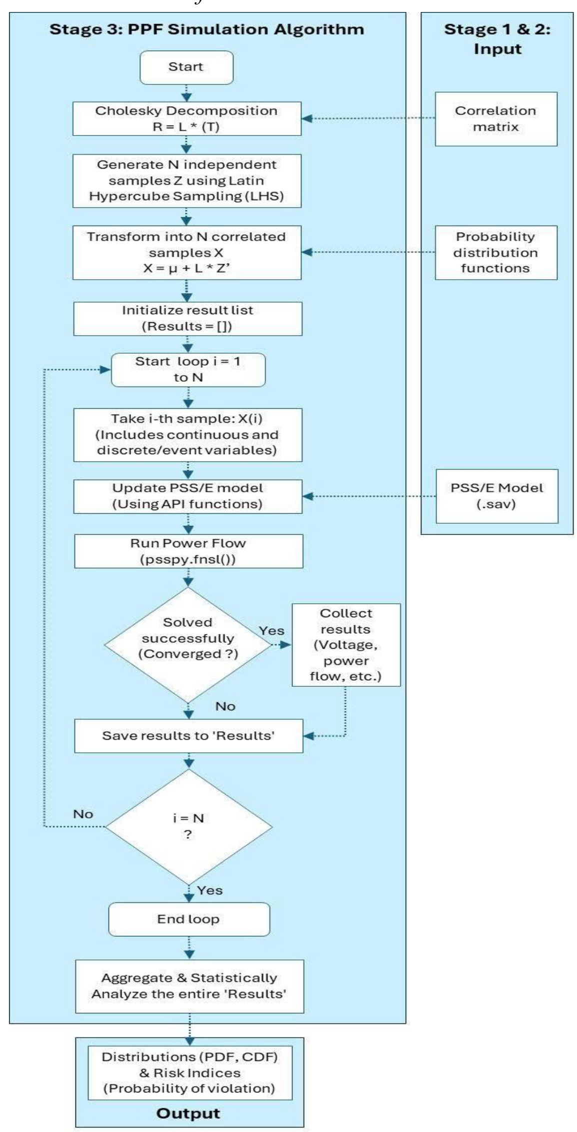

***Figure 5.** PPF simulation flowchart*

For each defined cluster, the internal correlation matrix is calculated from SCADA time-series data. This study uses the Spearman rank correlation coefficient (r_s) instead of the Pearson coefficient, as the latter only measures linear correlation and is sensitive to outliers and non-normal data. In contrast, operational data for renewables and loads often do not follow normal distributions (as shown in Stage 1), and their relationships are not always linear. Spearman's coefficient is non-parametric, operating on the ranks of data rather than their actual values, making it robust to outliers and capable of capturing monotonic relationships - i.e., when one variable tends to increase, so does the other (or vice versa), regardless of linearity.

The Spearman correlation coefficient rs between two random variables X_i and X_j (which can be source-source, load-load, or source-load) is calculated by applying the Pearson formula to their ranked data:

$$r_s = \rho_{rg(X_i),rg(X_j)} = \frac{cov(rg(X_i),rg(X_j))}{\sigma_{rg(X_i)}\sigma_{rg(X_j)}} \quad (4)$$

Where:

- X_i and X_j: Time series of random variables;
- $rg(X_i)$ and $rg(X_j)$: Ranked data of X_i and X_j;
- $cov(rg(X_i), rg(X_j))$: Covariance of ranked values;
- $\sigma_{rg(X_i)}, \sigma_{rg(X_j)}$: Standard deviations of ranked values.

After calculating Spearman correlation matrices for each cluster, they are combined into an overall block-diagonal correlation matrix. This matrix is a key input, determining the dependency structure of random samples generated in Stage 3, ensuring simulation scenarios accurately reflect the complex relationships among system elements.

2.4. PPF simulation and risk assessment

The probabilistic models and correlation structures built are integrated with the power system model to quantify operational risks. This process is executed via an automated simulation algorithm, integrating Python programming with the specialized power system analysis software PSS/E by Siemens.

2.4.1. Monte Carlo Simulation with Latin Hypercube Sampling (LHS)

The selected computational method is Monte Carlo Simulation (MCS), a powerful and flexible approach capable of handling complex problems with multiple random variables and arbitrary distributions. To improve computational efficiency and reduce the required number of iterations, this study applies an advanced sampling technique: Latin Hypercube Sampling (LHS).

