

Master Thesis

Probabilistic Framework for Assessing Cascading Failures in Power Systems

ET4300: Master Thesis

Ananth Rajagopal

Master Thesis

Probabilistic Framework for Assessing Cascading Failures in Power Systems

by

Ananth Rajagopal- 5912431

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Tuesday the 25th of March, 2025 at 15:00.

Thesis Committee: dr. J. L. Cremer, TU Delft, Chair and Supervisor

dr. M. Ghaffarian Niasar, TU Delft, Core Member

M. Karaçelebi, TU Delft, Daily Supervisor

Cover: Inside the ISO New England power grid control room - cnet.com
by Martin LaMonica

Style: TU Delft Report Style, with modifications by Daan Zwaneveld



Ideas alone have little value.

An innovation's importance lies in its practical implementation."

- Werner von Siemens, January 27, 1865

Acknowledgment

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Jochen Cremer, for his invaluable guidance, encouragement, and insightful feedback throughout my thesis. Your expertise and support have been instrumental in shaping this work and in my growth as a researcher. I extend my heartfelt thanks to Mert Karaçelebi for his exceptional dedication and daily guidance. Your patience, practical insights, and unwavering support have helped me navigate challenges and overcome obstacles at every step. I truly appreciate the time and effort you devoted to helping and am also sincerely grateful for your thoughtful advice and constructive suggestions, which greatly enriched the quality of this thesis.

To my family, I owe endless gratitude for their unwavering support, encouragement, and belief in me throughout this journey. Your love and sacrifices have been the foundation of my achievements. A special thank you to my friends, colleagues, and peers for their constant motivation, camaraderie, and shared moments of both challenge and celebration. Your support has made this experience truly memorable. Finally, I thank everyone who, directly or indirectly, contributed to this work. Your support has made this accomplishment possible.

Abstract

Cascading failures in power networks pose a significant threat, capable of escalating from isolated line outages to extensive blackouts with severe economic and societal impacts. The topic presents a probabilistic framework designed to assess and compute the risk of cascading failures within power network and rank various cascade contingencies, utilizing the IEEE39 bus 10-machine New England Power System. Using DigSILENT PowerFactory, a detailed contingency analysis was conducted, focusing on line loading conditions following faults as a starting point. This approach operates under the foundational assumption that cascading failures can be effectively modelled through the sequential analysis of line contingencies. Central to this framework is assessing topological vulnerabilities in the grid and determining frequently occurring outage patterns called probabilistic contingency motifs (PCMs). By analyzing the characteristics of the grid, an impact metric is proposed using short-circuit analysis, electrical distance and LODFs for the identified cascade contingency. The outage probabilities and the proposed impacts are used to compute the risks of cascade cases.

After ranking based on their associated risks, high-risk contingencies are dynamically simulated through time-domain simulations to assess dynamic security. This simulation approach validates the model's predictions and ensures that ranked contingencies reflect realistic cascades. Use cases can show that the model will enable Transmission System Operators (TSOs) to implement preventive measures and simulate corrective actions effectively. By systematically identifying and incorporating PCMs and leveraging grid topology, the model estimates the likelihood and impact of cascading events and delivers actionable insights for improving power system robustness. Future work will expand the framework's scalability to larger and more complex networks and integrate real-time data streams to facilitate dynamic risk assessment and proactive mitigation strategies.

Contents

1	Introduction	1
1.1	Background and Context	1
1.2	Literature Review on Cascading Failures and Protection	3
1.2.1	Failure Escalation in Power Systems: Examples and Mechanisms	4
1.2.2	Cascading Failure Dynamics: Modeling and Historical Perspectives	6
1.2.3	Cascade Protection Mechanisms: Preventive and Corrective Controls	21
1.2.4	Motifs and Probabilistic Frameworks for Risk Assessment	23
1.3	Research Direction and Contribution	25
1.4	Research Questions and Structure of the Thesis	27
2	Theory	28
2.1	Cascading Outages in Power Systems	28
2.1.1	Overview and Mechanisms of Cascading	29
2.1.2	$N - k$ Contingency Analysis and its importance	31
2.1.3	Contingency and Power Flow Analysis procedure	33
2.2	Probabilistic Modeling of $N - k$ Outages	37
2.2.1	Subgraph representation of line outages	38
2.2.2	Representation of subgraph patterns	39
2.2.3	Identification of an $N - k$ cascade contingency motif	41
2.2.4	Probabilities of multiple line outages	43
2.3	Risk Assessment of Cascading Events	46
2.3.1	Overview of Risk in Power Systems	47
2.3.2	Calculating Impact Metrics	47
3	Methodology	48
3.1	PCM for $N - k$ critical cascade risk rankings	48
3.2	Network structure based probability analysis workflow	48
3.2.1	Performing contingency analysis	50
3.2.2	Pattern Classification and identification of PCMs	51
3.2.3	Contingency enumeration and probability of line outages	53
3.3	Critical line impact assessment	56

3.3.1	Calculation of impact for risk	56
3.3.2	Generator short circuit contributions ($I_{SC,gf}$)	57
3.3.3	Weighting factors, LODF sensitivities (w_{gf})	58
3.3.4	Iteratively building the Z_{bus} matrix	59
3.3.5	Electrical distance between fault location and generator (D_{gf})	62
3.4	Risk rankings and verification	63
3.4.1	Verification of risks and associated contingency rankings	64
3.4.2	Analyzing operator decision making	66
4	Simulation and Results	67
4.1	Simulating cascades and obtaining dynamic impact	67
4.1.1	Project activation and power system dispatch	68
4.1.2	Sensitivity Analysis	69
4.1.3	Short Circuit analysis in PowerFactory and proposed impact	70
4.1.4	RMS Simulations in PowerFactory and maximum rotor angle impact	71
4.1.5	Cascade load flow and line overloading impact	72
4.2	Case Study Setup	73
4.2.1	Contingency analysis for critical line sets	73
4.2.2	Probabilities of critical line sets	75
4.3	Case Studies, simulations and results	78
4.3.1	CS1: Impact Metric	78
4.3.2	CS2: Risk calculation and analysis	80
4.3.3	CS3: Effects on dispatch on system risk	87
5	Discussion and Limitation	91
5.1	Interpretation of Results	91
5.1.1	Probability of a cascade contingency scenario	91
5.1.2	Effectiveness of dynamic impact metric	92
5.1.3	Implications of risk calculation and ranking	93
5.2	Answers to the Research Questions	94
5.3	Limitations	96
6	Conclusion	98
6.1	Recommendations	99
	References	100
A	Electrical Distance D_{gf} from Z_{bus}	104
B	Impact weight w_{gf} from $LODFs$	108

List of Figures

1.1	The impact of line removal on l_c vs RG, ρ , and μ_{N-1} in IEEE 30 power system. The grid is attacked based on betweenness centrality. [3]	2
1.2	Description of failure in terms of state transitions [6]	4
1.3	IEEE 39-bus coherent grouping of generators and islanding boundaries applied with (a) DBSCAN algorithm [11] and (b) Spectral Clustering Approach [12]	7
1.4	Flowchart of Cascading Failure Process [13]	8
1.5	Flowchart of cascading failure analysis based on different phases before, during, and after cascade. [10]	9
1.6	Representative sample of node-dependent damage for different tolerance values [14]	11
1.7	Leading causes for blackout events and component failures in continental and national datasets [15]	12
1.8	Critical loading threshold and spectral graph metrics for the IEEE-118 with test systems attacked with a betweenness centrality strategy [3]	14
1.9	The CCDF of lost load for IEEE-118 system with and without cascading [17]	15
1.10	Structure of LSTM for fault current and direction determination [19]	16
1.11	Comparison of cascading classes using various models and performance metrics [13]	17
1.12	Average blackout size vs. maximum capacity of initial line failures for different grid-loading ratios and Cascade-stability probability vs. Failures in the system [23]	18
1.13	With the bypass algorithm, weak nodes (red) are reinforced by creating triangles with the two closest neighbors (green) or doubling edges between neighboring weak nodes (blue links) as an example (left) and overall Europe (right) [26]	20
1.14	Critical loading threshold and spectral metrics for the IEEE 118 with node significance-based attack [3]	21
1.15	Comparison of NCC and MSCC models in the IEEE-39 and IEEE-118 bus systems [28]	22
2.1	Sequence of cascade transition	29
2.2	An event tree explaining cascade phases [32]	30
2.3	Cascade phase transition	30
2.4	Deterministic versus probabilistic reliability analysis in power systems [33]	32
2.5	IEEE 39-bus network with highlighted subgraphs illustrating examples of $N - 2$, $N - 3$ and $N4$ line outages (layout is not geographic)	41

3.1	PCM Workflow	49
3.2	Flow of Impact \mathcal{I}_f calculation using critical lines, SC-simulations and Z_{bus}	56
3.3	Verification after selecting critical contingency sets from risk rankings	64
4.1	Contingency scenarios with increasing k in N- k cascades for IEEE39 bus network	73
4.2	Heatmap of line loading under $N - 2$ contingency for IEEE39 network	74
4.3	Complementary Cumulative Distribution Function (CCDF) Plot showing diameter distribution of cascade sets	77
4.4	An example of probability of a specific line set from the cascade contingency list	78
4.5	Correlation plot between proposed impact \mathcal{I} versus verification impact $\Delta\theta_f$	79
4.6	Computational time comparison between \mathcal{I} versus $\Delta\theta_f$	80
4.7	Confusion matrix comparison of Contingency Selection Methods	84
4.8	Comparative Performance of Selection Methods for High-Impact Contingencies	85
4.9	Distribution of Spearman Rank Correlation coefficient between \mathcal{R} -based metric and $\Delta\theta$ impact based metric for random and top-10 ranked contingencies	86
4.10	Parallel Coordinate Plot: Generator Dispatch vs Risk Category	88
4.11	PCA-Based Contour Plot: Risk Across Dispatch Scenarios	89
4.12	Contour Plot of Load 03 and 07 in IEEE39 bus network: Risk Across Dispatch Scenarios	89
4.13	Feature Importance Plot: Key Dispatch Factors Influencing Risk	90

List of Tables

2.1	PATTERNS FOR $N - k$ CONTINGENCY (LINE OUTAGE SUBGRAPHS)	40
4.1	CONTINGENCY SAMPLES FOR CASCADE ORDERS IN IEEE39 BUS SYSTEM	73
4.2	PROBABILITIES OF PATTERNS IN IEEE39 SYSTEM FOR IDENTIFIED CONTINGEN- CIES	75
4.3	DISTRIBUTION OF PATTERNS (EMPIRICAL PROBABILITIES) $P(S_{k,i} k)$ FOR IEEE39 NETWORK	76
4.4	PROBABILITY $P(d S_{k,i})$ FOR DISCONNECTED SUBGRAPHS FOR DIFFERENT DI- AMETERS IN $S_{k,i}$ for IEEE39 NETWORK	76
4.5	NUMBER OF DISTINCT SUBGRAPHS WITH DIFFERENT DIAMETERS IN $S_{k,i}$ for IEEE-39 BUS	76
4.6	PROBABILITY OF OUTAGES ($P(s_{k,i})$) WITH DIFFERENT PATTERNS AND DIAME- TERS (d)	77
4.7	PERFORMANCE METRICS OF CONTINGENCY SELECTION	83
4.8	COMPUTATION TIME PER SCENARIO ($N - 2$) FOR DIFFERENT METHODS	85

Nomenclature

Abbreviation	Definition
PCM	Probabilistic Contingency Motif
OPF	Optimal power Flow
COI	Center of Inertia
LODF	Line Outage Distribution Factor
PLF	Probabilistic Load Flow
TSO	Transmission System Operator
DSO	Distribution System Operators
PDF	Probability Distribution Function
KVL	Kirchoff's Voltage Law
KCL	Kirchoff's Current Law
ELNS	Expected Load Not Served
CCDF	Complementary Cumulative Distribution Function
GMM	Gaussian Mixture Method
LSTM	Long Short-Term Memory
NDS	Networked Distribution Systems
RDS	Radial Distribution System
DNN	Deep Neural Networks
CDF	Cumulative Distribution Function
CTS	Corrective Transmission Switching
ELM	Extreme Learning Machine
OTS	optimal transmission switching
MSCC	Multi-Step Corrective Control
NCC	Non-Recurring Corrective Control
BPA	Bonneville Power Administration
NYISO	New York Independent System Operator
NR	Newton-Raphson

1

Introduction

1.1. Background and Context

Electric power grids are critical infrastructures in modern society, where disruptions can lead to severe societal and economic consequences. Cascading failures in these grids, wherein initial disturbances propagate through dependent components, are among the primary causes of widespread blackouts. This phenomenon has been responsible for some of the largest blackouts in history, such as the 2003 Northeast blackout in the United States and Canada and the 2012 blackout in India, which affected hundreds of millions of people [1]. These events highlight the catastrophic social and economic impacts of cascading failures, including disrupted critical infrastructure, billions in economic losses, and compromised public safety. These failures typically originate from an initial disturbance or fault in electrical components, compounded by human operator errors. Systematically identifying the root causes of cascading failures in power systems continues to be a significant challenge.

These failures occur as a sequence of dependent events where the outage of one element leads to redistributed power flows, overloading other components, and triggering subsequent failures. Renewable energy sources, while essential for sustainable development, introduce new challenges, such as reduced system inertia, increased variability in power supply, and reliance on inverter-based distributed energy resources (DERs). These factors increase the system's vulnerability to cascades, as even minor disturbances can propagate rapidly through a highly interconnected grid. Addressing this requires advanced models for grid-forming technologies and improved protection systems to prevent cascading events. The study of cascading failures is, therefore, fundamental to safeguarding the stability and reliability of critical infrastructure in a rapidly changing energy landscape. [2]

The propagation of cascading failures in power grids highlights a critical vulnerability, where localized disruptions escalate into widespread outages due to the interconnected nature of the grid. As observed in Figure 1.1 from [3], the removal of even a single line lowers the critical loading threshold of the system (l_c) from 87% to 86%, showing the grid's increasing susceptibility to cascading failures. With additional line removals, l_c decreases further, indicating earlier phase transitions and a higher risk of systemic collapse.

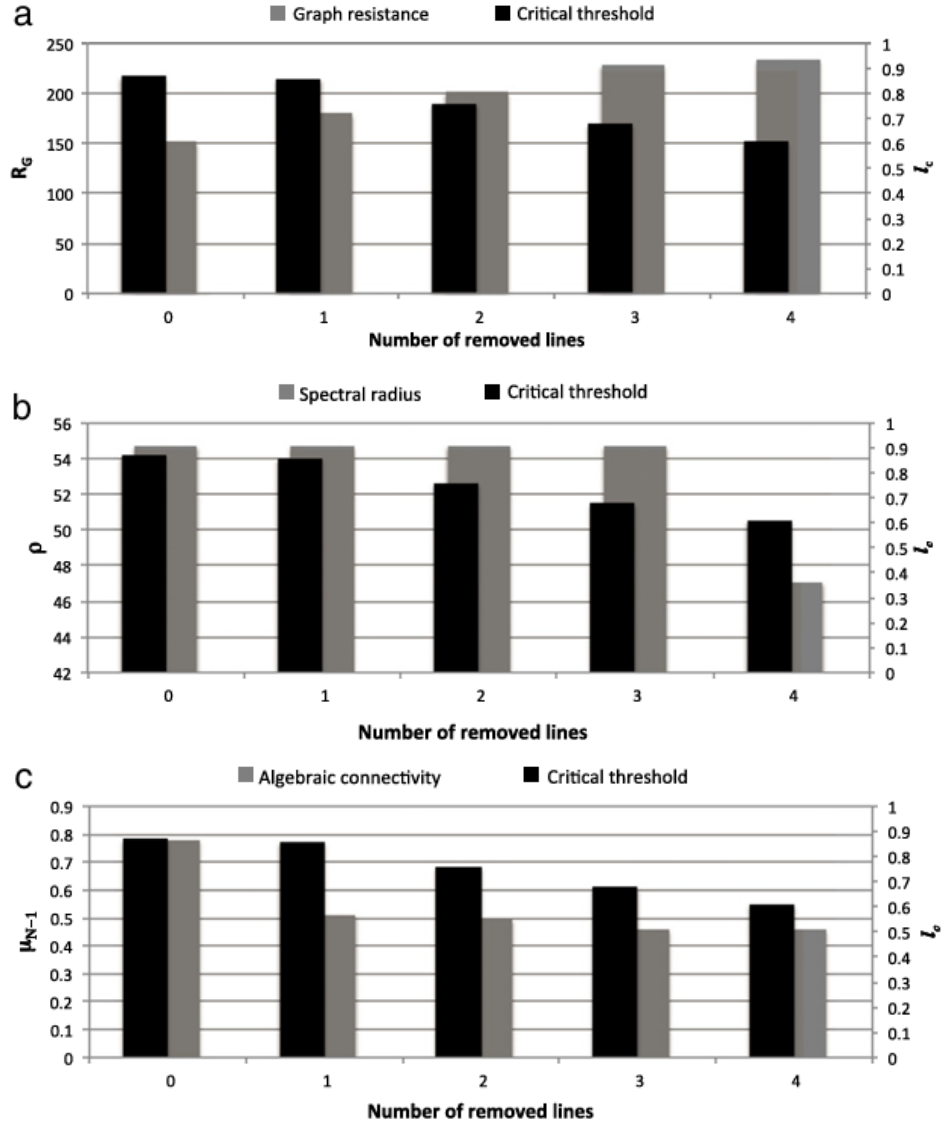


Figure 1.1: The impact of line removal on l_c vs R_G , ρ , and μ_{N-1} in IEEE 30 power system. The grid is attacked based on betweenness centrality. [3]

These changes are reflected in spectral graph metrics, effective graph resistance (R_G), which increases as electrical distances between buses grow, while algebraic connectivity (μ_{N-1}), defined as the second smallest eigenvalue of the graph's Laplacian matrix, decreases. Algebraic connectivity is a critical measure of the graph's robustness, reflecting how well-connected the network is and its ability to maintain

overall connectivity even when some connections are removed. A lower value of algebraic connectivity signifies reduced network resilience, making the system more vulnerable to disconnections and eventual cascades. This reduction shows a weaker topological cohesion, which describes the structural integrity of the network and its capacity to sustain efficient energy flow and coordination. Similarly, the spectral radius (ρ), the largest eigenvalue of the graph's Laplacian matrix, indicates the overall "spread" of node interactions and the influence of the network's structure on dynamic processes. It remains stable initially but drops sharply with extensive line removals. This captures the cumulative impact of topological degradation [3]. Together, these results emphasize the importance of understanding and mitigating topological vulnerabilities. Preventing the progression of cascading failures from local disruptions to global blackouts is crucial.

Preventing cascading failures is vital for safeguarding the stability and resilience of modern power grids. Addressing topological vulnerabilities can stop local disruptions from escalating into widespread blackouts. These efforts not only enhance reliability but also reduce economic losses, improve public safety, and facilitate the seamless integration of renewable energy, fostering a future defined by stability and sustainability.

1.2. Literature Review on Cascading Failures and Protection

This section shows the mechanisms behind failures, the progression of cascades, and protection strategies currently in use. The N-1 contingency criterion is discussed in [4] as a cornerstone for assessing power system stability, ensuring the system remains operational even if a single component fails. Mathematically, as per [5], the system must satisfy:

$$\sum_{i=1}^{N_g} P_{Gi} = P_D + P_{\text{loss}}, \quad P_{Gi}^{\min} \leq P_{Gi} \leq P_{Gi}^{\max}, \quad \forall i \quad (1.1)$$

$$V_i^{\min} \leq V_i \leq V_i^{\max}, \quad S_{km} \leq S_{km}^{\max} \quad (1.2)$$

where P_{Gi} is the real power generation at the i -th bus, P_D is the total demand, P_{loss} is the total real power loss, V_i is the voltage magnitude at the i -th bus, and S_{km} is the MVA power flow in branch km . While the N-1 criterion provides a robust baseline for the Transmission System Operator (TSO) to act on backup options, its limitations are evident when failures cascade, and the N-1 criterion is insufficient by itself. The cascades often result from overloading due to power redistribution after failure, relay miscoordination, or inadequate protection mechanisms such as load-shedding or islanding schemes. The culmination of cascading failures is a blackout, characterized by widespread loss of power, crippling critical infrastructure, and causing significant disruptions. Historical events described in [1], such as the 2003 North American blackout, illustrate the devastating consequences of cascading failures and highlight the need for enhanced predictive models and protective measures.

1.2.1. Failure Escalation in Power Systems: Examples and Mechanisms

Modern power systems face a wide range of failure mechanisms, ranging from isolated events to continuous disruptions that can compromise system stability. Understanding these mechanisms is crucial for designing resilient power grids capable of mitigating disruptions and maintaining operational integrity. Failures in power systems often stem from equipment malfunctions, overloading, or improper operation of protection devices. External factors, such as severe weather events or faults caused by animals and tree contacts, can also trigger disruptions. For instance, an overloaded transmission line may trip due to thermal stress, leading to further redistribution of power flows and reducing the system's stability margin. The failure of a component in a power network can be represented as an Element State Transition (EST), shown in [6], where the state of an element changes due to its operational condition. The element state $s_i(t)$ is binary: $s_i(t) = 0$ represents a connected and operational element, while $s_i(t) = 1$ indicates a tripped or disconnected element. The EST from 0 to 1 is significant in failures. The transition between these states is governed by two rates. The disconnection rate $\lambda_i(t)$ quantifies the likelihood of an element failing, while the reconnection rate $\mu_i(t)$ represents the probability of restoring the element to its functional state. These rates are influenced by factors such as equipment reliability, system load, and protective relay settings. Failures may occur naturally due to equipment degradation or externally through overloading. When an element's load $L_i t$ exceeds its capacity C_i then $\lambda_i(t) \neq 0$.

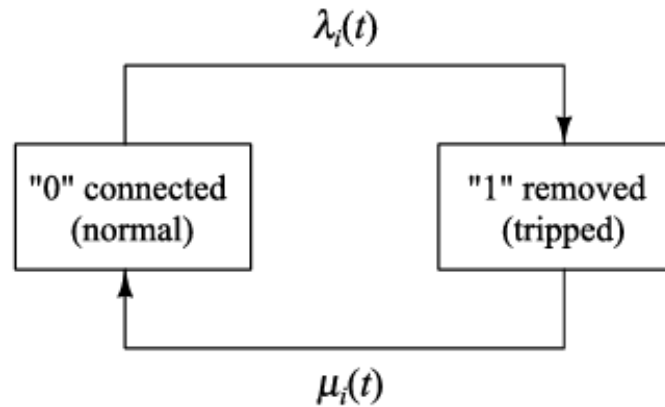


Figure 1.2: Description of failure in terms of state transitions [6]

Failures in power systems, whether localized or systemic, have significant consequences explained in [7] that extend beyond the immediate loss of functionality. Improper relay operations, changes in apparent impedance caused by converter-based generation, and latent or “hidden failures” (HFs) in protection schemes can all trigger serious disturbances. Under faulted conditions, the behaviour of Voltage Source Converter (VSC)-based High Voltage Direct Current systems deviates markedly from that of conventional synchronous generators, as they limit fault current contributions to protect power electronic components. The controlled current injection from VSC stations, which is often less than 1.15 p.u. modifies the positive and negative sequence fault currents. This altered fault signature confuses

conventional distance protection. For example, when a single line-to-ground (SLG) fault occurs on a transmission line connected to a VSC-based HVDC system, the positive, negative, and zero sequence networks must be considered to accurately compute the apparent impedance seen by the relay. The converter's constrained current injection, particularly its limitation of positive sequence current and suppression of negative sequence current, leads to atypical fault current waveforms. Analyzing an SLG fault on phase-A, the positive sequence apparent impedance, Z_{AG}^{app} , can be expressed as:

$$Z_{AG}^{app} = \frac{V_a^{R1}}{I_a^{R1} + 3KI_0^{tr}} \quad (1.3)$$

Where V_a^{R1} and I_a^{R1} are, respectively, the phase-A voltage and current at the relay location, I_0^{tr} is the zero sequence current through the transformer neutral, and K is the zero sequence compensation factor [8]. The faulted line's apparent impedance calculation, which traditionally assumes a known subtransient fault current magnitude and pattern, is complicated by the VSC's restricted fault current contribution. This current limitation, implemented to protect the converter's power electronic devices, alters the trajectory of the impedance locus in the $R-X$ plane during a fault event. Furthermore, the inability of the VSC station to provide adequate negative sequence current support due to the internal control strategy that prioritizes balanced positive sequence current injection affects the relay's interpretation of the faulted condition. As a result, the relay may 'underreach' or 'overreach' its intended protection zones, thus escalating a simple fault into a more severe event and leading to a cascade.

To understand how localized failures can develop into large-scale blackouts, deterministic power flow equations and probabilistic models of event durations can be combined. In this approach, localized failures are not seen as isolated incidents but as events within a probabilistic chain of state transitions. By representing EST using a Markovian framework, the collective behaviour of the entire system is analyzed in terms of failure propagation. An extended chemical master equation (CME) model is then used to capture these dynamics. Considering τ as the time interval between two consecutive network state transitions, the state transition probability density function (PDF), $f(\tau)$, represents the likelihood that the next transition occurs within an infinitesimal time interval. This is because the system is modelled in continuous time, where the interest is in the probability of the next transition occurring over an arbitrarily small interval, allowing for precise modelling of the instantaneous behaviour of the system. The equations from [6] demonstrate that the interval τ between state transitions follows an exponential distribution, with a mean interval inversely proportional to $\lambda^*(t_1)$. Formally, the likelihood is expressed in [6] as:

$$f(\tau) = \lim_{dt \rightarrow 0} \frac{H(\tau, dt) - H(\tau, 0)}{dt} = \lambda^*(t_1) \times e^{-\Lambda^*(t_1)\tau}, \quad (1.4)$$

where $H(\tau, dt)$ is the cumulative hazard function, which measures the accumulated likelihood of transitions up to time τ , and $H(\tau, 0)$ represents the hazard function at the starting point. Here, $\Lambda^*(t_1)$ denotes the network's overall failure rate at time t_1 , incorporating the contributions of all overloaded

and vulnerable elements in the system. Mathematically, $\Lambda^*(t_1)$ is the sum of the individual EST rates:

$$\Lambda^*(t_1) = \sum_{i=1}^N \lambda_i(t_1), \quad (1.5)$$

where $\lambda_i(t_1)$ is the individual state transition rate of the i -th element, and N is the total number of elements in the network. This sum reflects the system's overall overloading stress at time t_1 . The cumulative distribution function (CDF), which gives the probability that the next state transition occurs before $t_1 + \tau$, is:

$$F(\tau) = 1 - e^{-\Lambda^*(t_1)\tau}. \quad (1.6)$$

As the system becomes more stressed due to line overloads, hidden failures, or insufficient reactive support, $\Lambda^*(t_1)$ increases, causing transitions to occur more frequently and accelerating the cascading process. In essence, a higher $\Lambda^*(t_1)$ value means the system is more likely to experience rapid successive outages, pushing it closer to widespread instability and eventual large-scale service interruptions. This highlights that $\Lambda^*(t_1)$ is a key measure for understanding the growth of cascading failures in a power system and how increased stress on the system leads to more frequent failures and potentially widespread cascades.

HF described in [9] are permanent defects in protection schemes that remain latent until exposed by a fault, which may then trigger a switching event such as a line outage. These defects significantly influence fault escalation by converting otherwise manageable faults into cascading disruptions. When a relay with an HF misoperates, instead of isolating only the faulted component, it inadvertently removes additional non-faulted lines from service. This unintended outage redistributes power flows, pushing nearby lines closer to their limits, thereby increasing the likelihood of subsequent failures and exacerbating the cascading process, ultimately leading to complex chains of line tripping. Beyond identifying the initiating scenarios, severity indexing provides a means to rank HF-driven disturbances numerically. The severity indices presented in [4] help operators pinpoint the most vulnerable relays or areas. This information guides the implementation of protective measures such as self-checking digital relays, adaptive logic modifications, or enhanced monitoring. By reducing HF probabilities and prioritizing protective measures according to their severity, TSO operators can mitigate the progression from initial defects to large-scale blackouts.

1.2.2. Cascading Failure Dynamics: Modeling and Historical Perspectives

Cascading failures in power systems unfold through distinct phases, each characterized by specific dynamics and challenges in [10]. Understanding these phases is crucial for developing effective mitigation strategies to enhance grid resilience. The initial stage of cascading failures is the precursor phase, where components progressively fail as they reach their operational limits due to increasing stress, such as voltage fluctuations, overloads, or equipment malfunctions. Prompt intervention is

crucial to prevent disturbances from escalating into severe outages, enabling system operations to implement control actions effectively. One of the primary mitigation techniques during this phase is controlled islanding, where a strategic partitioning of the power grid into smaller, autonomous sections occurs. This aims to isolate failure sections, preventing their propagation across the entire network. Techniques such as density-based spatial clustering (DBSCAN) combined with non-linear programming described in [11] have been proposed to identify coherent generator groups suitable for islanding. Furthermore, a multi-layer spectral clustering approach in [12] enhances the accuracy of island formation by analyzing the spectral properties of the grid's adjacency matrix, thus minimizing the risk of imbalanced load-generation ratios within the isolated sections. This is shown in Figure 1.3b, where coherent generators are clustered in groups.

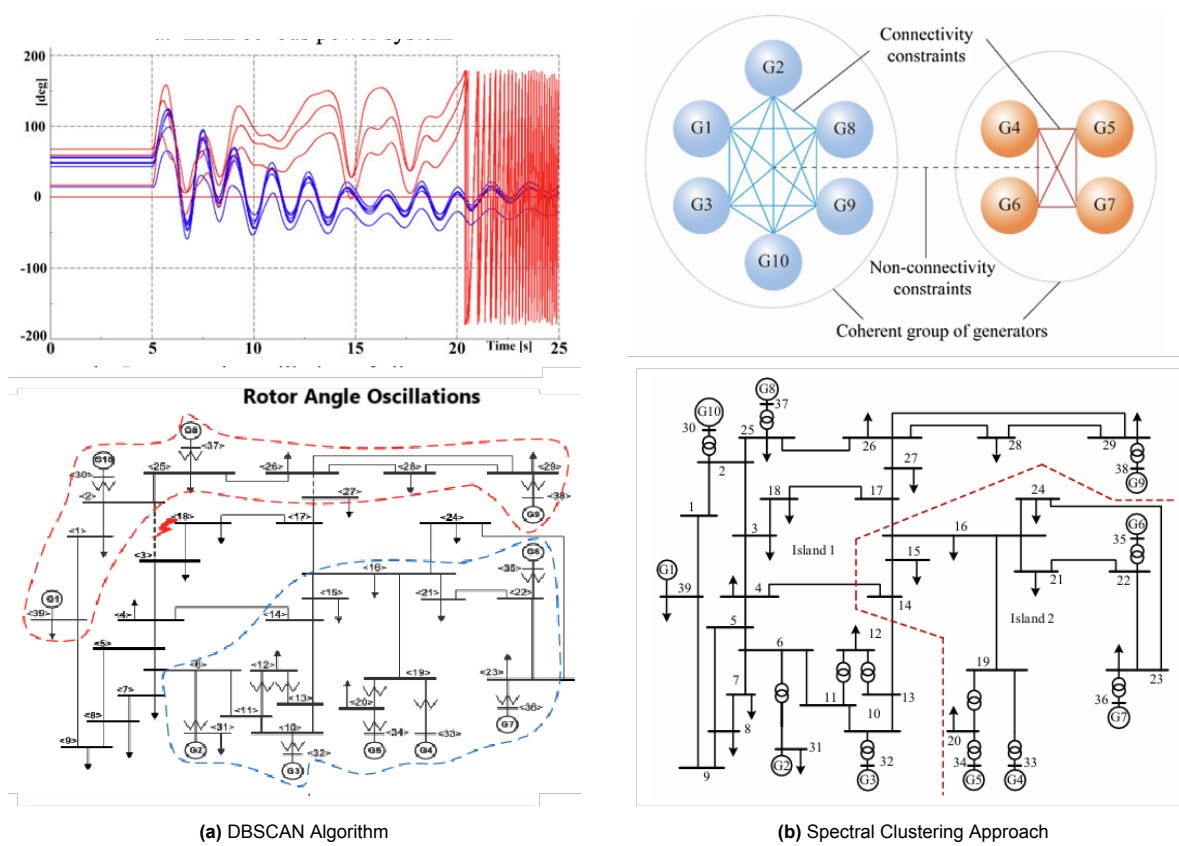


Figure 1.3: IEEE 39-bus coherent grouping of generators and islanding boundaries applied with (a) DBSCAN algorithm [11] and (b) Spectral Clustering Approach [12]

Load Shedding can be another critical mitigation strategy that can be used during the precursor phase. It involves reducing the load in specific areas to prevent overloads and maintain system stability. Machine learning (ML) models have been developed to determine the optimal load-shedding amount, ensuring minimal disruption while effectively stabilizing the grid. A deep neural network in [10] with specialized loss functions can be used to predict the required load-shedding amount, capturing the non-linear relationships during real-time operating states. The loss function used in this model is de-

defined as:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^M w_j^2 \quad (1.7)$$

where y_i represents the actual load shedding amount, \hat{y}_i is the predicted amount by the neural network, λ is the regularization parameter, and w_j are the network weights. This formulation ensures that the model not only minimizes prediction errors but also prevents overfitting by penalizing large weights. Figure 1.4 in [13] illustrates the flowchart of the cascading failure process in power grids, highlighting the steps involved in simulating and predicting cascading failures.

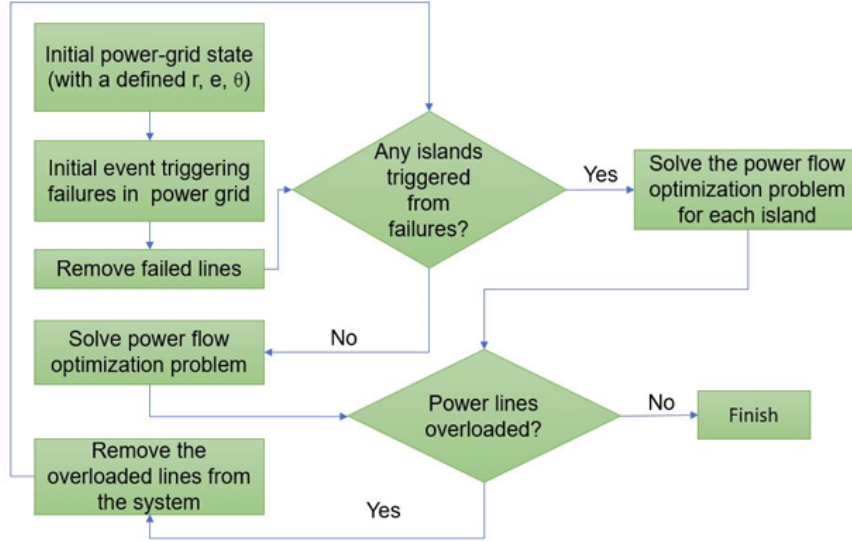


Figure 1.4: Flowchart of Cascading Failure Process [13]

The next phase is the Dispatching Phase, which refers to the real-time management of power generation resources to balance supply and demand. Effective dispatching is critical in maintaining system stability and preventing the escalation of cascading failures. Then comes the Escalation Phase, where cascading failures accelerate rapidly, leading to widespread outages and making blackout prevention exceedingly difficult. The swift progression leaves minimal time for effective intervention, exacerbating the impact on the power grid and society at large. Critical component failures often trigger this phase, facilitating the spread of disturbances across the network. During this phase, the component failure rate exponentially rises, overwhelming the control mechanisms. This phase gives us a sense of the importance of rapid detection and automated response systems before the cascade spirals out of control. Finally, the Cascade Phase-out is reached, where the failure propagation reduces as a significant portion of the grid's components have already failed. During this phase, the system reaches a plateau where the rate of new failures diminishes, and the remaining operational components attempt to stabilize the grid. This phase represents the culmination of the cascading process, where the extent of the blackout is determined. In this phase, recovery and restoration efforts are carried out. The focus shifts to re-establishing critical infrastructure and restoring power to affected areas. Understanding

the dynamics of the phase-out phase is essential for effective post-event analysis and for developing strategies to enhance system resilience against future cascading failures.

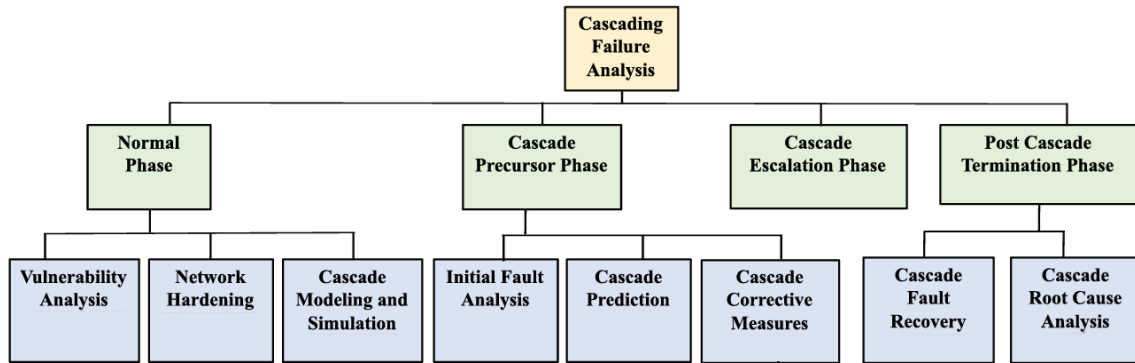


Figure 1.5: Flowchart of cascading failure analysis based on different phases before, during, and after cascade. [10]

Analyzing the impact of power measurements on failure dynamics gives us an idea of how active and reactive power measurements are crucial for monitoring the operational integrity and stability of power systems. Active power (P) and reactive power (Q) measurements obtained from Phasor Measurement Units (PMUs) are utilized to detect and predict cascading failures in power grids with renewable energy integration. Active power measurements provide real-time data on the balance between power generation and consumption, enabling the identification of transmission line overloads that may lead to component tripping. Reactive power measurements, on the other hand, are essential for maintaining voltage stability across the network. Sudden fluctuations or deviations in Q can indicate voltage instability, which is a precursor to voltage collapse and subsequent cascading outages. The interplay between P and Q measurements allows for the early detection of imbalance and instability, facilitating timely interventions to prevent the escalation of failures. For instance, the relationship between active power flow and voltage angles can be expressed by the DC power flow equation:

$$P_{ij} = \frac{\theta_i - \theta_j}{X_{ij}} \quad (1.8)$$

where P_{ij} is the active power flow from bus i to bus j , θ_i and θ_j are the voltage angles at buses i and j , respectively, and X_{ij} is the reactance of the transmission line connecting them. System operators measure voltage and current in real-time, from which power flows (P , Q) are computed. Monitoring these parameters enables the detection of abnormal operating conditions that could contribute to cascading failures, thereby enhancing the resilience and reliability of the power grid. By analyzing deviations in voltage levels and current magnitudes, operators can assess risks related to voltage instability and thermal overloading. This information helps prioritize intervention strategies, ensuring that control actions are both timely and effective in preventing widespread outages.

Historical Perspectives on Cascading Failures: Case Studies and Mitigation Strategies

Examining major blackout events provides critical insights into the dynamics and vulnerabilities of power transmission systems. The case studies in [14] show us the mechanisms that drive failures and evaluate the effectiveness of various strategies. An in-depth analysis of cascading failures within the North American power grid utilized a network efficiency framework to quantify system robustness. The results demonstrated that the probability distribution of disturbances follows a power law with an exponent close to -1.1. This finding suggests a low overload tolerance within the grid, where even minor disturbances can propagate, increasing power losses and reducing overall transmission efficiency. Figure 1.6 illustrates how the removal of nodes from the North American power grid, based on centrality measures such as betweenness and degree, impacts overall system stability. Centrality measures help identify nodes that play a crucial role in facilitating power flow across the network.

- **Betweenness centrality** quantifies how often a node appears on the shortest paths between other nodes in the network. A node with high betweenness acts as a critical connector, meaning its failure can significantly disrupt power transmission and force inefficient rerouting.
- **Degree centrality** measures the number of direct connections a node has. Nodes with a high degree serve as major hubs, and their removal can disconnect large portions of the grid.

The analysis categorizes nodes into three distinct classes:

1. **Non-impact nodes** ($\approx 60\%$ of the total): These nodes have low betweenness and low degree, meaning they play a minimal role in power transmission. Their removal causes negligible or no disruption to the grid.
2. **Critical nodes**: These nodes exhibit high betweenness and high degree, making them essential for efficient power flow. Their removal leads to significant disruptions, with the extent of damage depending on the system's built-in tolerance levels.
3. **Transition nodes**: These nodes initially contribute to the tolerance-dependent damage curve, meaning their removal weakens the grid to some extent. However, beyond a critical tolerance level, removing additional transition nodes no longer causes further damage, suggesting redundancy in the network structure.

This classification highlights the structural vulnerabilities of the power grid. While most nodes have limited individual impact, a small subset of critical nodes disproportionately influences grid stability. The presence of transition nodes suggests that strategic reinforcements or reconfigurations could enhance the grid's resilience, reducing the risk of cascading failures when key nodes are lost.

The graph from [14] reveals two primary behaviours: nodes whose removal causes no significant damage and those that follow a universal damage-versus-tolerance curve, leading to substantial efficiency losses. Additionally, a transition category is identified, where nodes shift from causing significant damage to having a negligible impact as tolerance levels increase. It shows the vulnerability of the power

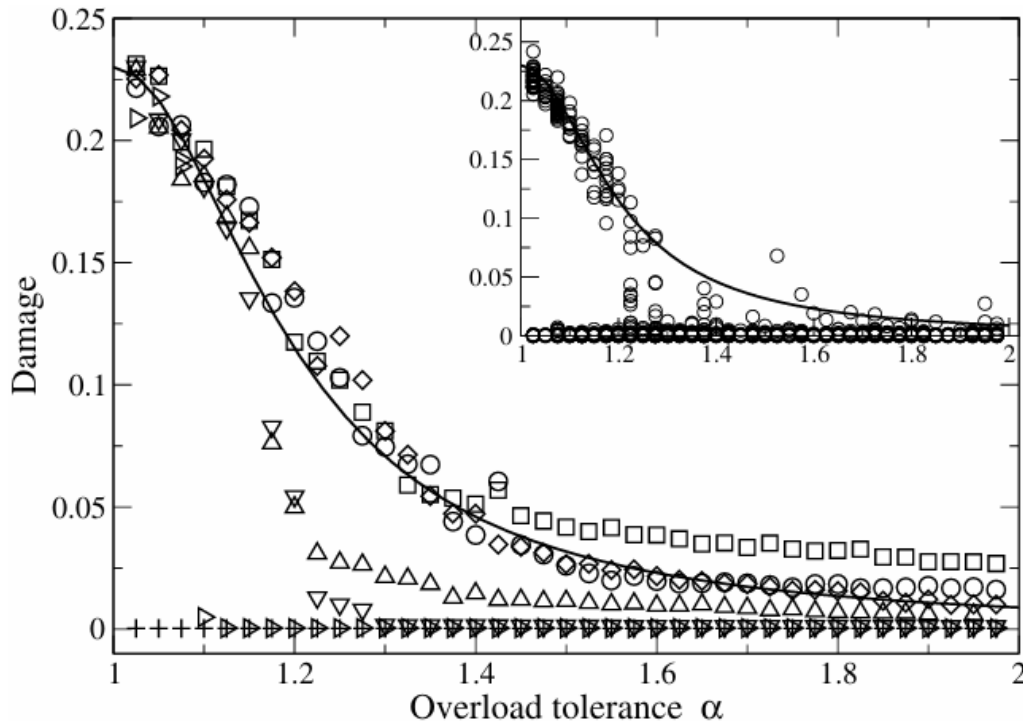


Figure 1.6: Representative sample of node-dependent damage for different tolerance values [14]

grid to targeted failures of high-centrality nodes and emphasizes the necessity of identifying and reinforcing these critical components to prevent large-scale cascading failures.