Unlike simple random sampling (which may produce uneven sample clusters), LHS is a stratified sampling method. It divides the cumulative distribution function (CDF) of each random variable into N intervals with equal probability, ensuring each interval is sampled exactly once. This results in random sample sets that are more evenly distributed across the entire probability space, thereby significantly accelerating the convergence of output statistics.

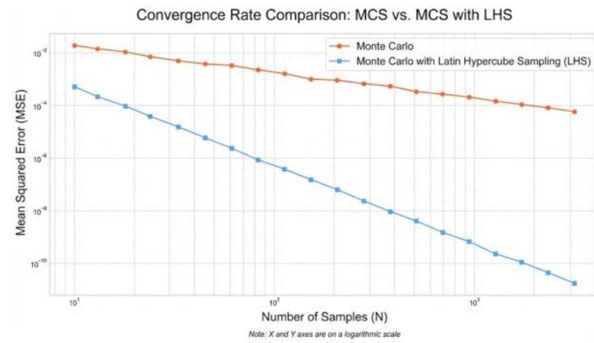

Figure 6. *Comparison of convergence speed between MCS and MCS with LHS*

2.4.2. Generation of correlated random samples

To ensure simulated scenarios accurately reflect the dependency structure built in Stage 2, random samples must be correlated. This is achieved using the mathematical technique of Cholesky decomposition [10].

The correlation matrix R (output of Stage 3) is decomposed into the product of a lower triangular matrix L and its transpose L^T:

$$R = L.\ L^T \quad (5)$$

An independent standard normal random vector Z_{ind} (generated by LHS) is transformed into a correlated standard normal vector Z_{corr} via matrix multiplication:

$$Z_{corr} = L\cdot Z_{ind} \quad (6)$$

Finally, actual random variables X_i (generation, load) are determined by applying the inverse cumulative distribution function (Inverse CDF) of each distribution defined in Stage 1 to Z_{corr}:

$$X_i = F_i^{-1}(\Phi(Z_{corr}) \quad (7)$$

Where:

- Φ: CDF of the standard normal distribution;
- F_i^{-1}: Inverse CDF of the actual probability distribution (e.g., Student-t, Beta) for element i.

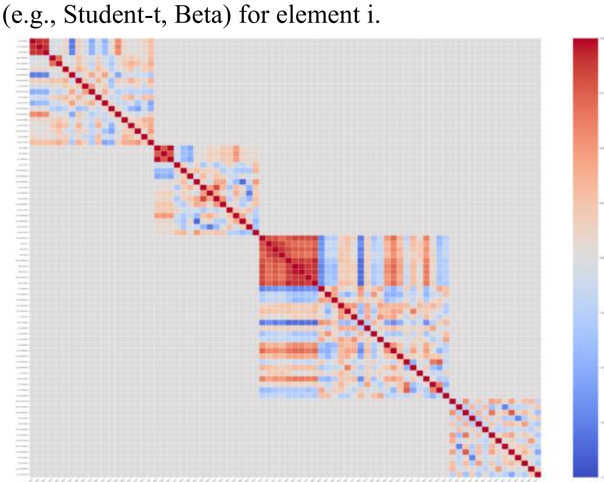

Figure 7. *Heatmap of the correlation matrix*

2.4.3. Automation loop and PSS/E integration

The entire simulation process is controlled by a Python script executing an N-iteration loop. In each iteration i = 1, ..., N, the following stages are performed automatically:

1. Scenario generation: Generate a set of correlated random samples X_i for all continuous variables (load, renewable generation) and sample discrete states (generator/line outages) based on their FOR.

2. Model Update: Using PSS/E API functions, the Python script updates power values from sample set X_i into the power system model. If a failure is sampled, the corresponding element is set to “out of service”.

3. Computation Execution: Call the PSS/E power flow solution function (e.g., psspy.fnsl()) to calculate the system operating state for the current scenario [11-12].

4. Result Collection: Upon successful computation, the Python script retrieves key output results (bus voltages, power flows, branch loadings) and stores them.