The understanding of cascading failures was further expanded by compiling a comprehensive database of European blackout events in [15], encompassing both continental and national scales. Their empirical analysis revealed that cascading failures are the predominant threat to modern transmission systems, with weather events being the most common initiators. The study identified significant correlations between component failure rates and system demand, as well as wind speeds, emphasizing the role of external factors in aggravating system vulnerabilities. Furthermore, the weaknesses in recovery prioritization, where low-impact events disproportionately contribute to prolonged recovery times, are highlighted.

From Figure 1.7 obtained from [15], the analysis of blackout events in the continental and national datasets highlights key insights into failures and their impacts. (A and B) show the contribution of different component sets to total failures, with the national dataset also indicating failure rates, distinguishing components owned by the national TSO (red line) and others (blue line). (C and D) focus on the impact of demand not served (DNS), categorized by seasonality, event type, and main cause, revealing conditions that intensify service disruptions. Finally, (E and F) examine recovery times, also influenced by seasonality, event types, and root causes, providing a comprehensive understanding of factors affecting system resilience and recovery. Beyond this, several other historical blackouts offer additional perspectives. Notable instances include the Texas blackout of 2021 and the separations of

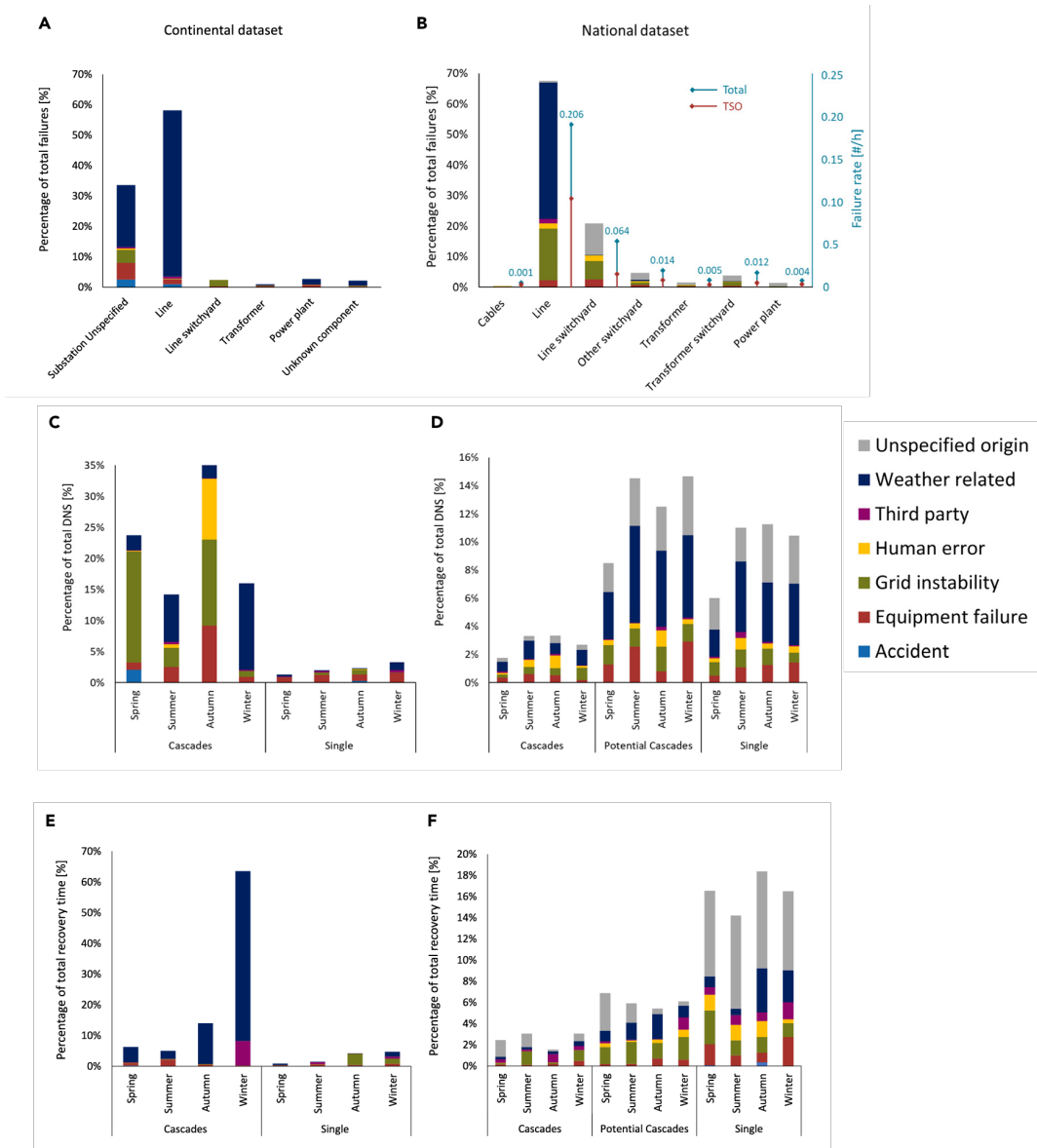


Figure 1.7: Leading causes for blackout events and component failures in continental and national datasets [15]

the European synchronous area in January and July 2021. These events show how interconnected networks can rapidly deteriorate under stress, leading to widespread power outages. For example, the Texas blackout mentioned in [14] highlighted the susceptibility of power grids to extreme weather conditions combined with high-demand scenarios.

The evolution of mitigation strategies against cascading failures has transitioned from primarily structural interventions to more sophisticated approaches that integrate both structural and temporal considerations. An average network efficiency metric in [15] quantifies the impact of node removals, facilitating the identification of critical components whose failure would disproportionately affect system performance. It measures how efficiently information (or, in this context, power) is transferred over the network by averaging the inverse of the shortest path lengths between all pairs of nodes. Higher

network efficiency indicates a more robust and resilient power grid, as the removal of nodes has a diminished effect on the overall connectivity and performance of the system.

$$\text{Network Efficiency (E)} = \frac{1}{N_G N_D} \sum_{i \in G_G} \sum_{j \in G_D} \epsilon_{ij} \quad (1.9)$$

Where N_G and N_D are the number of generation and distribution parameters, ϵ_{ij} is the efficiency of the most efficient path between the generator i and the distribution substation j . This quantitative approach supports targeted interventions, such as reinforcing high-load nodes or enhancing network redundancy. Meanwhile, the integration of advanced technologies, including real-time monitoring systems, artificial intelligence (AI), and machine learning (ML) algorithms, to predict and respond to potential cascading failures is also employed. These findings highlight the utility of comprehensive blackout databases and predictive analytics in identifying early warning signs and critical component correlations, enabling proactive measures to be implemented before failures cascade through the network.

Modeling Cascading Failures, Traditional and Probabilistic Approaches

Effectively modelling cascades in power systems is crucial for understanding their dynamics and developing robust strategies. Traditional approaches for cascade modelling predominantly use network-based models and dynamic models. Network-based models in [3] represent the grid as a complex network where buses are the nodes and edges are transmission lines. This helps in analyzing the grid's structural properties and its susceptibility to cascades. Graph theory can be employed to investigate phase transitions, utilizing metrics such as effective graph resistance (R_G), spectral radius (ρ), and algebraic connectivity (μ_{N-1}) to assess grid robustness as mentioned earlier. A study on the IEEE 118 bus system demonstrated that algebraic connectivity (μ_{N-1}) strongly correlated with critical loading thresholds (l_c), indicating that it is efficient in predicting the onset of phase transitions under various conditions, including δ -based and betweenness centrality (CB)-based attacks. The effective graph resistance is calculated as:

$$R_G(u, v) = L_{uu}^+ + L_{vv}^+ - 2L_{uv}^+ \quad (1.10)$$

where L^+ is the pseudoinverse of the laplacian matrix L and provides a measure of connectivity between nodes u and v . This metric, along with others, enables the identification of critical components whose failure can cause widespread outages.

Visualization of these relationships is depicted in Figure 1.8 from [3], which illustrates the correlation between critical loading thresholds and spectral graph metrics.

Dynamic models use real-time operating conditions and topological changes within the power grid. Unlike static network-based models, [16] shows that dynamic models use the temporal evolution of the grid's state, enabling the prediction of how initial disturbances propagate and evolve. An adaptive PMU-

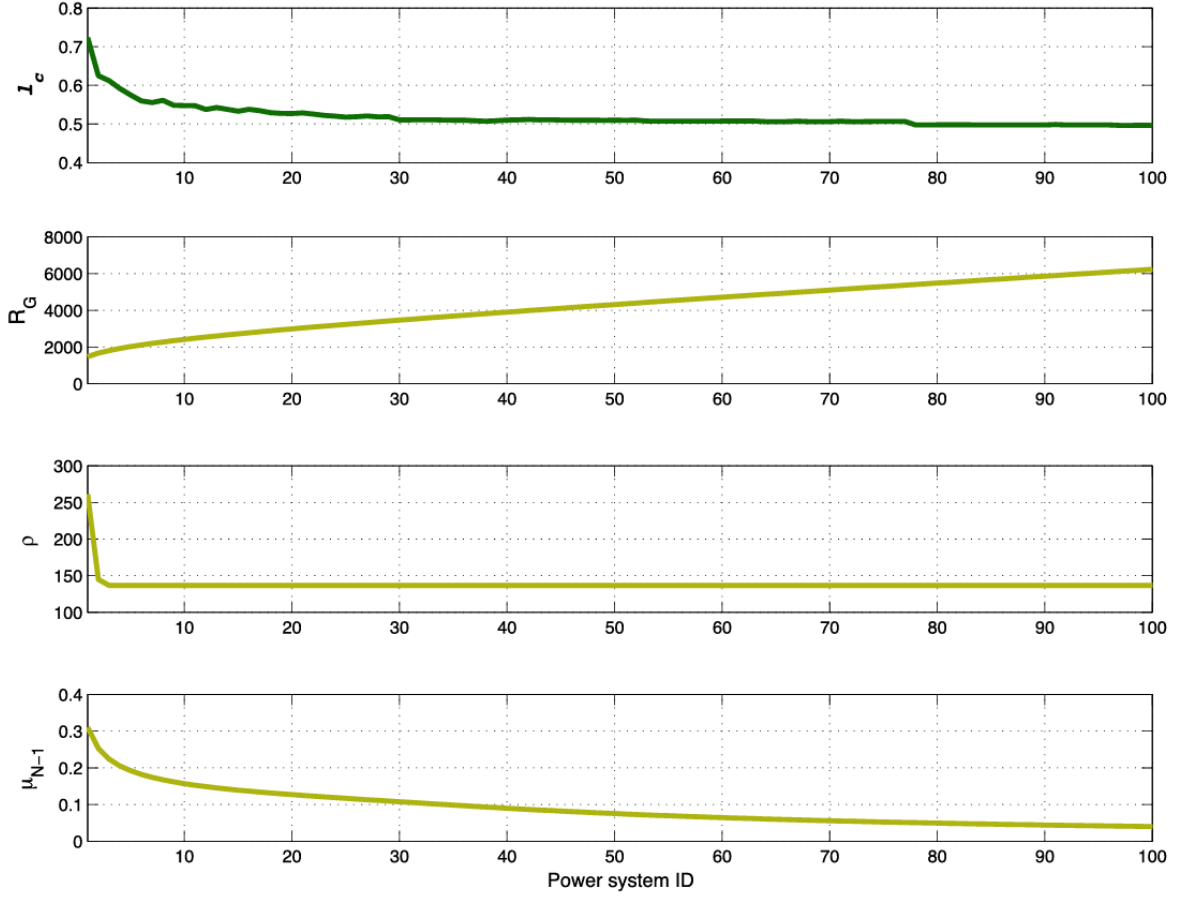


Figure 1.8: Critical loading threshold and spectral graph metrics for the IEEE-118 with test systems attacked with a betweenness centrality strategy [3]

based Wide Area Backup Protection Scheme utilizes PMUs to provide real-time voltage and current measurements. It also optimizes PMU placement to form Backup Protection Zones (BPZs). When a fault occurs within a transmission network, the sum of zero- and positive-sequence currents entering the affected BPZ increases significantly, and the algorithm first identifies the faulted BPZ by comparing the measured currents against predefined adaptive thresholds. The faulted line and its precise location within the BPZ are then determined using a linear least squares method whose mathematical formulation involves constructing an observability function and solving for the unknown fault parameters using $Hu = m$. H is a matrix composed of voltage and current measurements from PMUs, u is the vector of unknown fault parameters, and m is the measurement vector. The solution yields the fault location and the fault current. The integration of PMUs into the Wide Area Measurement System (WAMS) facilitates the collection of high-frequency data. Simulation results from the WSCC 9-bus and IEEE 118-bus test systems demonstrated the scheme's effectiveness in identifying faulted zones and lines with limited measurement points, highlighting its practical applicability in large-scale power systems.

Moreover, dynamic models can be augmented with stochastic and Monte Carlo simulations (MCS), as explained in [17], to account for the probabilistic nature of cascading failures. These simulations

allow for the assessment of various fault scenarios and their potential impacts on the grid, providing a more comprehensive understanding of failure propagation under uncertainty. This MCS methodology involves several key steps. Firstly, it generates numerous system states by simulating random combinations of generation and transmission outages, capturing the inherent uncertainties and variabilities. Secondly, for each generated scenario, the system's frequency deviation is calculated by considering the primary frequency control responses of generators and the frequency sensitivity of loads. This step assesses the ability to restore generation and load balance following disturbances. Further, integrating MCS with DC OPF to evaluate the risk can be used. It serves as remedial action to minimize lost load by adjusting generation dispatch and shedding loads where necessary. The simulation tracks the amount of lost load (Expected Load Not Served, ELNS) and constructs the Complementary Cumulative Distribution Function (CCDF) of blackout sizes. To accurately estimate the Probability Distribution Function (PDF) and smooth the CCDF of the lost load data, the Gaussian Mixture Method (GMM) can be applied. This combination of MCS and optimization provides a framework for evaluating blackout risks, demonstrating that cascading outages significantly influence the distribution of blackout sizes. The CCDF, used to evaluate the risk associated with lost load, is calculated using CDF ($F(x)$) and PDF ($f(t)$) from the estimated PDF using the following equation:

$$CCDF = 1 - F(x) = 1 - \int_{-\infty}^x f(t)dt \quad (1.11)$$

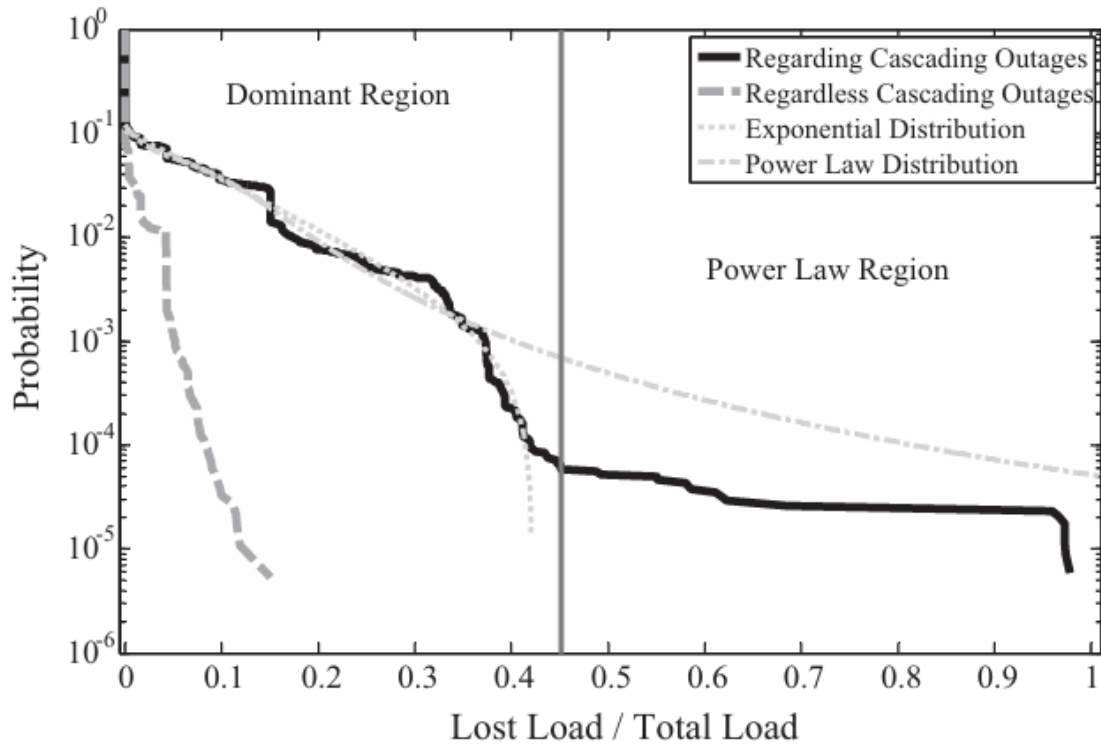


Figure 1.9: The CCDF of lost load for IEEE-118 system with and without cascading [17]

Figure 1.9 from [17] presents the CCDF of lost load data for the IEEE 118-bus test system, comparing scenarios with and without cascading outages. In the absence of cascading failures, the CCDF follows an exponential distribution, indicating that large-scale blackouts are relatively rare and their probabilities decrease exponentially with increasing lost load. Conversely, when cascading outages are incorporated into the model, the CCDF transitions to a power-law distribution in the tail region, signifying a higher probability of substantial load losses. This shift underscores the profound impact of cascading failures on system reliability by introducing dependencies between component outages, allowing small initial disturbances to propagate and amplify, thereby increasing the likelihood of large-scale blackouts. Overall, dynamic models that incorporate cascading outages and system responses are essential for accurately assessing the risks of blackout.

Machine Learning Techniques in Cascading Failure Prediction

In recent years, ML has offered improved accuracy and adaptability and can efficiently process large-scale datasets. One of the primary advantages is the predictive capability of ML models for fault detection and classification, which are the initial causes of cascades. As per [18], a model such as the Conformer Convolution-Augmented Transformer can effectively perform transmission line fault classification by processing raw time series data directly, bypassing the need for manual feature extraction and achieving a testing accuracy of 88.88%, outperforming the Extreme Learning Machine (ELM) model, which can attain only 62% accuracy. Similarly, [19] suggests using Long Short-Term Memory (LSTM) networks employed for fault direction estimation in Networked Distribution Systems (NDS), significantly improving robustness and accuracy of fault detection compared to conventional Radial Distribution System (RDS) methods.

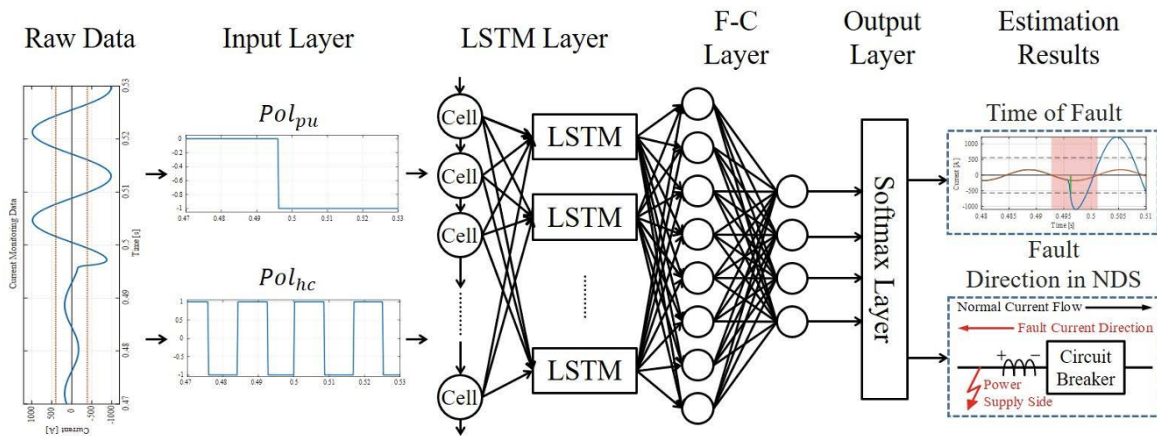


Figure 1.10: Structure of LSTM for fault current and direction determination [19]

Figure 1.10 illustrates an LSTM-based fault direction estimation system for power networks given in [19] that uses features of the current waveform such as C_{norm} , Pol_{pu} , Pol_{hc} to predict fault direction and time of detection. It optimizes fault detection accuracy using time-series simulations and adaptive learning, aiding in precise protection coordination.

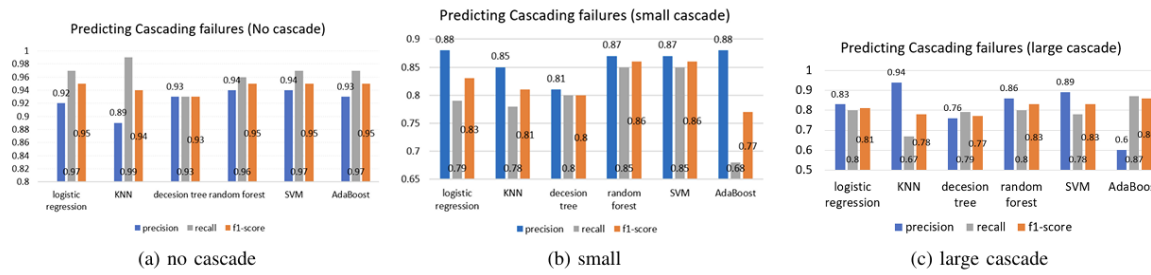


Figure 1.11: Comparison of cascading classes using various models and performance metrics [13]

Another significant advantage of ML techniques given by [20] is their ability to identify critical features from raw data automatically, reducing the dependence on feature extraction. Deep Neural Networks (DNNs) that detect and classify faults in 3-phase long transmission lines are utilized as the model in [21] to process voltage and current measurements, effectively learning features without manual intervention. Additionally, such models exhibit exceptional scalability and efficiency as they are integrated with meta-heuristic techniques, such as the Wild Horse Optimization (WHO), for fault detection, classification, and localization in power grids to achieve high accuracy with minimal computational cost.

A cascading failure simulator framework using MATPOWER in [13] generates extensive datasets based on the IEEE 118-bus system, and ML models trained on these comprehensive datasets can effectively predict and classify cascading failures with high accuracy. Here, the prediction of cascade consists of three classes, no cascade, small cascade and large cascade, thereby making it a multi-class classification problem.

Figure 1.11 in [13] compares various ML models using performance metrics and reveals that the classification of "no cascade" scenarios gives a higher precision compared to "high cascade" scenarios. This is due to high cascade scenarios having fewer test samples than no cascade scenarios, showing a limitation of ML models due to the limited availability of high cascade data and creating a class imbalance. The next challenge is managing high-dimensional and complex datasets. As [10] suggests, computational cost also poses a big challenge, particularly when training complex ML models on large-scale power systems and the impact of cyber-attacks on ML models has not been explored adequately. Lastly, the interpretability and trustworthiness of ML models remain significant considerations, particularly in high-stake power system applications. Inaccuracies in power flow measurements can lead to wrong fault classification and unstable responses, making robust data preprocessing and noise mitigation necessary. An analysis in [22] shows that deep learning models, while powerful, often operate as black boxes, making it challenging to interpret decision-making. As power grids evolve, real-time conditions, such as changing grid topologies, generation variability, and diverse operations, may vary significantly from the data used to train these models. This can lead to decreased accuracy of prediction and reliability, undermining the potential benefits of these ML models predicting outages.

Probabilistic models, as suggested by [23], have emerged as essential tools in identifying blackouts due

to cascades, offering frameworks that account for the inherent uncertainties and complex interactions within power systems. Regeneration-based probabilistic models are designed to capture the stochastic nature of cascading failures by leveraging critical parameters such as the maximum capacity of failed lines (C_{max}) and the grid-loading ratio ($R_{D/G}$), defined as the ratio of total demand to total generation capacity. These models partition the topological space of a power grid into similar classes, simplifying the complex configuration space into manageable grid states defined by C_{max}, F , where F represents the number of outaged lines. This reduction in state-space complexity allows the model to scale to large power grids, making it computationally feasible.

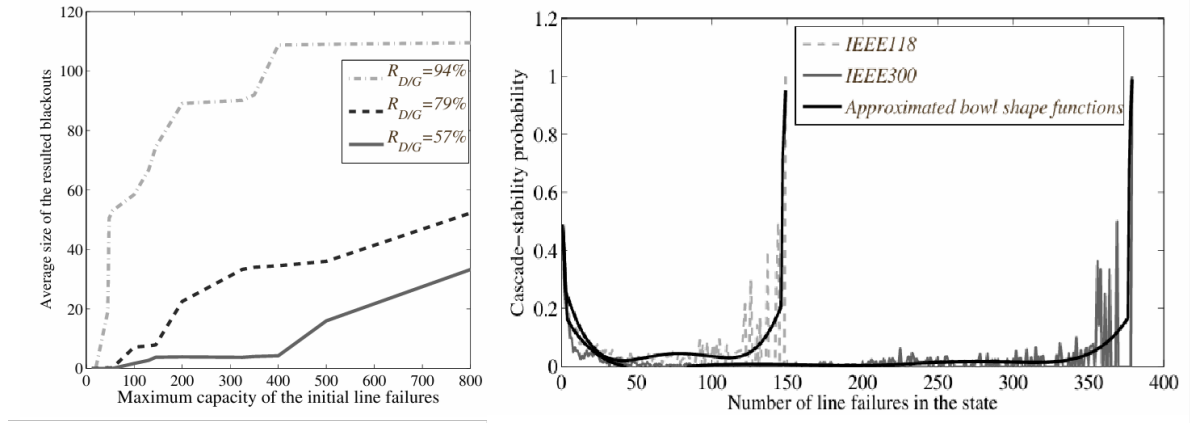


Figure 1.12: Average blackout size vs. maximum capacity of initial line failures for different grid-loading ratios and Cascade-stability probability vs. Failures in the system [23]

The effectiveness of this probabilistic model for blackout given in [23] and illustrated in Figure 1.12 shows the average size of blackouts as a function of the maximum capacity of initial line failures for various grid-loading ratios. The results show that higher C_{max}, F values lead to larger blackouts, with grid-loading ratios near maximum capacity multiplying the cascading severity. The cascade-stability probability, $P_{stab}(S, R_{D/G})$, quantifies the likelihood that a grid state S remains stable. P_{stab} follows a bowl-shaped curve in the second image, indicating higher stability for grid states with very few or many failed lines. In contrast, grid states with intermediate failure counts are more likely to transition to subsequent failures and decrease with increasing C_{max} , showing a destabilizing effect of such high-capacity failures. The model in [23] has used coupled differential equations to compute blackout probabilities over time. These equations incorporate cascade-stability probabilities and transition rates between grid states and reveal a two-phase failure process: an initial slow phase followed by rapid escalation, consistent with observed real-world blackouts.

Probabilistic load flow analysis offers another approach to modelling cascading failures by assessing the statistical properties of power flows under normal and perturbed grid conditions. This method uses parameters such as mean, variance, skewness, and kurtosis to characterize the transition from stable to unstable grid states. This framework captures cascading failures by analyzing line outages based on their probability distribution function (PDF) and cumulative distribution function (CDF). In normal grid

conditions, PLF assumes a Gaussian distribution for power flows. The transmission lines operate within their installation capacity, and grid variables such as voltage, current, and power flow follow a stable probability distribution. When a disturbance, such as a line tripping, occurs, the load is redistributed among the remaining lines, often leading to overloads and further failures. Probabilistic models used in the IEEE 30-bus system predict subsequent failures by analyzing shifts in statistical distributions [24]. After the tripping of specific transmission lines, the power flow distribution shifts from Gaussian to non-Gaussian, indicating increased load stress and system instability. The PDF shows a sharp rise in load stress, with vulnerable lines having higher cumulative values. Higher-order moments, such as skewness and kurtosis, increase significantly, signaling the onset of instability. When a line trips, the PDF can exhibit a heavy-tailed distribution, and the CDF may reach unity, indicating the approach of a blackout. This shift from Gaussian to non-Gaussian distributions is a key indicator of cascading failures. Future applications may integrate real-time data and topological features to enhance predictive accuracy and resilience.

Integration of Dynamic Topological Changes

Dynamic topological changes in power grids are critical to addressing cascading failures and improving resilience. It can be observed in [25] that the ability to adapt and optimize network topology dynamically has profound implications for the stability, synchronization, and efficiency of power systems. Networks with a higher topological disorder characterized by the presence of "shortcut" links tend to exhibit two contrasting behaviours. First, they are more robust in handling increased loads, reducing the likelihood of cascading outages. Second, when failures do occur, these networks break apart more quickly compared to regularly structured networks. Using synthetic small-world network models and IEEE 57 and IEEE 118 bus test cases, a trade-off was identified between robustness to initial failures and vulnerability to grid breakup. This trade-off was expressed through the clustering coefficient C and characteristic path length L , which shows the network interconnectivity and average distance between nodes, respectively. Simulations performed using DC Power Flows revealed that increasing "shortcut" links in a network reduces congestion but accelerates fragmentation once failures propagate. Dynamic modelling was considered by applying the Kuramoto model to large-scale European power grids. The swing equation, a second-order Kuramoto equation, was employed to study synchronization under various topological configurations:

$$\ddot{\theta}_i + \alpha \dot{\theta}_i = P_i + \sum_j W_{ij} \sin(\theta_j - \theta_i) \quad (1.12)$$

where α is the damping parameter, P_i represents nodal power, and $W_{i,j}$ defines the adjacency matrix of the network. This model accounted for both local synchronization and global dynamics, identifying weak nodes based on local order parameters r_i :

$$r_i = \frac{1}{N_{\text{neigh}}} \sum_j A_{ij} e^{i\theta_j} \quad (1.13)$$

Here N_{neigh} is the number of neighbours of node i , and A_{ij} is the adjacency matrix indicating the connection between nodes i and j . Thus, it is concluded in [26] that by adding bypasses or reinforcing critical nodes with low synchronization, a significant improvement in grid stability near the synchronization transition point can be obtained..

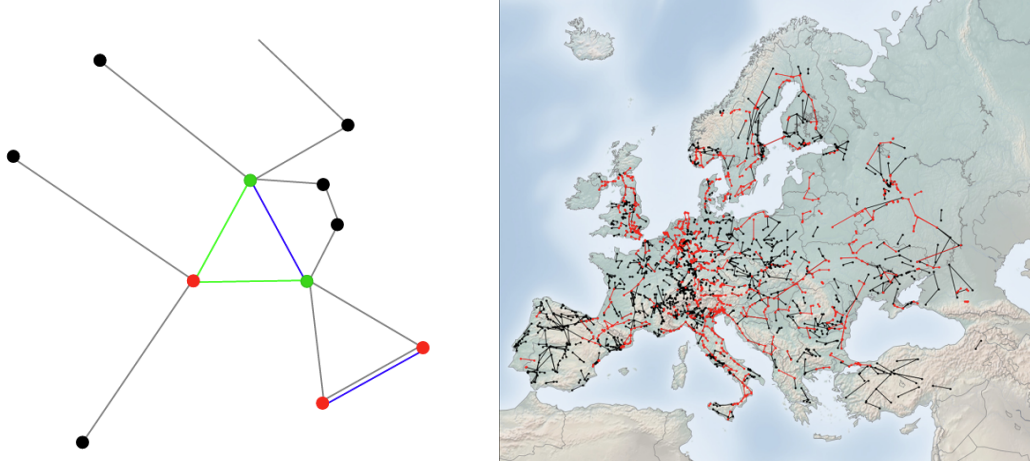


Figure 1.13: With the bypass algorithm, weak nodes (red) are reinforced by creating triangles with the two closest neighbors (green) or doubling edges between neighboring weak nodes (blue links) as an example (left) and overall Europe (right) [26]

Weak nodes, shown in red in Figure 1.13, obtained from [26], are prioritized for reinforcement. For each red (weak) node, its two closest neighbours (marked in green in the figure) are selected for intervention. If the weak node has no other weak nodes in its immediate vicinity, a triangle is created by adding a new edge (blue) between the two closest neighbours. This triangular structure enhances the local clustering coefficient and provides alternative paths for power flow, reducing potential overloads. On the other hand, if the weak node has two weak nodes neighbouring it, the edge connecting them is doubled, giving parallel pathways (original grey link reinforced with a blue link). This ensures stability during power redistribution under stress. By creating triangles or doubling edges at critical weak points, the algorithm ensures robust synchronization and enhances overall grid resilience.

Braess's Paradox, mentioned in [26], is a common and counterintuitive phenomenon in networked systems where adding new links or increasing capacity in a network can degrade overall system performance. In the context of power grids, this paradox manifests when additional links or modifications aimed at improving connectivity inadvertently cause instability or promote cascading failures. Using simulations of the European high-voltage power grid, the paradox was observed at lower coupling constants (K), where the system's power flow is limited and adding new links increases instabilities by redistributing loads in non-optimal ways. When W_{ij} in Equation 1.11 is modified to add new links, the system dynamics deviate from optimal synchronization, particularly when operating away from critical coupling thresholds. Exploring Braess's Paradox in terms of spectral graph metrics on the IEEE 118 power system shows that increasing impedance values in specific transmission lines or introducing new high-impedance links can lower the critical loading threshold (l_c).

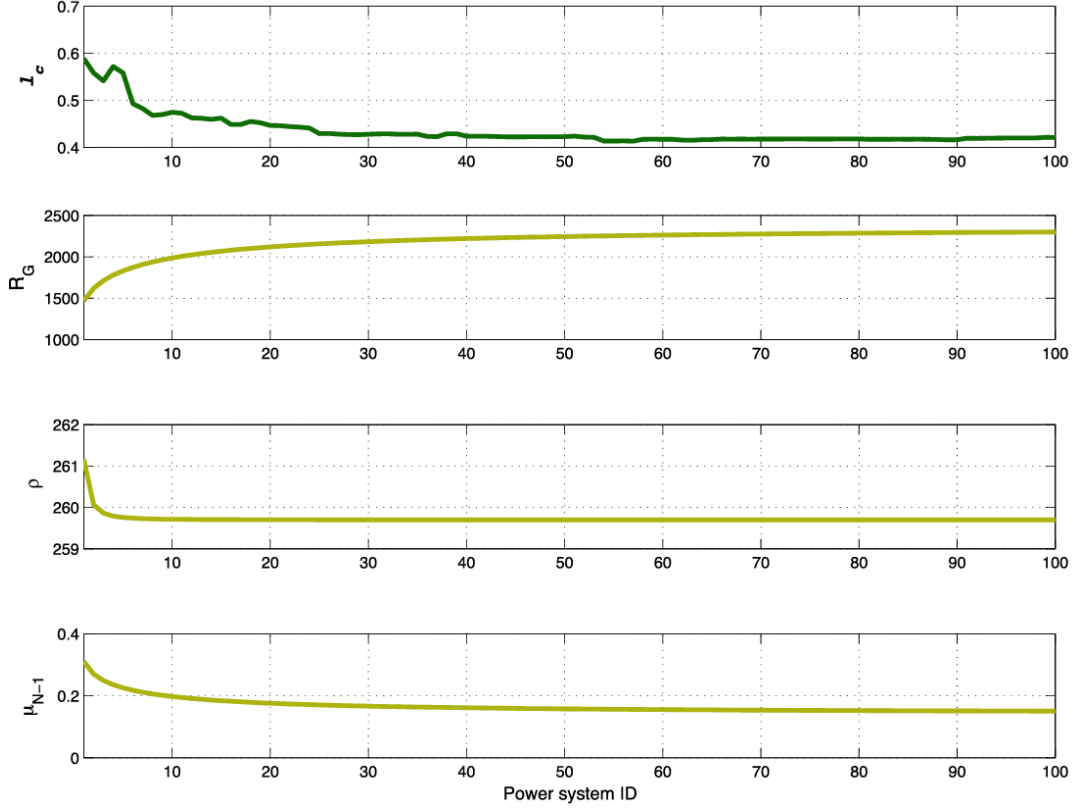


Figure 1.14: Critical loading threshold and spectral metrics for the IEEE 118 with node significance-based attack [3]

Figure 1.14 from [3] shows Braess's Paradox in action on the IEEE 118 system, where even the slight topological adjustments caused significant shifts in l_c . The addition of capacity to specific nodes caused earlier phase transitions, reducing the grid's ability to sustain power demand after cascades.

1.2.3. Cascade Protection Mechanisms: Preventive and Corrective Controls

To mitigate cascading risk, various protection mechanisms have been developed, broadly categorized in [27] into preventive and corrective control strategies. Preventive control aims to enhance the robustness of the system and prevent the initiation of cascade events, while corrective control focuses on stopping the progression of cascades. Corrective Transmission Switching (CTS) is one of the viable techniques for managing contingencies and preventing cascading failures in power systems. Conventional CTS methods often rely on optimal formulations or heuristic approaches for solving the corrective actions required to prevent cascading failures. However, in scenarios involving rapid failure propagation where temperature relays actuate swiftly, there may not be sufficient time to implement these corrective measures, making some solutions not useful in practical applications. To address this limitation, a modified optimal CTS can be used. This approach integrates generation re-dispatch and accounts for the available correction time derived from a continuous temperature evolution model. It involves validating CTS against the time constraints imposed by relay. Specifically, the temperature of transmission lines

is modeled using a thermal inertia framework, governed by the differential equation:

$$\frac{dT_o(t)}{dt} = \beta I^2 - \alpha(T_o(t) - T_\alpha) \quad (1.14)$$

where $T_o(t)$ is the temperature of line o at time t , I is the current, T_α is the ambient temperature with α and β are coefficients representing heat dissipation and generation rates, respectively. [27] also shows a modified optimal transmission switching (OTS) formulation with generation re-dispatch to mitigate post-contingency overloads and prevent cascading failures. By altering network topology through corrective transmission switching (CTS), the approach redistributes power flows to maintain system stability. Validation methods ensure that the corrective actions are feasible within relay operation times, enhancing resilience and cost-effectiveness.

Another method introduced in [28] is the Risk-Based Multi-Step Corrective Control (MSCC) method, which is designed to mitigate cascading failures by dynamically adjusting nodal power injections through generation rescheduling and load shedding. Unlike traditional Non-Recurring Corrective Control (NCC) models that confine corrective actions to specific cascading steps, the MSCC approach distributes control efforts across multiple steps, enhancing both economic efficiency and operational resilience. The MSCC model is grounded in a fault chain model combined with DC power flow equations, allowing for the prediction and interaction with cascading failure processes over several steps. Through numerical simulations on IEEE 39-bus and IEEE 118-bus systems.

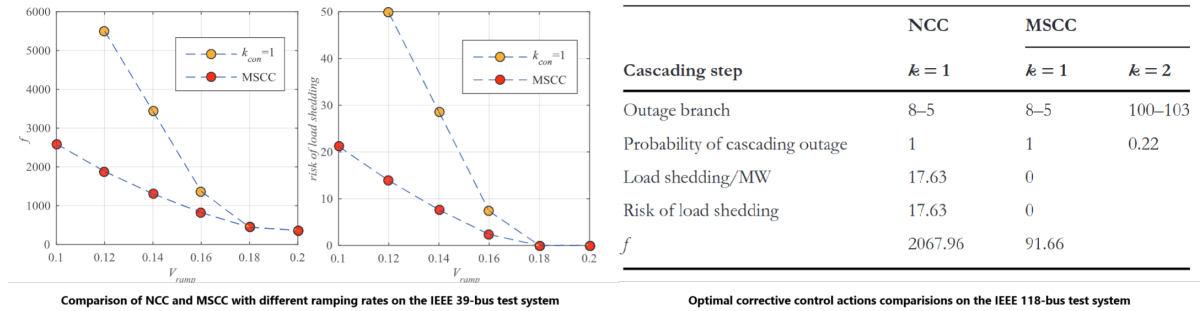


Figure 1.15: Comparison of NCC and MSCC models in the IEEE-39 and IEEE-118 bus systems [28]

Figure 1.15 compares the NCC and MSCC models in terms of load-shedding and economic costs for varying generator ramping coefficients (V_{ramp}) in the IEEE-39 bus system. As V_{ramp} increases, both load-shedding and the objective function decrease due to improved generation dispatch. For $V_{ramp} < 0.18$ in [28], the MSCC model outperforms NCC, which fails to prevent cascading outages when $V_{ramp} < 0.11$, as critical branch loadings exceed limits despite adjustments. At higher V_{ramp} , both models converge, with no load shedding and equal objective values, highlighting the importance of fast-ramping units for improved system flexibility and resilience. This analysis is extended to the IEEE-118 bus system, where the removal of Branch 8–5 triggers cascading failures. MSCC achieves significantly lower objective values (91.66 vs. 2067.96) and prevents load shedding, while NCC incurs 17.63 MW of load

shedding. In the second cascading step ($k = 2$), MSCC mitigates risks by removing Branch 100–103, reducing the probability of cascading outages to 0.22 and alleviating overloading on other branches. This demonstrates MSCC's superior coordination in minimizing risks, costs, and disruptions compared to NCC. Both studies emphasize the importance of integrating corrective measures with the dynamic nature of cascading failures and emphasize the role of power flow redistribution and thermal relay mechanisms in the propagation of cascades. In addition to these corrective strategies, understanding common patterns of initial failures leading to cascading events and developing robust frameworks are essential for comprehensive analysis.