2.4.4. Statistical analysis and risk assessment

Upon completion of the simulation loop, the result is a large dataset containing N possible operating states of the system. From this dataset, empirical probability distributions (as histograms or cumulative distribution functions - CDFs) for output variables are constructed. Operational risk indices are directly quantified, for example:

- Probability of voltage limit violation: $P(V_{bus}<V_{min}$ or $V_{bus}<V_{max})$;
- Probability of line overload: $P(\%Loading_{line} >100\%)$.

These indices provide a quantitative and clear view of system security under uncertainty.

3. Application to the 220kV Central Vietnam power system

3.1. System Under Study

The selected system for applying and validating the analytical framework is the Central Vietnam power system, focusing on the 220kV voltage level. This system is chosen as a representative case study for the following reasons:

- High penetration of renewable energy: This region has experienced a rapid expansion of solar and wind power sources in recent years, creating a complex operational environment with high uncertainty - an ideal scenario for testing a probabilistic risk assessment method.
- Representativeness: The system encompasses all essential components of a modern power grid, including conventional power plants, large-scale renewable energy farms, and diverse load centers (industrial, residential, etc.), enabling a comprehensive and generalizable risk evaluation.

3.2. Model and data

Power system model: The simulation model of the 220kV Central Vietnam power system is developed using the specialized PSS/E software. The model includes complete grid topology parameters (buses, branches, transformers) and equipment characteristics (generators, loads). This foundational model serves as the physical object to which the uncertainty scenarios generated by the analytical framework are applied.

Analytical data: As described in the methodology, the input dataset consists of three years of SCADA operational data, stored in Parquet format. This dataset provides time series of active power (P) and reactive power (Q) for hundreds of measurement points, including loads at 220kV and 110kV buses, as well as connection points for major solar and wind plants in the region.

3.3. Simulation and result analysis

The scenario analyzed is the midday off-peak during the dry season, characterized by low system load while solar generation can reach its maximum. The primary objective of analyzing this scenario is to assess the risk of overload on transmission lines near renewable energy centers.

The element under consideration is the 220kV Quy Nhon – Tuy Hoa transmission line, which is responsible for supplying electricity to the eastern areas of Gia Lai and Dak Lak provinces. Additionally, this line plays a crucial role in evacuating renewable energy (especially solar power) from the region. The loading level of the line is represented by probability density function (PDF) and cumulative distribution function (CDF) curves.

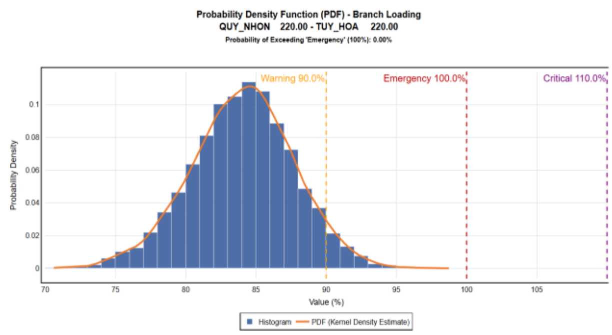

Figure 8. *PDF curve of loading level for the 220kV Quy Nhon - Tuy Hoa transmission line*

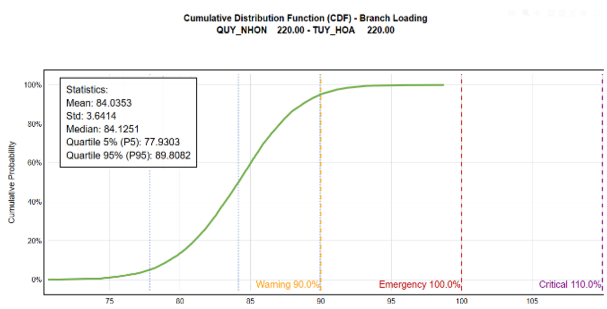

Figure 9. *CDF curve of loading level for the 220kV Quy Nhon - Tuy Hoa transmission line*

Analysis of the PDF and CDF curves yields the following observations:

- **Operational trend:** The peak of the PDF curve indicates that the most probable loading level for this line is in the range of 80–85%. This represents the typical operating state during midday off-peak in the dry season when solar generation is high.
- **Degree of fluctuation:** The distribution curve spans from about 70% to nearly 98%, indicating significant variation in power flow on the line. This variability accurately reflects the impact of renewable energy sources and load.