1.2.4. Motifs and Probabilistic Frameworks for Risk Assessment

Use of PCMs

Risk assessment of cascading outages in power transmission systems is pivotal for enhancing system reliability and preventing large-scale blackouts. Central to contemporary methodologies are the concepts of PCM frameworks mentioned in [29], which collectively facilitate the identification and evaluation of high-risk outage scenarios. Motifs are specific patterns of multiple line outages that recur with higher frequency than expected under random conditions within a power network. They have been introduced as a means to improve the efficiency and accuracy of contingency selection for risk assessment. By analyzing historical outage data from large transmission systems such as the Bonneville Power Administration (BPA) and New York Independent System Operator (NYISO), [29] identified recurrent subgraph patterns, which are referred to as motifs that significantly exceed their uniform probability assumptions. For instance, analysis performed on the BPA network data highlighted that sub-graphs examples of multiple line outages and certain 2-edge and 3-edge sub-graphs occur disproportionately often, indicating inherent dependencies in the system's physical or engineered structure. These motifs serve as foundational elements in constructing a probabilistic model that accurately represents the likelihood of various cascading outage scenarios.

Probabilistic frameworks help in the quantitative assessment of cascading outage risks by modelling initial uncertainties, cascade propagation, and impacts. In the context of PCMs, a probabilistic framework can be used where the occurrence of specific motifs is incorporated into the risk assessment model. This approach involves estimating the probability of each motif based on historical data and using these probabilities to prioritize contingency lists for simulation. The probabilistic model leverages:

$$P(S_{k,i} \mid k) = \frac{|S_{k,i}|}{\binom{N}{k}} \quad (1.15)$$

where $P(S_{k,i} \mid k)$ denotes the probability of a particular subgraph pattern $S_{k,i}$ occurring among all possible k -edge subgraphs in a network with N nodes. Focusing on high-probability motifs enhances the efficiency of risk assessments, enabling more targeted and resource-effective simulations. Fur-

thermore, the importance of probabilistic approaches in capturing the stochastic nature of cascading failures is highlighted, advocating for methodologies that can systematically incorporate historical data and uncertainty quantification. PCMs align with these recommendations, offering a robust mechanism for prioritizing and analyzing high-risk contingencies.

Risk Assessment in Cascading Outages

Risk assessment methodologies must account for the probability of initiating events. It is emphasized in [30] that risk is defined as the product of the probability of an event and its associated impact cost. Mathematically, the risk $R(s)$ posed by contingency s , outage cost or impact $C(s)$ and probability $P(s)$ is expressed as:

$$R(s) = P(s) \times C(s) \quad (1.16)$$

An approach to estimating system risk is to select a subset of all possible contingencies and simulate each to estimate the resulting blackout. Aggregating system risk by summing individual risks can mix low-probability, high-cost events with high-probability, low-cost events, making it more useful to categorize risks by blackout size.

Moreover, there are various uses of risk assessment in the context of cascading outages. It aids in making informed decisions regarding actions to mitigate system stress, whether in real-time operations, operational planning, or investment planning. System operators can leverage these risks to predict and rank contingencies, which will be explored further in the thesis, enabling targeted mitigation measures such as adjusting protection settings or dispatching reserves. Additionally, [30] shows the critical contingencies that need to be addressed to ensure that system reliability can be identified. The PCM framework aligns with these uses by providing a structured and data-driven approach to identifying and prioritizing high-risk contingencies. This integration not only improves the accuracy of risk assessments but also ensures that resources are allocated efficiently to address the most significant threats to system reliability. Furthermore, leveraging motifs allows for a reduction in the computational complexity associated with simulating all possible outage scenarios. Instead of exhaustively enumerating every potential combination of line outages, PCMs prioritize those motifs that have historically demonstrated higher probabilities of occurrence, and this strategic sampling conserves computational resources while ensuring that the most critical and impactful outage scenarios are thoroughly analyzed.

1.3. Research Direction and Contribution

The literature about cascading failure analysis is extremely vast and has been experimented on with various techniques to improve accuracy. The review shows the overall progress of research till now, and it is summarized using the following points:

- The underreach or the overreach of the relays from its intended protection zones or hidden failures (HFs) can escalate even simple faults into continuous cascading scenarios.
- Techniques such as controlled islanding, DBSCAN with non-linear programming or a multi-layer spectral clustering along with load shedding can be used to analyse and mitigate cascade by leveraging the identification of cascade phase dynamics.
- Analyzing PMU on failure dynamics is crucial for monitoring the operational integrity and stability of power systems. It can provide various targeted approaches for grid operators to prioritize intervention strategies more effectively.
- Significant correlations between component failure rates, system demand, and external factors like wind speeds highlight how these factors can worsen system vulnerabilities and trigger a cascade leading to a blackout.
- ML techniques, such as LSTM-based fault direction and DNNs, can be used to predict and classify cascading failures with high accuracy. But these techniques face the challenge of class imbalance due to the limited availability of high cascade data and difficulty in interpreting decision making,
- Graph theory is one of the techniques to investigate cascades by utilizing metrics such as effective graph resistance (R_G), spectral radius (ρ), diameter and algebraic connectivity (λ_1) to assess grid robustness. These metrics enable the identification of critical components whose failure can cause widespread outages.
- Probabilistic models have emerged as essential tools in identifying blackouts by using the cascade contingency PDFs and CDFs, transition states and impact metrics to identify extreme vulnerability cases and pre-indicate the system operator of the presence.

This thesis proposes a risk metric-based ranking of N-k contingencies and the identification of critical cascade contingency scenarios. The approach allows for swift contingency ranking without the need for complex dynamic simulations or machine learning models, enhancing decision-making for Transmission System Operators (TSOs) and improving grid reliability and resilience under various operating conditions. Building on these conclusions, further research can explore promising methods for analyzing cascades. One approach involves pre-determining cascade contingency probabilities by first performing a contingency analysis to identify critical cascade scenarios. This can then be followed by identifying contingency motifs for each dispatch scenario and calculating the probabilities of these motifs to assess the likelihood of a cascade. Additionally, the development of an impact function could help quantify the effects of specific cascade contingency scenarios, considering factors such as topological

structure, line failures, and generator contributions.

We propose a *PCM-based risk metric*, calculated using a probabilistic model that incorporates cascade contingency probabilities based on $N - k$ contingency motifs, as described in [29]. These motifs represent frequent critical structural patterns in outage scenarios. The risk metric also includes a dynamically developed cascade impact function that considers distribution factors, electrical distances, and generator short-circuit current contributions. This approach utilizes the grid's topological data to predict the risk and extent of cascading events and rank contingencies accordingly.

Key contributions of this work include:

- **Identification of critical contingencies** through contingency analysis, enabling the prioritization of critical scenarios for further evaluation.
- **Calculation of cascade probabilities** based on contingency motifs, providing a clear measure of the likelihood of cascading events in various system states.
- **Development of a cascade impact function** that quantifies the effect of specific contingencies by considering distribution factors, electrical distances, and generator short-circuit contributions, leading to more accurate risk predictions.
- **Validation of the proposed risk metric** by comparing the cascade impact function's results with dynamic RMS simulations, such as the assessment of rotor angle deviation in case of cascade scenarios.
- **Evaluation of the computational efficiency** of the risk metric to ensure it is feasible for real-time application in system operations.
- **Comparison of risk rankings** to demonstrate the ability of the proposed approach to produce reliable and accurate cascade contingency rankings, which are essential for effective decision-making.
- **Analysis of generator dispatch** and its impact on the variability of risk, ensuring that the model can adapt to changing grid conditions and provide robust risk assessments.
- **Development of a plot** to help system operators quickly visualize high-risk scenarios, aiding faster decision-making to mitigate cascading failures.

These contributions lay the foundation for a more efficient, reliable, and adaptable approach to assessing cascading failures in power systems.

1.4. Research Questions and Structure of the Thesis

Following the research direction mentioned previously, we can outline the main research question of this thesis with several sub-questions:

How effectively can a probabilistic framework assess and rank $N - k$ cascading failure risk by leveraging contingency motifs and dynamic impact metrics?

- How can contingency motif subgraphs be identified and their probabilities be used to model and predict cascading failures?
- How does the proposed impact metric correlate with traditional impact metrics when ranking contingencies?
- How well does this risk-based probabilistic model's ranking of cascading events align with the rankings obtained by RMS simulation results in evaluating system security and resilience?

These questions outline the thesis, and they will be answered using the following structure.

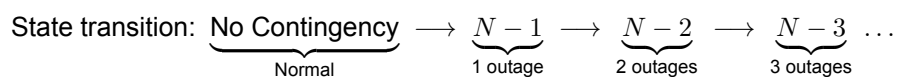
We summarize the necessary theory for the thesis in Chapter 2, where we provide an introduction to cascading failures in power systems, followed by an overview of contingency motifs and probabilistic modeling techniques for $N - k$ contingencies. We also introduce methods for risk assessment, including conventional impact metrics. In Chapter 3, we present the risk estimation methodology, detailing the probabilistic contingency model, the network structure-based probability analysis workflow, and the novel dynamic impact metric developed in this thesis, along with its implementation to calculate the risks for ranking cascading events. Chapter 4 focuses on case setup, simulations, and results, where we analyze the performance and obtained rankings of our proposed framework compared to baseline risk assessment methods. Finally, in Chapter 5, we provide a discussion and conclusion, addressing the research questions and highlighting the significance of our findings.

2

Theory

2.1. Cascading Outages in Power Systems

Cascading outages in power systems have drawn significant attention because of the devastating costs of large-scale blackouts. Unlike ordinary faults or single-element contingencies, cascading failures involve a chain reaction of events where the tripping of one component (e.g., a transmission line or generator) triggers additional stress and possible failures elsewhere in the system. During the last two decades, major blackouts in North America, Europe, Asia and beyond have demonstrated the complexity and unpredictability of these events [4]. The cascades consist of sequential outages where, in the initial state, they start from the position of no contingency and then proceed to show an outage, transitioning from $N-1$ (one line outage), potentially to $N-2$, $N-3$, and so on. For example, $k = 2$ corresponds to pairs of lines outaged at once or an $N-2$ contingency, whereas $k = 4$ involves quadruples of outaged lines or an $N - 4$ contingency scenario. At the occurrence of each step, additional stress is added to the remaining lines, resulting in the tripping of the remaining healthy lines one by one. An ineffective intervention or ignorance can significantly accentuate the occurrence of a cascade and eventually lead to a system-wide blackout. The sequence shown below in Figure 2.1 shows the state transitions of cascades.



Various physical and operational circumstances can cause a cascade to spread, the major one being

Figure 2.1: Sequence of cascade transition

the redistribution of power flows following the loss of a transmission line. When a line is lost due to a short circuit, natural disaster, poor maintenance, or infrastructure failure, the power flow redistributes across the grid to compensate. If this redistribution overloads other lines, it can trigger a cascading failure, spreading disruptions further. This redistribution occurs according to Kirchhoff's current and voltage laws. Further line failures or load shedding required are determined by checking if line limits are violated. A key aspect of power redistribution during cascading failures is the non-monotonicity of changes in line flows. Not all lines experience an increase in load; some may experience decreased flows, while others may even exhibit a reversal in the direction of flow. This behavior is shown in [31] and illustrated with two examples. The first example involves the impact of redistribution on adjacent lines. When a line fails, the resulting redistribution of power flows can either increase or decrease loading on nearby lines. If this redistribution causes new violations of operational limits, it can trigger further failures, propagating the cascade. The second example involves load shedding, which is used as an emergency action to relieve stress on the network. However, improper load-shedding can inadvertently shift power flows in a way that increases stress on certain critical lines, particularly those connecting multiple hubs within the network. Another way to quantify the redistribution of power flows is by using line outage distribution factors (LODFs), which represent the sensitivity of line flows to outages. LODFs are used in sensitivity analysis to assess the impact or cost of different cascade scenarios.

2.1.1. Overview and Mechanisms of Cascading

Cascades present multiple challenges, making their dynamics complex to study and even harder to simulate. They can be triggered by various factors, including natural disasters, human errors, cyber or physical attacks on power grids, or system failures. Here, the focus is on systemic failures, specifically short circuits, to perform contingency analysis and gather outage data. Simulations are also conducted using short circuits as failure events caused by systemic issues. One way to model external disturbances in this context is through a short-circuit model. According to the N-1 security rule, a power grid must remain electrically stable after the loss of any single component.

The cascade process, which is based on operational constraints and dynamics of the system, occurs in phases such as initiation, slow cascade, fast cascade and blackout. The 2003 U.S.-Canada blackout [32] was started by a single line trip due to contact with a tree, thereby triggering sequential overloads and thermal failures, leading to a cascade that ultimately affected 50 million people. Similarly, the 2003 Italy blackout occurred because of the import of high power, where quick cascades occurred in the form of multiple line trips, frequency collapse, and eventual islanding, highlighting the vulnerabilities of the interconnected system. In both cases, the slow cascade phase started with an initiating event, such as the loss of a network element, which led to power flow redistribution in the grid. This redistribution

increased thermal stress on other elements, raising their failure probabilities.

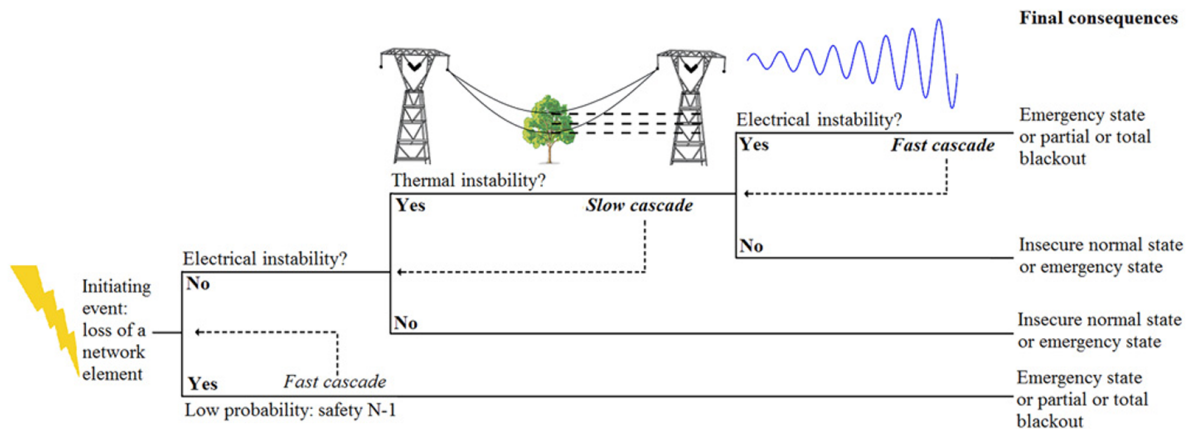


Figure 2.2: An event tree explaining cascade phases [32]

In Figure 2.2, thermal instability marks the transition into the slow cascade phase, where events occur on timescales ranging from tens of seconds to hours. During this phase, preventive actions by operators compete with the occurrence of additional contingencies. Independent external events can also arise, compounding cascades. Failures propagate gradually, resulting in minimal immediate impact on overall power system stability. However, this makes it harder to detect potential instability from future failures. Most of the failures in the slow cascade phase are often overlooked by operators, which diminishes the likelihood of timely preventive action. Additionally, HFs can cause the incorrect operation of critical equipment, such as circuit breakers. Such failures may function normally under stable conditions but fail when stressed by a disturbance. If a transmission line trips, nearby lines connected to its terminals may also become susceptible to misoperations or overloads to the redistribution of power flows. Overloaded lines may heat up due to thermal effects, sag, and eventually come into contact with trees or the ground, contributing to the cascade.

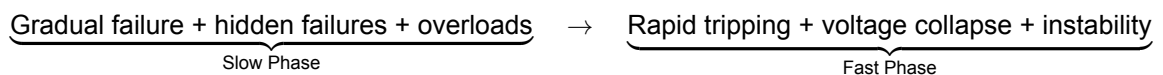


Figure 2.3: Cascade phase transition

Without corrective control, the system may escalate into the fast cascade phase. The fast cascade phase, characterized by electrical instability (e.g., protection system activations, angular instability) and successive tripping of overloaded transmission lines, occurs on a much shorter timescale, ranging from milliseconds to tens of seconds. The dynamic nature of this phase and extremely low time spans such as milliseconds to tens of seconds, illustrated in Figure 2.2, makes TSO interventions ineffective, resulting in widespread. However, TSOs can implement severe corrective actions, such as load shed-

ding, to mitigate the risk of blackouts during the fast-cascade phase. Instability events such as voltage collapse and frequency oscillations are observed in this phase. A generator tripping can cause this instability, further disturbing the load-generation balance and worsening the cascade. While transmission lines can typically withstand overloads for extended periods, dynamic events such as large power swings may trigger zone 2 or zone 3 protective relays. These relays, with short time delays, can trip transmission lines so quickly that operators have no opportunity to reclose the lines or halt the cascading sequence. As a result, failures propagate rapidly like falling dominoes, ultimately culminating in a widespread blackout.

When a system faces multiple line failures simultaneously in a cascade, particularly in the case of higher-order contingencies, it becomes increasingly vulnerable to widespread disruptions. These can cause large-scale load and generation shedding, further compromising the stability of the grid. Thus, they often give rise to an $N - k$ contingency scenario. The analysis of such a scenario helps in identifying critical lines whose simultaneous failure could lead to the most severe consequences, prompting the need for robust protection mechanisms and resource allocation to mitigate the risk of cascading outages. $N - k$ contingencies, therefore, provide a foundation for assessing the vulnerability of the system to cascading events.

2.1.2. $N - k$ Contingency Analysis and its importance

Controlling the power flow redistribution is essential to enhance the capacity of the system and protection against disturbances. Contingency analysis is a critical process used in power system operations to evaluate how the system behaves when a component (such as a generator, transmission line, transformer, or load) fails or experiences an outage. It is employed to understand power system behavior during equipment outages, ensuring security against single-component failures. Various contingency conditions can be analyzed, such as the loss of a generator, a transmission line, a transformer, or a load. In this thesis, the focus is on the analysis of contingency when a transmission line is under outage. $N - 1$ contingency analysis is a standard used by system operators to assess power system security. According to this criterion, the power system must be capable of withstanding the failure of any single component without violating the operational constraints of other components while continuing to supply both critical and non-critical loads. Contingency analysis models the effects of single or multiple outages to predict the state of various system variables, such as voltage, load, and power flows. This analysis is essential for evaluating the grid's ability to maintain stability and supply power during critical periods. Given the importance of grid security and reliability, system operators commonly apply $N - 1$ contingency analysis to ensure that the grid can remain stable even after the loss of any individual component. By simulating potential outages and analyzing the system's response, operators can identify vulnerabilities and prepare corrective actions to maintain reliable operation under stressed conditions.

Most of the time, the $N - 1$ contingency of the system may not be sufficient as multiple failures can take

place simultaneously. Here, the $N - k$ contingency analysis comes into the picture, which considers multiple component failures. The $N - k$ contingency implies that the power system should be able to withstand k-number of random/independent component failures without the violation of operational constraints of remaining healthy components. For this reliability analysis of power systems, both deterministic and probabilistic approaches can be used. In the deterministic approach, the focus is on analyzing specific events or worst-case scenarios. Each and every scenario is evaluated independently of the other to determine its associated impact, answering the question, "If event X occurs, what is the outcome Y?" The results are then assessed to ensure that they are acceptable under all scenarios, thereby identifying the most severe impacts. Deterministic methods rely on predefined criteria, such as ensuring specific components do not fail under conditions like typical lightning surges, as specified by standards. While deterministic criteria are simple to interpret and implement, their limitations are increasingly evident. They often result in over-designed systems by neglecting economic factors and risk considerations. For instance, questions such as "Should a remote customer be connected with $N1$ redundancy, or would they accept some level of risk for a reduced cost?" or "Is it necessary to connect offshore wind farms with $N1$ redundancy, or can operational risks be managed during maintenance under $N0$ conditions?" often arise.

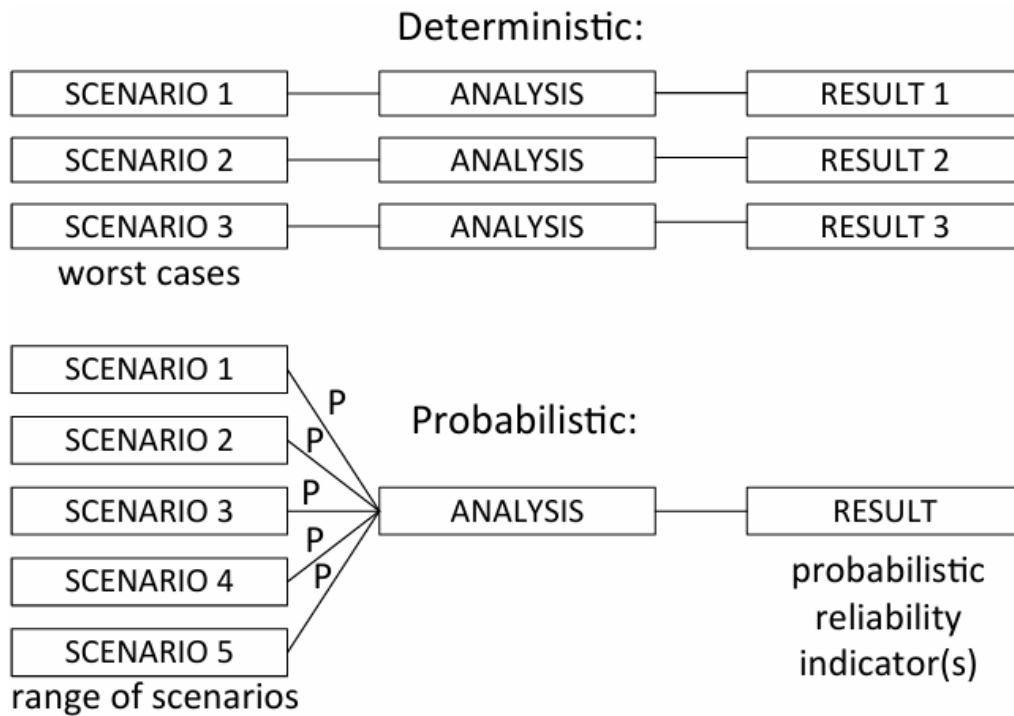


Figure 2.4: Deterministic versus probabilistic reliability analysis in power systems [33]

In contrast, probabilistic approaches incorporate the likelihood of the occurrence of various events, thereby allowing the analysis of a broader range of scenarios. Each scenario is assigned a probability (p), and the outcomes are combined into probabilistic reliability indicators as shown in Figure 2.4. This approach provides a more realistic analysis by capturing the variability of renewable generation,

such as high-wind and low-load scenarios, since their variability can be included in the probabilities by considering historical data. Considering the random nature of component failures in probabilistic analysis can prove vital. For instance, all possible generator failures can be included, rather than only worst-case scenarios like the loss of the largest unit. This approach addresses the shortcomings of the deterministic methods mentioned above by balancing the reliability of the system with economic (impact) considerations. For example, the approach allows system operators to make decisions by evaluating trade-offs between reliability and cost, such as whether to install a spare transformer in a substation or use underground cables instead of overhead lines. While deterministic criteria have historically proven effective for ensuring reliability and are widely prescribed in grid codes, they lack flexibility and do not account for economic and risk-based insights. Probabilistic approaches are an extension of deterministic methods and offer actionable reliability indicators that reflect real-world uncertainties.

2.1.3. Contingency and Power Flow Analysis procedure

Contingency analysis is best described in [34] as a systematic procedure used to determine possible power system states after component or line failure. It is primarily performed in specialized software within a power management system, which simulates hypothetical test cases that could result in power flow, voltage, or limit/capacity violations in active and reactive power. If P_{ij} is the power flow on the line between buses i and j , $F_{max,ij}$ can be considered as the transmission capacity limit. Contingency analysis is thus performed on some or all potential outages occurring in the system. There are three major steps to perform contingency analysis, as shown in [35]. These are contingency creation, contingency selection, and contingency evaluation. Primarily, the goal is to reduce an extensive list of contingencies by identifying those outages with serious limit violations. The individual outages are then evaluated by analyzing the power flows.

Contingency creation is the first stage of power system contingency analysis. Let

$$E = \{e_1, e_2, \dots, e_{n_L}\}$$

Denotes the set of all transmission lines in the system. A contingency is defined as a subset of \mathcal{E} , representing the components assumed to fail simultaneously. In the simplest single-outage scenario,

$$\mathcal{C}_{\text{single}} = \{\{e_1\}, \{e_2\}, \dots, \{e_{n_L}\}\},$$

While general multiple-outage contingencies appear in

$$\mathcal{C} = \{S \subseteq E\}.$$

A quantitative measure $\delta(S)$ characterizes each contingency S by tracking an indicator, such as max-

imum loading under that outage scenario. The extended contingency set pairs each subset S with its associated metric:

$$\hat{\mathcal{C}} = \left\{ (S, \delta(S)) \mid S \subseteq E \right\}.$$

The last component of each pair, $\delta(S)$, is compared against a threshold Θ to determine whether further analysis is required. The set of severe contingencies is given by

$$\mathcal{S} = \left\{ (S, \delta(S)) \in \hat{\mathcal{C}} \mid \delta(S) > \Theta \right\}.$$

A contingency $(S, \delta(S))$ moves into detailed evaluation when its metric $\delta(S)$ exceeds the threshold Θ . This approach provides a structured way to represent both the elements of each contingency and the associated loading or voltage measure needed to assess system reliability.

Any contingency causing one or more limit violations is classified as severe. The last stage of contingency evaluation is used to determine the extent of violations for these severe cases and to identify if subsequent outages (based on new system conditions) must be considered. Detailed power flow studies and short-circuit analyses are carried out here to confirm or quantify the violations [36]. The analysis is performed by first conducting power flow studies and then simulating short circuits after simulating the removal of system components such as transmission lines. Power flow studies are essential for assessing steady-state stability, ensuring that voltages and currents remain within acceptable limits. Short circuit studies, on the other hand, are used for protection and coordination purposes. These studies are typically performed after the power flow analysis, allowing for a more focused approach to evaluating and improving protection strategies.

Power Flow analysis using Newton-Raphson method

Power flow analysis is an essential starting point before contingency analysis is performed. Power flow involves determining the steady-state operating conditions of a power system, specifically the voltage magnitudes, phase angles, active power (P), and reactive power (Q) at each bus in the network. The Newton-Raphson (NR) method is one of the most widely used numerical methods for solving the non-linear equations arising in power flow studies due to its robust convergence properties and high computational efficiency. The power flow problem revolves around solving the non-linear algebraic equations derived from Kirchhoff's laws. The power injections at a bus i are given in terms of the voltage $|U|$ and the elements of the $|Y_{bus}|$ matrix:

$$P_i = \sum_{n=1}^N |U_i| |U_n| (G_{in} \cos(\theta_i - \theta_n) + B_{in} \sin(\theta_i - \theta_n)) \quad (2.1)$$

$$Q_i = \sum_{n=1}^N |U_i| |U_n| (G_{in} \sin(\theta_i - \theta_n) - B_{in} \cos(\theta_i - \theta_n)) \quad (2.2)$$

where for N number of total buses, P_i and Q_i are active and reactive power injections on bus i , $|U_i|$ and θ_i are the voltage magnitude and phase angle at bus i , G_{in} and B_{in} are the real and imaginary parts of the Y_{bus} matrix. The Y_{bus} matrix, or the nodal admittance matrix, represents the relationship between nodal voltages and injected currents. By applying Kirchhoff's Current Law (KCL) at each node i ,

$$I_i = \sum_{j=1}^N Y_{ij} V_j, \quad (2.3)$$

where I_i is the net current injected at node i , V_j is the voltage at node j , and Y_{ij} are the elements of the Y_{bus} matrix. The diagonal elements (self-admittances) are

$$Y_{ii} = \sum_{k \neq i} y_{ik} + y_i^{\text{shunt}}, \quad (2.4)$$

indicating that each node's diagonal entry is the sum of all admittances (including shunt components) connected to that node. The off-diagonal elements (mutual admittances) are

$$Y_{ij} = -y_{ij} \quad (i \neq j), \quad (2.5)$$

reflecting the negative of the admittance between nodes i and j . Each y_{ij} is typically derived from the per-unit impedance z_{ij} of the branch connecting i and j :

$$y_{ij} = \frac{1}{z_{ij}} = g_{ij} + j b_{ij}. \quad (2.6)$$

This structure produces a sparse matrix because $Y_{ij} = 0$ for buses i and j with no direct connection, and it is usually symmetric under reciprocal network conditions ($y_{ij} = y_{ji}$). Thus, Y_{bus} is both computationally efficient and serves as the foundation for describing how bus voltages and currents interact in a power system..

Power flow analysis classifies buses into three types [37]. If the voltage magnitude at bus i is denoted by $|U_i|$ and its phase angle by θ_i , and P_i and Q_i are the active and reactive power injections at bus i , respectively. The Slack/Reference bus is

$$|U_1| = V_{\text{slack}}^*, \quad \theta_1 = 0, \quad (2.7)$$

where V_{slack}^* is a specified reference voltage magnitude, and the angle is set to zero for a common reference frame. The slack bus accounts for real and reactive power imbalances in the system, so P_1 and Q_1 are determined by the overall network solution rather than being specified. The PV/Generator Bus is

$$P_i = \bar{P}_i, \quad |U_i| = \bar{V}_i, \quad (2.8)$$

where \bar{P}_i (active power) and \bar{V}_i (voltage magnitude) are specified. The unknown variables at this bus

are Q_i and θ_i , which the power flow equations solve for. The PQ/Load bus is

$$P_j = \bar{P}_j, \quad Q_j = \bar{Q}_j, \quad (2.9)$$

where \bar{P}_j and \bar{Q}_j are known active and reactive power demands (loads). In this case, $|U_j|$ and θ_j remain unknown and must be determined by the power flow algorithm.

The Newton-Raphson method iteratively solves the power flow equations by linearizing them around an operating point using a first-order Taylor series expansion [38]. The state vector is defined as

$$x = \begin{bmatrix} \theta_2 \\ \theta_3 \\ \vdots \\ \theta_N \\ |U_{m_1}| \\ |U_{m_2}| \\ \vdots \end{bmatrix}, \quad (2.10)$$

where θ_i are the voltage angles at each bus (excluding the slack bus angle, typically $\theta_1 = 0$) and $|U_{m_j}|$ are the voltage magnitudes of buses that are not fixed (e.g., PQ buses). In general, the exact composition of x depends on which bus variables are already specified (slack or PV buses) versus those that are unknown. The mismatch vector, $h(x)$, is the difference between specified and calculated power injections:

$$h(x) = \begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix}, \quad \Delta P = P_{\text{specified}} - P_{\text{calculated}}, \quad \Delta Q = Q_{\text{specified}} - Q_{\text{calculated}}. \quad (2.11)$$

In each iteration k , the state vector is updated according to

$$x^{(k+1)} = x^{(k)} - [J(x^{(k)})]^{-1} h(x^{(k)}), \quad (2.12)$$

where

$$J(x) = \frac{\partial h(x)}{\partial x} \quad (2.13)$$

is the Jacobian matrix of partial derivatives relating changes in the state variables to changes in the active and reactive power mismatches. The Jacobian matrix is typically partitioned into four submatrices:

$$J(x) = \begin{bmatrix} \frac{\partial P}{\partial \theta} & \frac{\partial P}{\partial |U|} \\ \frac{\partial Q}{\partial \theta} & \frac{\partial Q}{\partial |U|} \end{bmatrix}, \quad (2.14)$$

reflecting the sensitivities of P and Q with respect to the angles θ_i and voltage magnitudes $|U_i|$. For example, the partial derivative of the active power P_i at bus i with respect to the voltage angle θ_n of bus n is given by:

$$\frac{\partial P_i}{\partial \theta_n} = |U_i| |U_n| \left[-G_{in} \sin(\theta_i - \theta_n) + B_{in} \cos(\theta_i - \theta_n) \right], \quad (2.15)$$

where G_{in} and B_{in} are the real and imaginary parts of the Y_{bus} matrix element connecting buses i and n . Similar expressions apply for the other submatrices of $J(x)$, ensuring that the Newton-Raphson iteration captures how changes in voltage magnitudes and angles affect the calculated power injections.

The Newton-Raphson method converges quadratically, making it computationally efficient for large networks. However, it requires an accurate initial guess for the state variables to ensure convergence. The computational complexity arises from inverting the Jacobian matrix at each iteration, but this is mitigated by efficient numerical techniques such as sparse matrix factorization (e.g., LU factorization with pivoting) and optimized linear solvers that exploit the structure of power system matrices. Overall, the NR method provides a robust framework for solving power flows with high accuracy and reliability.

Once the power flow analysis is complete and the steady-state operating conditions are determined, contingency analysis can be applied by performing RMS simulations of short circuits and examining system transients. The outcomes of these simulations reveal potential violations of system constraints under fault conditions and estimate the probability of subsequent line outages or cascading events. These insights form a natural bridge to the next phase of investigation, where probabilistic uncertainties in multiple simultaneous outages are analyzed.

2.2. Probabilistic Modeling of $N - k$ Outages

In the first step of contingency selection, each single-line outage $c \in \mathcal{C}_{\text{single}}$ is assumed to be independent, and each operating condition $x \in \mathcal{X}$ is equally likely. Under a uniform assumption,

$$P(c, x) = \frac{1}{|\mathcal{C}_{\text{single}}| \cdot |\mathcal{X}|}, \quad (2.16)$$

where $\mathcal{C}_{\text{single}}$ denotes the set of single-line outages, and \mathcal{X} is a set of possible operating conditions or time instants. For the example of an $N - k$ contingency, the first outage is simulated and then the second outage is simulated considering the absence of the first outaged component, the third outage is simulated considering the absence of both the outaged components and so on. The definition of multiple contingencies ($N - k$) varies across contexts, leading to different interpretations of k in $N - k$. In this thesis, $N - k$ specifically denotes a contingency involving the outage of k transmission lines.

2.2.1. Subgraph representation of line outages

A convenient way to perform line outage simulations in a power transmission network is to view each outage combination as a subgraph of the underlying network graph. In this framework, the network is modelled as a graph $G = (J, E)$, where for a graph G , the set of nodes/junctions (J) and edges (E) corresponds to transmission lines. When $N - k$ outages are represented by choosing k transmission lines out of the E edges are outaged simultaneously. Suppose k is an integer satisfying $1 \leq k \leq |E|$. To capture a specific combination of k outaged lines, $E_{k,i}$ is defined as

$$E_{k,i} = \{e_1, e_2, \dots, e_k\} \subseteq \mathcal{E}, \quad (2.17)$$

where i indexes the different k -element subsets of E . Denoting by $J_{k,i} \subseteq J$ the set of nodes (buses) incident to the edges in $E_{k,i}$. Then the k -edged subgraph arising from these outaged lines is

$$s_{k,i} = (J_{k,i}, E_{k,i}). \quad (2.18)$$

Here, i runs over all possible ways to choose k edges from E , so each $s_{k,i}$ is a subgraph containing exactly k edges. Formally, let

$$\mathcal{S}_k = \{s_{k,i} \mid i \in \mathcal{I}_k\} \quad (2.19)$$

denote the collection of all edge-induced subgraphs of size k in G , where \mathcal{I}_k is an index set enumerating each distinct choice of k edges.

Two subgraphs are said to be isomorphic if there is a one-to-one relabeling of their nodes that preserves adjacency. If two subgraphs $s_{k,i}$ and $s_{k,j}$ are isomorphic, they belong to the same isomorphism class. Denoted here by $S_{k,\alpha}$,

$$S_{k,\alpha} \quad (\alpha \in \mathcal{A}_k) \quad (2.20)$$

the distinct isomorphism classes among the k -edged subgraphs, where \mathcal{A}_k indexes these classes. Each structural pattern $S_{k,\alpha}$ thus represents a unique configuration of k outaged lines in G , up to graph isomorphism. In addition to the topological structure of each subgraph, the functional connectivity of the network can be incorporated by utilizing an adjacency matrix that accounts for line loading. Let P_{ij} denote the power flow on the line between buses i and j , and let $F_{\max,ij}$ be its transmission capacity. A threshold parameter $\lambda \in (0, 1)$, chosen by the operator to indicate an acceptable fraction of line loading, is defined. The resulting adjacency matrix A is given by

$$A_{ij} = \begin{cases} 1, & \text{if } 0 < \frac{P_{ij}}{F_{\max,ij}} < \lambda, \\ 0, & \text{otherwise.} \end{cases} \quad (2.21)$$

An entry of $A_{ij} = 1$ signifies that the line (i, j) remains in service and is not heavily loaded (i.e., it

is below the threshold λ), whereas $A_{ij} = 0$ indicates either an outaged line or one that exceeds the permissible loading limit. The corresponding edges or elements A_{ij} are effectively set to zero, capturing the reduction in the network's functional connectivity when multiple lines are outaged or heavily loaded.

2.2.2. Representation of subgraph patterns

Building on this functional representation of line loading, subgraph patterns provide a systematic view to capture and categorize the network's evolving structure under various outage scenarios. A contingency motif is an outage subgraph set and refers to a recurring structural pattern that appears significantly more often than other possible outages and depends on various factors, such as frequency of occurrence and number of failed components (k). Formally, $s_{k,i}$ is a subgraph formed by the outage of k lines, where k is the number of edges in the subgraph and i indexes the various distinct subgraphs of size k , then

$$S_{k,i} = \{ s_{k,i}^{(1)}, s_{k,i}^{(2)}, \dots \} \quad (2.22)$$

is the set of all subgraphs in the network that are isomorphic to $s_{k,i}$. In a real power system, each outage configuration involving k lines can be mapped to one such isomorphism class.

In general, whenever an $N - k$ contingency occurs, it is visualized in the image by highlighting the pattern with k transmission lines (edges). The examples provided below are representative patterns $S_{k,i}$ that may appear most in the network graph. These are the crucial contingencies in a cascade from $N - 1$ to $N - 5$ because they are prioritized for detailed analysis. While the specific shapes and isomorphism classes depend on the topology of G , the following common subgraph types are broadly illustrative.

The mathematical descriptions in Table 2.1 represent different contingency patterns. $S_{1,1}$ describes a single edge (e) removal, where $e \in E$. For $k = 2$, $S_{2,1}$ represents two edges sharing a common node, i.e., $(e_1, e_2) \in E$ where $\{e_1, e_2\}$ share a node, whereas $S_{2,2}$ corresponds to two edges that do not share a node, meaning $(e_1, e_2) \in E$ but $\{e_1, e_2\}$ do not share a node. Moving to $k = 3$, $S_{3,1}$ is a three-edge star where all three edges meet at a single node, such that $(e_1, e_2, e_3) \in E$ and $\{e_1, e_2, e_3\}$ share a node. $S_{3,2}$ consists of two edges sharing a node while the third edge is disconnected, meaning $(e_1, e_2, e_3) \in E$, where $\{e_1, e_2\}$ share a node, but e_3 is disconnected. $S_{3,3}$ forms a path where each edge shares a node with the next, such that $(e_1, e_2, e_3) \in E$, where e_1 shares a node with e_2 , and e_2 shares a node with e_3 . $S_{3,4}$ represents a closed loop or triangle, where $(e_1, e_2, e_3) \in E$ form a cycle. In contrast, $S_{3,5}$ consists of three disconnected edges, meaning $(e_1, e_2, e_3) \in E$, with no shared node among them. For $k = 4$, $S_{4,1}$ represents four edges meeting at a single node, meaning $(e_1, e_2, e_3, e_4) \in E$ with all edges sharing a node. $S_{4,2}$ consists of three edges sharing a node while the fourth edge is either disconnected or forms a small chain, i.e., $(e_1, e_2, e_3, e_4) \in E$. $S_{4,3}$ consists of four edges with two common nodes, where three edges share one node and the fourth shares another, meaning $(e_1, e_2, e_3, e_4) \in E$. $S_{4,4}$ forms an open quadrilateral or 4-cycle, meaning $(e_1, e_2, e_3, e_4) \in E$ where the edges form an open


















Pattern	Visual	Outages	Description
k = 1: N-1 Contingency			
$S_{1,1}$		1	$e \in E$
k = 2: N-2 Contingency			
$S_{2,1}$		2	$(e_1, e_2) \in E, \{e_1, e_2\}$ share a node
$S_{2,2}$		2	$(e_1, e_2) \in E, \{e_1, e_2\}$ do not share a node
k = 3: N-3 Contingency			
$S_{3,1}$		3	$(e_1, e_2, e_3) \in E, \{e_1, e_2, e_3\}$ share a node
$S_{3,2}$		3	$(e_1, e_2, e_3) \in E, \{e_1, e_2\}$ share a node, e_3 disconnected
$S_{3,3}$		3	$(e_1, e_2, e_3) \in E, e_1$ with e_2, e_2 with e_3
$S_{3,4}$		3	$(e_1, e_2, e_3) \in E, \text{closed loop}$
$S_{3,5}$		3	$(e_1, e_2, e_3) \in E, \text{no shared node}$
k = 4: N-4 Contingency			
$S_{4,1}$		4	$(e_1, e_2, e_3, e_4) \in E, \text{all share a node}$
$S_{4,2}$		4	$(e_1, e_2, e_3, e_4) \in E, \text{three share a node, } e_4 \text{ forms chain}$
$S_{4,3}$		4	$(e_1, e_2, e_3, e_4) \in E, \text{three share a node, fourth shares another}$
$S_{4,4}$		4	$(e_1, e_2, e_3, e_4) \in E, \text{open loop}$
$S_{4,*}$	-	4	Other subgraphs of four edges

Table 2.1: PATTERNS FOR $N - k$ CONTINGENCY (LINE OUTAGE SUBGRAPHS)

loop. Finally, $S_{4,*}$ is an umbrella representation and includes all other possible four-edged subgraphs that are not explicitly covered by $S_{4,1}$ to $S_{4,4}$. As k increases, the number of isomorphism classes grows rapidly. The specific definitions of subgraphs with $k > 4$ are not considered due to their rare occurrence and the system operator's ability to spot a cascade before this and take appropriate action. These subgraph patterns provide insights into how real outages tend to cluster around certain key network features, such as single high-degree nodes (substations) or small loops, and can guide future analysis to improve contingency screening.

An example of such subgraph patterns is illustrated in Figure 2.5. It can be observed from the IEEE 39-bus power network schematically shown in the image and used in this study. Substations (or buses) correspond to the network nodes and are represented as blue circles with node numbers, and transmission lines correspond to the edges represented as black lines connected to these blue nodes.