- **Asymmetry:** The distribution is slightly skewed to the right, suggesting a higher likelihood of high loading conditions compared to low loading cases.
- **Risk assessment:** The probability that the line exceeds the warning threshold (loading level at 90% of rated capacity or higher) is 4.57%.

4. Conclusion

This paper has presented the development and validation of a comprehensive analytical framework for power system operational risk assessment using probabilistic power flow based on actual data. The proposed framework addresses the limitations of traditional methods by integrating an automated process, from raw SCADA data processing to systematic quantification of risk indices. The core methodological contributions include the automated identification of the most appropriate statistical models for each uncertain factor, construction of realistic cluster-based correlation structures, and the application of Monte Carlo simulation combined with Latin Hypercube Sampling (LHS).

Through the case study on the 220kV Central Vietnam power system model, the effectiveness of the analytical framework has been clearly demonstrated. The results quantitatively assessed risks under characteristic operational scenarios. The study also emphasizes the critical importance of modeling the correlation between renewable energy sources and loads to avoid underestimating system risks.

This analytical framework has practical application potential for entities such as Power System Operation Centers and Power Companies. Thanks to its open architecture, Python-based platform, and industry-standard PSS/E computational engine, it can be readily deployed as a decision support system. In actual operation, this tool can effectively assist short-term operational planning by providing early probabilistic risk warnings, as well as support medium- and long-term system security assessments. The adoption of this framework marks a significant shift from “deterministic” operational thinking to “probabilistic risk management”, aligning with the increasingly uncertain nature of modern power systems.

Future research and development directions are proposed to focus on expanding the scope of application. An important direction is to integrate statistical models from historical SCADA data with scenarios of load growth and generation development. This approach will enable a transition from current operational risk assessment to near-future risk evaluation, providing quantitative information for short-term power grid planning.

REFERENCES

[1] A. İskenderoğlu and Ö. Gül, “Probabilistic Power Flow Analysis of Power Systems with PV and Wind Using Monte Carlo Method”, in *Proc. 10th Int. Conf. Electr. Electron. Eng. (ICEEE)*, 2023.

[2] Z. Ren, K. Wang, W. Li, L. Jin, and Y. Dai, "Probabilistic Power Flow Analysis of Power Systems Incorporating Tidal Current Generation", *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1195-1203, Jul. 2017.

[3] D. D. Le, V. D. Ngo, T. A. N. Nguyen, and V. K. Huynh, “Study on uncertainty of operating parameters to calculate and analyze working condition of power systems", *Journal of Science and Technology - The University of Danang*, vol. 5, no. 90, pp. 34-38, May 2015.

[4] D. S. Kumar, H. Quan, K. Y. Wen, and D. Srinivasan, "Probabilistic risk and severity analysis of power systems with high penetration of photovoltaics," *Solar Energy*, vol. 230, pp. 1156-1164, Dec. 2021.

[5] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, 2nd Ed. Wiley, Hoboken, NJ, USA, 2008.

[6] J. C. Helton and F. J. Davis, “Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems”, *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23-69, Jul. 2003.

[7] V. W. Berger and Y. Zhou, “Kolmogorov–Smirnov Test: Overview”, in *Encyclopedia of Statistics in Behavioral Science*. Wiley, 2014.

[8] J. E. Cavanaugh, “The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements”, *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 11, no. 3, Art. no. e1460, May/Jun. 2019.

[9] J. Zang, B. M. Hodge, and A. Florita, “Investigating the Correlation Between Wind and Solar Power Forecast Errors in the Western Interconnection”, in *Proc. ASME 7th Int. Conf. Energy Sustain. and 11th Fuel Cell Sci., Eng., Technol. Conf.*, USA, Jul. 2013.

[10] L. Li, Y. Fan, X. Su, and G. Qiu, “Probabilistic Load Flow Calculation of Power System Integrated with Wind Farm Based on Kriging Model”, Energy Eng., vol. 118, no. 2, pp. 565-580, 2021.

[11] Siemens Power Technologies International, *Program Operation Manual PSS®E 36.1.0*. USA: Siemens PTI, Jul. 2024.

[12] Siemens Power Technologies International, *Application Program Interface PSS®E 36.1.0*. USA: Siemens PTI, Jul. 2024.