A k -edge subgraph in the IEEE 39-bus network is defined by choosing k edges (lines) and the incident nodes. In Figure 2.5, the set of lines *Line* 01 – 02 and *Line* 01 – 39, which are lines between buses 1, 2 and 1, 39, forms a two-edge subgraph. This corresponds to an N-2 contingency and is a subgraph of pattern $S_{2,1}$ . Similar examples of N-3, and N-4 are represented in pink and green in figure 2.5 where the green one represents pattern $S_{4,3}$  while pink represents $S_{3,4}$ . Meanwhile, purple ($S_{3,1}$ ) and blue ($S_{3,3}$ ) represent other types of subgraphs that can occur when an N-3

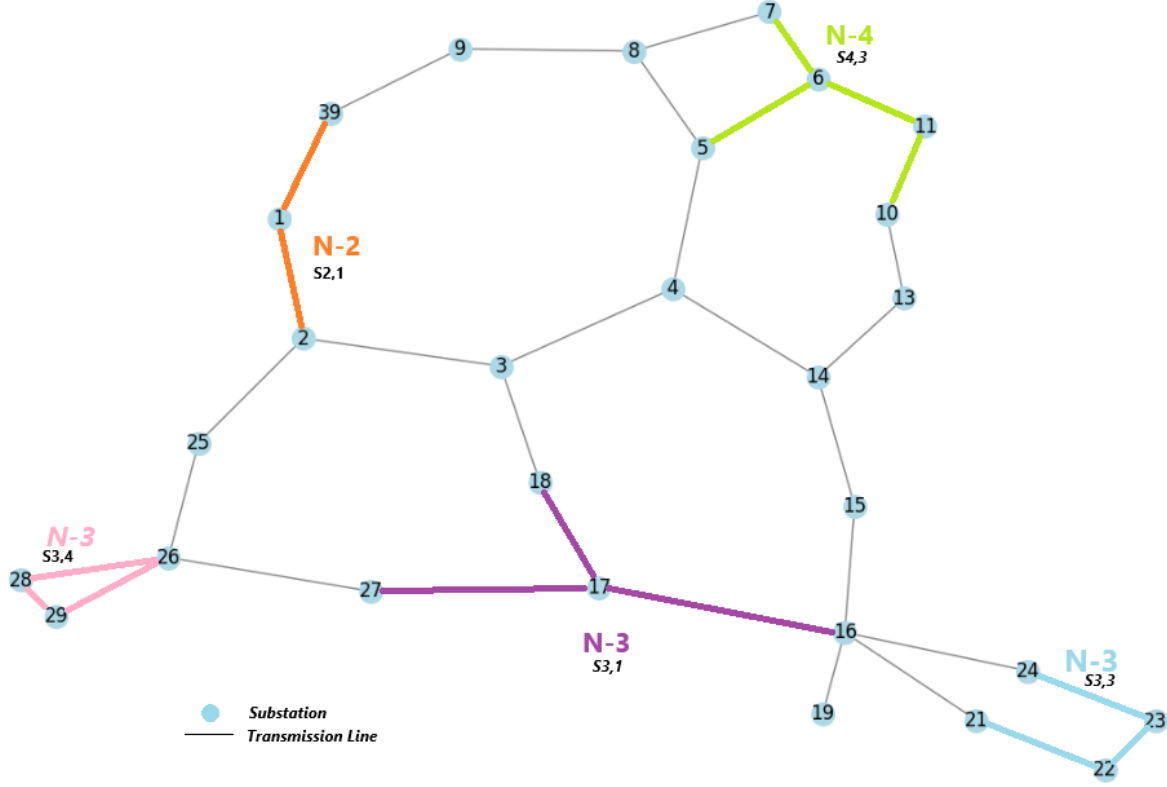



Figure 2.5: IEEE 39-bus network with highlighted subgraphs illustrating examples of $N - 2$, $N - 3$ and $N - 4$ line outages (layout is not geographic)

contingency happens.

In Figure 2.5, consider a subgraph induced by the edges *Line* 01 – 02, *Line* 01 – 39 and *Line* 02 – 03, *Line* 02 – 04. These edges might be arranged in the same pattern as , yet they may differ in which specific nodes of the IEEE39 system they connect.

2.2.3. Identification of an $N - k$ cascade contingency motif

In this subsection, using the probabilities of patterns corresponding to cascading scenario, motifs will be identified. The pattern of the outage is critical and needs to define the probability of the pattern $S_{k,i}$.

Uniform Probability

A simple baseline assumption is that, for a total number of buses N , there are $\binom{N}{k}$ possible k -edged subgraphs that can be formed, each corresponding to a potential set of transmission line failures. This assumption follows from the idea that each subset of k lines is equally likely to be outaged or a uniform distribution. This uniform probability of a specific subgraph occurring or $p_{S_{k,i}}^{\text{uni}}$ is given by:

$$p_{S_{k,i}}^{\text{uni}} = \frac{1}{\binom{N}{k}} \quad (2.23)$$

When we group isomorphic k -edge subgraphs into a pattern $S_{k,i}$, the probability for a given pattern under the uniform distribution is similarly defined. This reflects the equal likelihood of each pattern (or isomorphism class) being selected when a failure scenario occurs.

$$P^{uni}(S_{k,i} | k) = \frac{|S_{k,i}|}{\binom{N}{k}} \quad (2.24)$$

$|S_{k,i}|$ is used primarily as it is the number of distinct k -edged subgraphs in the isomorph class $S_{k,i}$.

Empirical Probability

In case of where the outage data, especially in [29], certain groups of transmission lines share the same pattern frequently than the uniform assumption would suggest. The probability of the subgraph can be estimated directly from the outage data using the two parameters n_k and $n_{k,i}$ and is given as:

$$P(S_{k,i} | k) = \frac{n_{k,i}}{n_k} \quad \text{where} \quad \sum_i n_{k,i} = n_k \quad (2.25)$$

where n_k is the count of total observed contingencies involving k lines and $n_{k,i}$ is the number of those n_k contingencies whose induced subgraph is $s_{k,i}$. Topological and operational factors could increase the probability of certain multi-line outages.

Adapting this idea to power grids, a motif is a k -edge subgraph ($S_{k,i}$) whose empirical probability ($P_{S_{k,i}}$) of occurrence exceeds its expected probability under the uniform assumption. If

$$P(S_{k,i} | k) > a \times P^{uni}(S_{k,i} | k) \quad (2.26)$$

then $S_{k,i}$ is a motif. The parameter a defines the threshold above which the observed occurrence of a pattern is considered statistically significant when compared to its expected frequency under the uniform baseline distribution. Bayesian hypothesis testing can determine if a subgraph falls under the bracket of a motif. The problem for a is given as

$$H_0 : P(S_{k,i} | k) \leq a \times P^{uni}(S_{k,i} | k) \quad (2.27)$$

$$\text{versus } H_1 : P(S_{k,i} | k) > a \times P^{uni}(S_{k,i} | k) \quad (2.28)$$

Rejecting H_0 indicates that the pattern's empirical frequency is unlikely to be explained by the uniform baseline alone, i.e., the pattern is a genuine motif.

2.2.4. Probabilities of multiple line outages

Out of all the motifs, certain subgraphs may be disconnected, indicating all those lines that do not share a common bus and are sometimes separated by several nodes. Disconnected motifs can span different physical regions of the network and, thus, have a larger diameter. The importance of disconnected subgraphs lies in their role in understanding power system vulnerabilities. They can indicate weakly connected regions of the grid, where failures may propagate differently compared to highly interconnected areas. The diameter of a subgraph ($d^{S_{k,i}}(E_x, E_y)$) is the largest distance in the network which separates the two lines x and y .

$$d^{S_{k,i}}(E_x, E_y) = \max_{x, y \in S_{k,i}} d(E_x, E_y) \quad (2.29)$$

Now, this network distance is defined using nodes, where the distance between two lines in a disconnected subgraph is given by the number of lines connecting the nodes that link these lines [39]. Empirical studies in [46] show that the diameter of a subgraph in a network follows Zipf's law, which describes a power-law relationship between an element's rank and its frequency. This means that the distribution of subgraph diameters exhibits a heavy-tail behavior, where smaller diameters are more common, while larger diameters occur with lower probability. However, despite their rarity, high-diameter subgraphs do appear, indicating that line outages can occur far apart within the network structure. Most of the multi-line outages occur among lines that are physically closer to one another because of the topological properties. These are the patterns that are more likely to show up as motifs.

Probability of k line Outages - $P(k)$

An important part of the model for determining multi-line outage probability is understanding $P(k)$, which represents the likelihood that exactly k transmission lines fail at the same time in an $N - k$ contingency event. In power grids, different k -line outages occur with varying frequency. For example, $k = 1$ or typical single-line outages are far more common than $k = 4$ or quadruple-line outages. Formally, estimating the rate of each k -contingency in historical or simulated data gives us:

$$P(k) = \frac{n_k}{\sum_l n_l} \quad (2.30)$$

where n_k is the counted or estimated number of k -line outages and the denominator sums all the l -line events considered. Thus,

$$P(k) = \frac{n_k}{n_2 + n_3 + n_4}, \quad k = 2, 3, 4 \quad (2.31)$$

The distribution for k is observed next, where a network is considered, and the outages are simulated to estimate their probabilities.

Probability of a pattern given k -line outages - $P(S_{k,i} | k)$

Within the network space, divided into k -line subgraphs, some of these subgraphs are motifs, denoted as $S_{k,i}$. Mathematically, let $\mathcal{G}_k = \{E_x \subseteq E \mid |E_x| = k\}$ represent the set of all k -line subgraphs, where E_x is a subset of k transmission lines from the set $E = \{e_1, e_2, \dots, e_{n_L}\}$ in the network. Thus, the set of all possible k -line subgraphs in the network is given by:

$$\mathcal{G}_k = \{S_{k,i} \subseteq E \mid |S_{k,i}| = k\} \quad (2.32)$$

The subgraph $S_{k,i}$ is a subset of \mathcal{G}_k , and it is characterized by its unique configuration in the network. The probability of a pattern given k -line outages can be denoted as $P(S_{k,i} | k)$, which is the empirical probability dependent on the count of total observed contingencies involving k -lines and the number of those contingencies whose induced subgraph belongs to the pattern $S_{k,i}$. Mathematically, this can be expressed as:

$$P(S_{k,i} | k) = \frac{n_{k,i}}{n_k} \quad \text{where} \quad \sum_i n_{k,i} = n_k \quad (2.33)$$

Probability of the contingency diameter given its pattern - $P(d | S_{k,i})$

In addition to identifying whether a particular subgraph $S_{k,i}$ belongs to a specific pattern $S_{k,i}$, incorporating diameter information can enhance the probabilistic model. This modeling step refines the multi-line outage representation by distinguishing not only whether lines fail in a given topological pattern but also how widely separated those lines can be in the network. The diameter of a subgraph provides valuable insight into the network's structural properties, such as the maximum distance between any two lines within the subgraph. This information is crucial because it helps quantify the spatial extent of the failure and how it might influence the propagation of disruptions throughout the network. The diameter of a subgraph ($d^{S_{k,i}}(E_x, E_y)$) is the largest distance in the network which separates the two lines x and y . Hence, a disconnected subgraph has $d > 1$, while a fully connected subgraph has $d = 1$. Thus, it can be assigned as,

$$P(d | S_{k,i}) = 1 \quad \text{for} \quad P(d | \text{connected}) \quad (2.34)$$

Other patterns that are disconnected form multiple clusters of 1 or more lines in separate parts of the network. In larger outages, some lines might be grouped around one substation, while others are far away, leading to subgraphs with multiple components. In such cases, the diameter can vary more widely. Empirically, to capture this variability, the $P(d | \text{disconnected})$ can be estimated from the observed data on the disconnected subgraphs. Specifically if there are $N_{\text{disconnected}}$ total disconnected

subgraphs in the data and $N_{disconnected,d}$ of those have a diameter d then

$$P(d \mid disconnected) = \frac{N_{disconnected,d}}{N_{disconnected}} \quad (2.35)$$

$P(d)$ can be the collection of all disconnected patterns.

Higher-order patterns denoted as $S_{k,*}$ have a relatively low probability $P(d \mid S_{k,i})$ compared to the others. This low probability often corresponds to rare occurrences, where the exact diameter d of a subgraph is difficult to predict due to its infrequency. In these cases, it is practical to simplify the model by treating these rare subgraphs similarly to connected ones, assigning a probability $P(d \mid S_{k,*}) = 1$ to represent that they are likely fully connected or exhibit a near-fully connected structure. Since $S_{k,*}$ rarely occurs, any small inaccuracy in the diameter calculations does not substantially affect the overall outage distribution.

In summary, the diameter conditional distribution of a pattern $S_{k,i}$ can be estimated by

$$P(d \mid S_{k,i}) = \begin{cases} 1, & S_{k,i} \text{ is connected subgraph,} \\ P(d \mid disconnected), & S_{k,i} \text{ is a disconnected subgraph,} \\ 1, & S_{k,*} \text{ is rarely observed subgraph.} \end{cases}$$

A probability of 1 indicates certainty, meaning the event always occurs. In this model, it reflects simplifications: for connected subgraphs, the diameter is fixed; for disconnected subgraphs and rare patterns, it's assumed deterministic based on observed behavior. Thus, assigning a probability of 1 helps simplify the modeling process.

Probability of a contingency given its pattern and diameter ($P(s_{k,i} \mid S_{k,i}, d)$)

The assumption that an individual subgraph $s_{k,i}$ belonging to a specific pattern $S_{k,i}$ with diameter d is uniformly distributed can be expressed mathematically as:

$$P(s_{k,i} \mid S_{k,i}, d) = \frac{1}{|S_{k,i}^d|} \quad (2.36)$$

where $|S_{k,i}^d|$ is the number of distinct k -edge subgraphs in pattern $S_{k,i}$ with diameter d . This is an approximate estimate gotten by enumerating or stochastically sampling subgraphs that meet the pattern and diameter criteria, as directly calculating the exact probability distribution can be computationally expensive and complex.

Total probability of a multi-line outage subgraph

The probability of a cascading failure involving multiple line outages can be formally expressed using conditional probabilities. For a set of outages, the probability of a cascading failure, where the failure of the first $k - 1$ lines causes the failure of the k^{th} line, can be written as:

$$P(S_k | S_{k-1}) = P(\{\ell_1, \ell_2, \dots, \ell_k\} | \{\ell_1, \ell_2, \dots, \ell_{k-1}\}) \quad (2.37)$$

where $S_k = \{\ell_1, \ell_2, \dots, \ell_k\}$ is the event where k^{th} line fails and $S_{k-1} = \{\ell_1, \ell_2, \dots, \ell_{k-1}\}$ is the event where the first $k-1$ lines fail. This conditional probability represents the likelihood of the k -th line failing given that the first $k-1$ lines have already failed, capturing the cascade effect where the failure of one line increases the probability of subsequent failures. The probability of observing any particular k -line subgraph $s_{k,i}$ in the network:

$$P(s_{k,i}) = P(k, S_{k,i}, d, s_{k,i}) \quad (2.38)$$

$$= P(k) \times P(S_{k,i} | k) \times P(d | S_{k,i}) \times P(s_{k,i} | d, S_{k,i}) \quad (2.39)$$

This systematically accounts for which pattern arises among the k -line outages and how far apart those outaged lines are in the network (d). Thus, multi-line outage events can be precisely characterized and modelled by partitioning the contingency space, first by k , then by pattern $S_{k,i}$, and finally by diameter, including those that form motifs or span widely separated parts of the grid.

2.3. Risk Assessment of Cascading Events

The stochastic and sequential nature of these events, coupled with growing system stresses from market dynamics, renewable integration, and operational policies, necessitates advanced techniques. Effective risk analysis must address computational challenges, quantify initial probabilities, and analyze the impact of failures to provide actionable insights. In real-time or near real-time operation, the risk assessment informs critical actions such as imposing power transfer limits, arming or disarming system integrity protection schemes, and adjusting relay settings to manage short-term risk while balancing speed and accuracy. During operational planning, actions like dispatching reserves to alter power flows, providing voltage support, warming up generating units, and preparing restoration plans can be made using the risk rankings. Over longer timescales, risk assessment aids in infrastructure investments, protection setting evaluations, regulatory adjustments, and system improvements.

The risk of a cascading outage is the expected value of impact considering two major factors: the probability of a k -line outage (cascade) and the impact/cost associated with the outage. The cascade impacts are estimated using time domain solutions of a disturbance. The risk assessment for cascading outages is crucial for enhancing the reliability of power systems. While security assessments, as required by standards such as NERC PRC-023 and FAC-011-2, focus on ensuring that operating limits are established to prevent cascading failures, they do not directly assess the associated risk value of such events. As [30] suggests, current tools for assessing and mitigating the risks of cascading outages are still underdeveloped.

2.3.1. Overview of Risk in Power Systems

Risk in power systems is fundamentally concerned with the uncertainty of future events and the potential impact of these events on the system performance. The challenge is the uncertainty of operation scenarios, where prospective scenarios are difficult to predict with certainty. Data on historical operations offers a retrospective view, providing insights into past system behaviour, but is limited in predicting future outcomes.

In the context of cascading failures, risk assessment involves understanding how initial failures, such as generator outages or line faults, may propagate through the system, potentially leading to large-scale outages. This requires a model of the system and the events that could occur within it, with performance measures incorporating the uncertainty inherent in future events to ensure reliability and system security. Reliability in this context is the ability of the system to meet the load demand without failure, while system security refers to the system's capacity to respond to faults and disturbances. The concept of risk arises from the possibility of failures within this system that may disrupt its functionality. With the liberalization of energy markets, power systems have become more fragmented, with various actors involved, including producers, consumers, transmission system operators (TSOs), and distribution system operators (DSOs). Each of these actors may have different views on risk. For example, losing stability is equivalent to a disruption in supply, which is prioritized by both the TSOs and consumers.

Understanding and managing risk in this context requires considering both the likelihood (probability) of an event occurring and its potential impact, which can vary from load losses and system security estimation to safety hazards or environmental consequences. In this context, where the perspective of a TSO is taken into account, the risk of the cascade is represented as a product of the probability of the occurrence of a particular cascade scenario and the impact that scenario causes:

$$\text{Risk} = \text{Probability} \times \text{Impact} \quad (2.40)$$

Since probability is dimensionless, the risk inherits the same units as the impact. As a result, risks can be directly compared to other impact metrics without any unit conversion.

2.3.2. Calculating Impact Metrics

Impact, in the context of cascading failures in power systems, refers to the effects of a specific set of line outages on the stability and performance of the grid. This impact metric can be assessed using key performance indicators such as generator short-circuit (SC) current contributions or peak rotor angle deviations. Based on [40] and its emphasis on dynamic security metrics, a suitable impact metric is proposed for estimating the risks using the probabilities and compared with another metric for verification.

3

Methodology

3.1. PCM for $N - k$ critical cascade risk rankings

The entire workflow is provided in Figure 3.1. First, a contingency analysis is performed to select critical contingencies based on a threshold. The selected contingencies are checked for motifs based on their uniform and empirical probabilities. By focusing on motifs, the most impactful and recurrent patterns are prioritized, significantly enhancing the efficiency of risk assessment. Once the cascade probabilities have been obtained, an impact function is formulated for each operating condition using the I_{SC} contributions of generators on the cascade, the electrical distance from the generators to fault location calculated using the modified Z-Bus matrix and the LODF sensitivities obtained with power flow redistribution. Both probability and impact are used to calculate risks of contingencies, which are ranked, and the contingencies with high risk are selected to perform dynamic security assessments and develop control strategies.

3.2. Network structure based probability analysis workflow

In [29], the authors perform a PCM analysis on the BPA transmission line network of 528 lines with data obtained from historical outages. The outcome was to obtain the multi-line outage probability of various k -edge subgraph patterns occurring in the network. This probability used various important calculations such as the probability of k line outages - $P(k)$, empirical probability of a pattern - $P(S_{k,i} | k)$, the probability that specific pattern $S_{k,i}$ has diameter d - $P(d | S_{k,i})$ and the probability of specific multiple contingencies given its pattern and diameter - $P(s_{k,i} | d, S_{k,i})$. The probability was calculated

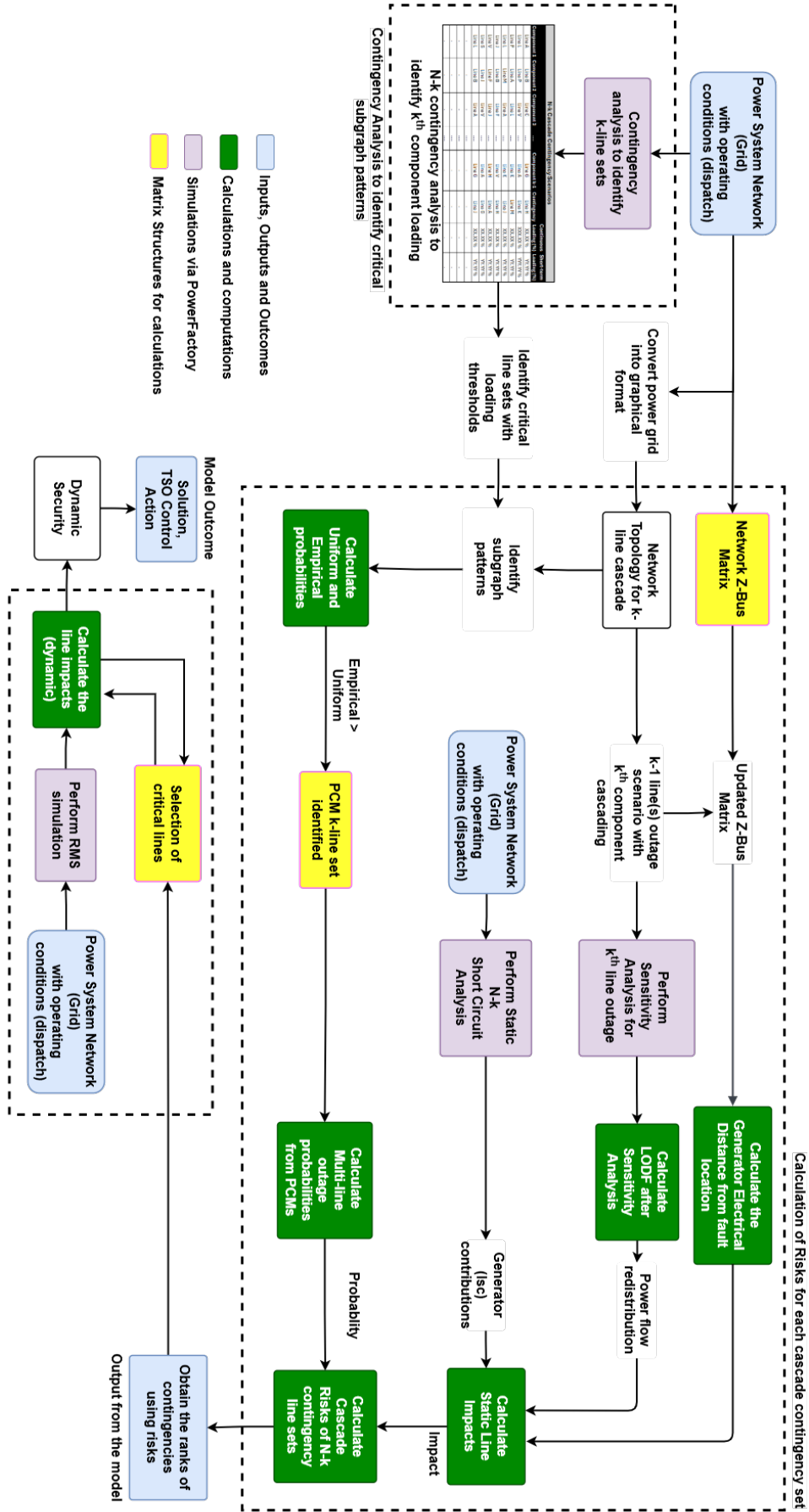


Figure 3.1: PCM Workflow

as

$$P(s_{k,i}) = P(k) \times P(S_{k,i} | k) \times P(d | S_{k,i}) \times P(s_{k,i} | d, S_{k,i}) \quad (3.1)$$

However, to consider a subgraph pattern as a contingency motif, the authors use uniform probability - $P^{uni}(S_{k,i} | k)$ and empirical probability - $P(S_{k,i} | k)$. The empirical probabilities are compared to the uniform probabilities, and only those subgraphs occurring in networks which have a higher empirical probability by a factor of a , are considered PCMs.

$$P(S_{k,i} | k) > a \cdot P^{uni}(S_{k,i} | k) \quad (3.2)$$

$$\frac{n_{k,i}}{n_k} > a \cdot \frac{|S_{k,i}|}{\binom{N}{k}} \quad (3.3)$$

The threshold factor **a** determines the deviation required for a pattern to be considered significant. Its choice determines the selection of contingency motifs: higher values of **a** filter out less significant patterns, while lower values include more patterns. This comparison highlights patterns that are not random but instead represent topological dependencies in the grid, making them critical for analyzing cascading failures. The objective here is to outline PCM formations with the help of graph theory and network structure. The probabilistic analysis is performed using the `NetworkX` package in Python. It is a Python package for the creation, manipulation, and study of complex networks of nodes and edges. It is preferably used here as it excels in handling large-scale graph-based problems by supporting various graph representations.

3.2.1. Performing contingency analysis

Power system contingency analysis plays a crucial role in assessing the resilience of the grid against outages. Here, the DigSILENT PowerFactory is used to perform contingency analysis [41], evaluating the impact of line outages on system stability. The process begins with defining contingency scenarios, where different levels of failures (N-1, N-2, N-3, etc.) are considered for E transmission lines in the network.

$$\begin{aligned} \text{N-1 contingency} &= \{e_i\}, \quad e_i \in E \\ \text{N-2 contingency} &= \{e_i, e_j\}, \quad e_i, e_j \in E, \quad i \neq j \\ \text{N-k contingency} &= \{e_1, e_2, \dots, e_k\}, \quad e_i \in E, \quad i = 1, 2, \dots, k \end{aligned} \quad (3.4)$$

Each scenario is analyzed using power flow calculations to identify overloaded transmission lines and quantify their loading percentages under contingency conditions. Once the contingency analysis is

finished on the whole network, the results are extracted into a structured database. The database records each affected transmission line along with its corresponding loading percentages (continuous and short-term) and the contingency name, which encodes the sequence of outaged lines (1 line in N-2, 2 lines in N-3 and so on). This structured dataset allows for a systematic evaluation of overloading patterns across different contingency levels. The extracted information is then used to filter high-risk scenarios based on continuous overloading thresholds, such as identifying components exceeding $\omega\%$ of their rated capacity.

The contingency database serves as the foundation for building a network representation of cascading failures. Using the `NetworkX` package, a graph $G = (J, E)$ is constructed, where J represents transmission lines and E captures dependencies between contingencies. The connectivity of G is given by the adjacency matrix A :

$$A_{ij} = \begin{cases} 1, & (i, j) \in E \\ 0, & \text{otherwise} \end{cases} \quad (3.5)$$

Similarly, the contingency-line relationships are stored in the incidence matrix B :

$$B_{ik} = \begin{cases} 1, & i \in k \\ 0, & \text{otherwise} \end{cases} \quad (3.6)$$

This network model represents the dependencies between affected components, enabling further identification and classification of contingency patterns.

3.2.2. Pattern Classification and identification of PCMs

The first step is to classify subgraphs within the network into various patterns ($S_{k,i}$) mentioned in the theory. Each subgraph represents a specific configuration within the network's structure, such as individual connections between specific nodes or common patterns. These patterns are identified within the graph to analyze dependencies and interactions among the affected components. The transmission network, represented as a graph $G = (J, E)$ with J nodes and E edges, is defined using the function `Graph()` in the `networkx` package. k -edged subgraphs ($s_{k,i}$) are extracted efficiently with methods `G.edge_subgraph(E)`. Once these subgraphs are extracted, they are classified according to predefined patterns based on their topological features. $s_{k,i}$ are then classified based on their characteristics such as connectivity to assign them their respective subgraph name using various functions such as `networkx.is_connected()` to check if the subgraph is fully connected and `networkx.spring_layout()`. After classifying the subgraphs $s_{k,i}$ into predefined patterns $S_{k,i}$, the frequency of each pattern in the graph G is recorded. Let $N(S_{k,i})$ denote the number of occurrences

of pattern $S_{k,i}$ in G . Then, we define:

$$N(S_{k,i}) = \sum_{s \in G} 1(s \cong S_{k,i}) \quad (3.7)$$

where $1(s \cong S_{k,i})$ is an indicator function that equals 1 if the subgraph s is isomorphic to the pattern $S_{k,i}$, and 0 otherwise. These counts $N(S_{k,i})$ serve as input for computations in subsequent steps.

Uniform Probability Calculation

The uniform probability calculation begins with determining the total number of possible k -edge subgraphs, denoted as N_k . The uniform probability of a specific k -edge subgraph $S_{k,i}$ occurring is given by:


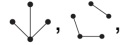
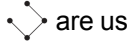
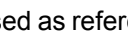
$$p_{S_{k,i}}^{\text{uni}} = \frac{1}{N_k} = \frac{1}{\binom{N}{k}} \quad (3.8)$$

When considering isomorphic k -edge subgraphs grouped into subgraphs $S_{k,i}$, the probability of a given subgraph under the uniform assumption is:

$$P^{\text{uni}}(S_{k,i} \mid k) = \frac{|S_{k,i}|}{N_k} = \frac{|S_{k,i}|}{\binom{N}{k}} \quad (3.9)$$

where $|S_{k,i}|$ represents the number of distinct k -edge subgraphs belonging to the isomorphism class $S_{k,i}$. This formulation ensures that each subgraph is weighted according to its occurrence frequency while maintaining the uniform probability assumption. This approach reflects the assumption that any subset of k edges is equally likely to be selected, leading to a uniform probability distribution over all possible failure scenarios.

Empirical Probability Calculation

The empirical probabilities are determined by analyzing unique subgraphs containing k edges in the network. The process begins by extracting all edges from the graph and generating combinations of edges of size k using the `itertools.combinations(edges, k)` module in Python, where `edges` is the list of all edges in graph G , obtained using `G.edges()` in NetworkX. To maintain uniqueness, each edge is included in only one subgraph. Once the unique k -edge subgraphs are identified, each subgraph is analyzed to determine its structural pattern based on node relationships. Patterns such as , , ,  are used as reference structures. For each identified pattern $S_{k,i}$, the count of its unique occurrences, $n_{k,i}$, is determined. The empirical probability $P(S_{k,i} \mid k)$ is then computed as:

$$P(S_{k,i} \mid k) = \frac{n_{k,i}}{n_k} \quad (3.10)$$

where n_k is the total number of observed patterns, calculated as the sum of all pattern counts in the subgraphs. In this context, a key distinction between empirical and uniform probabilities is the treatment of subgraph occurrences. The empirical approach focuses on analyzing unique subgraphs without

considering the sequence in which failures occur. In contrast, the uniform probability approach assumes an equal likelihood of all possible failure sequences, considering the order in which edges may be removed. This differentiation is crucial in interpreting the empirical results in contrast to uniform expectations.

Hypothesis testing for identifying PCMs

According to Equations 2.27 and 2.28, hypothesis hypothesis-testing methodology to detect PCMs is implemented programmatically. The null hypothesis H_0 assumes that the probability of observing a specific contingency pattern under empirical conditions is less than or equal to a uniform condition. Conversely, the alternate hypothesis H_1 is also defined. To implement this methodology programmatically, a Boolean function is defined to evaluate whether a given subgraph satisfies the motif condition. The function returns:

- **True**, if $P(S_{k,i} | k) > a \times P^{\text{uni}}(S_{k,i} | k)$, indicating that the subgraph is statistically significant or a PCM, rejecting H_0 and accepting H_1 .
- **False**, otherwise, implying that the subgraph does not meet the significance criterion, failing to reject H_0 .

This automated approach provides a reproducible and scalable framework for hypothesis testing across various k -edge subgraph patterns in large networks. The output of the function can be represented as a list of sets:

$$\mathcal{M} = \{S_{k,i} | P(S_{k,i} | k) > a \times P^{\text{uni}}(S_{k,i} | k)\} \quad (3.11)$$

or equivalently as a set of lines, where each line corresponds to a detected PCM.

3.2.3. Contingency enumeration and probability of line outages

The PCM set \mathcal{M} is used further to determine the probability of k -line outages $P(k)$ for each subset. This is the first critical component in estimating the probability of multi-line outage estimation. It is calculated using Equations

$$P(k) = \frac{n_k}{\sum_l n_l} \quad (3.12)$$

where n_k is the counted or estimated number of k -line outages and the denominator sums all the l -line events considered. Thus,

$$P(k) = \frac{n_k}{n_2 + n_3 + n_4}, \quad k = 2, 3, 4 \quad (3.13)$$

Contingencies involving more than four line outages ($k > 4$) are excluded from the analysis due to their very rare occurrence. In typical systems, the probability of such higher-order outages is extremely low because the conditions required for multiple line failures are highly unlikely. This inherent rarity is a consequence of the system's robustness and the protective measures in place, which prevent widespread disruptions. As such, excluding $k > 4$ from the analysis is justified in most practical scenarios.

Pattern-based probability of contingency-diameter

The probability $P(d \mid S_{k,i})$ is implemented using the `NetworkX` library by differentiating between connected and disconnected subgraphs. This is because, as explained earlier, this probability for a connected subgraph pattern $S_{k,i} \in S_C$ is constant and directly taken to be 1 due to its diameter being 1. For disconnected subgraphs $S_{k,j} \in S_D$, the diameter is calculated empirically. To achieve this, all possible subgraphs with up to k edges are generated from the main network G using combinations of edges. Subgraph connectivity is checked using `networkx.is_connected(subgraph)`. If the subgraph is connected, no further processing is needed; otherwise, the subgraph's diameter must be calculated.

For disconnected subgraphs, the diameter is determined by measuring the maximum shortest path distance between any two edges in the subgraph. This process starts by considering all pairs of edges in the subgraph and using `networkx.shortest_path(G, source=node1, target=node2)` to evaluate the shortest path between their nodes. If no path exists between the nodes, `networkx.NetworkXNoPath` exception is raised and caught, skipping the pair to prevent errors. For each pair of edges, the shortest path is computed, and the length of the path (measured in terms of the number of nodes) is considered the distance between the two edges. The smallest distance between any two nodes of an edge pair is recorded, and the maximum of these minimum distances across all edge pairs is taken as the diameter of the disconnected subgraph. The diameter d for a disconnected subgraph $S_{k,j}$ is given by:

$$diameter(S_{k,j}) = \max_{e_i, e_j \in S_{k,j}} \min_{v_i \in e_i, v_j \in e_j} dist(v_i, v_j), \quad (3.14)$$

where $dist(v_i, v_j)$ is the shortest path between nodes v_i and v_j , and the minimum is taken over all edge pairs e_i, e_j in the disconnected subgraph $S_{k,j}$. Finally, the probability $P(d \mid disconnected)$ for each diameter d is calculated by dividing the number of subgraphs with diameter d by the total number of disconnected subgraphs. In mathematical form:





$$P(d \mid disconnected) = \frac{N_d}{N_{total}}, \quad (3.15)$$

where N_d is the number of disconnected subgraphs with diameter d , and N_{total} is the total number of disconnected subgraphs.

Pattern and diameter based contingency probability

The primary assumption is that an individual subgraph $s_{k,i}$ belonging to a pattern group $S_{k,i}$ with a given diameter d is uniformly distributed. This can be expressed mathematically as:

$$P(s_{k,i} \mid S_{k,i}, d) = \frac{1}{|S_{k,i}^d|}, \quad (3.16)$$

The process begins by identifying disconnected subgraphs that match specific PCMs, such as $S_{2,2}$ () , $S_{3,2}$ () , $S_{3,5}$ () , and $S_{4,2}$ () . The disconnected subgraphs with $k \geq 4$ are not con-

sidered due to their rare occurrence, cascade identification before this moment, and the fact that they cannot be a PCM. This exclusion can be justified under any setup, especially in typical power system networks, where subgraphs with high edge counts are less likely to form power system contingency patterns. Additionally, these larger subgraphs are generally part of cascading events, which are outside the scope of initial PCM identification. This is done using the following combinatorial approach: `itertools.combinations(G.edges(), len(subgraph.edges()))`, where a specific pattern is targeted based on the number of nodes, edges, and the types of connections it possesses. For each such pattern, the diameter is calculated as:

$$diameter(S_{k,i}) = \max_{e_i, e_j \in S_{k,i}} \min_{v_i \in e_i, v_j \in e_j} dist(v_i, v_j), \quad (3.17)$$

The frequency of each diameter across the subgraphs is tracked. This frequency is the number of distinct subgraphs for the pattern $S_{k,i}$ with diameter d , denoted by $|S_{k,i}^d|$. The probability $P(s_{k,i} | S_{k,i}, d)$ for each diameter d is the inverse of this frequency, expressed as:

$$P(s_{k,i} | S_{k,i}, d) = \frac{1}{|S_{k,i}^d|}. \quad (3.18)$$

These probabilities are computed for all diameters in the range of interest (typically from 2 to 11), as diameters larger than these are of very rare occurrence in any typical power system network. Mathematically, this can be represented as:

$$P(d | S_{k,i}) = \frac{1}{|S_{k,i}^d|}, \quad \text{for } d \in \{2, 3, \dots\}. \quad (3.19)$$

This probability distribution provides insight into the likelihood of specific contingency patterns occurring in the system based on their diameter.

Multi-line outage probability or probability of a specific cascade scenario

Thus, using all four calculations, the multi-line probability for each specific PCM and for all diameter values from d_{min} to d_{max} is calculated as - $P(s_{k,i})$.

$$P(s_{k,i}) = \frac{n_k}{\sum_l n_l} \cdot \frac{n_{k,i}}{n_k} \cdot P(d | disconnected) \cdot \frac{1}{|S_{k,i}^k|} \quad (3.20)$$

This results in a matrix where for each type of subgraph pattern, the probabilities of outages for each unique diameter value based on the diameter of that specific subgraph are obtained.

3.3. Critical line impact assessment

This section provides the methodology for identifying the impact \mathcal{I}_f of experiencing a contingency of the selected critical lines using short circuit analysis and is aimed at identifying the potential risk of a specific set of cascade line outages.

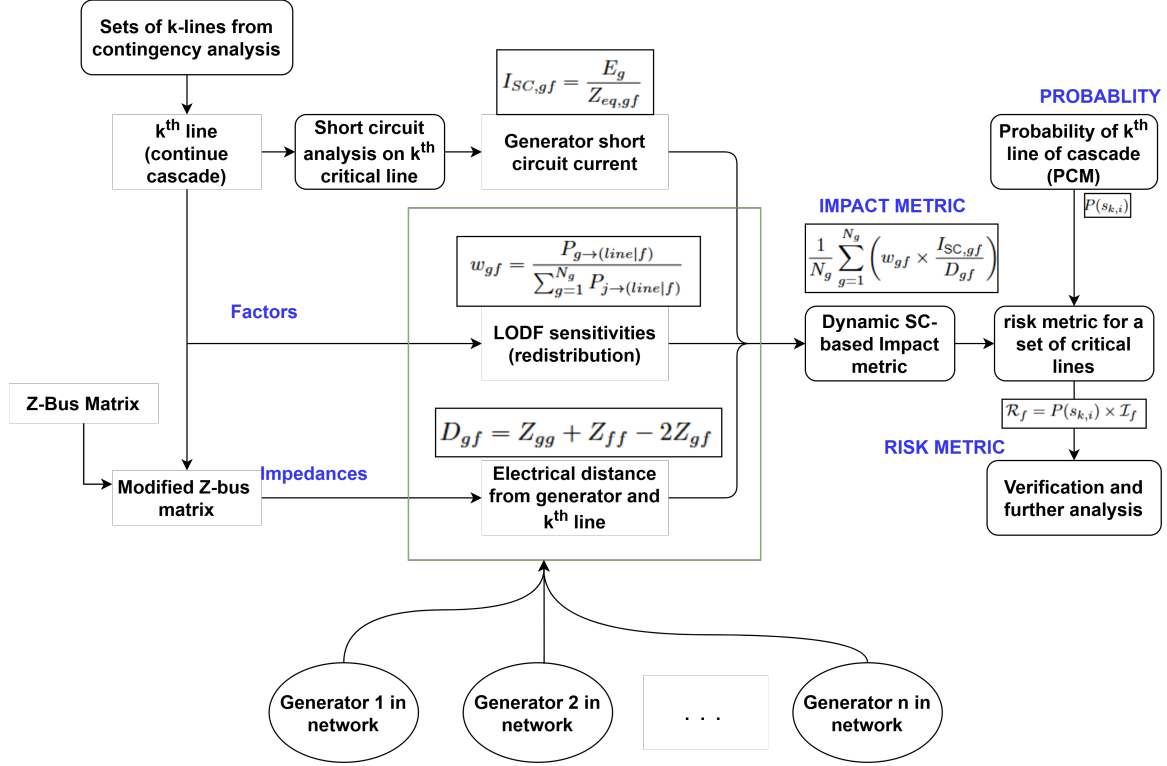


Figure 3.2: Flow of Impact \mathcal{I}_f calculation using critical lines, SC-simulations and Z_{bus}

The section shows systematic steps that provide a precise and quantifiable value of the impact \mathcal{I}_f of a cascade scenario. The method of calculating \mathcal{I}_f is provided in Figure 3.2. Once critical lines are identified using contingency analysis, simulated outages such as short circuit and switching events are introduced into the network model within DigSILENT PowerFactory. Python scripts automate the process of systematically removing these lines and performing short circuit analysis along with powerflows, sensitivity analysis and electrical distance calculations for obtaining \mathcal{I}_f and allowing for the examination of various outage scenarios for various dispatch conditions.

3.3.1. Calculation of impact for risk

The assessment of the impact \mathcal{I}_f of a specific fault or contingency scenario is a critical aspect of power system stability analysis. In this work, we introduce a novel short-circuit-based impact factor to measure and rank contingencies. Unlike traditional state-of-the-art metrics, our proposed approach explicitly incorporates the contribution of each generator to the short-circuit current, weighted by its

sensitivity and electrical distance from the fault location. This formulation provides a more intuitive and physically meaningful measure of the system's vulnerability to faults. To define \mathcal{I}_f , we first consider the Line Outage Distribution Factor (LODF), which quantifies the sensitivity of power flow redistribution following a line outage. Given a fault at location f , \mathcal{I}_f is computed by integrating the short-circuit contributions of all generators. Each generator's contribution is scaled according to its LODF sensitivity (weight) and electrical distance from the fault. The formulation is expressed as:

$$\text{Impact}_f = \frac{1}{N_g} \sum_{g=1}^{N_g} \left(\text{LODF}_{gf} \times \frac{I_{SC,gf}}{D_{gf}} \right) \quad (3.21)$$

$$\mathcal{I}_f = \frac{1}{N_g} \sum_{g=1}^{N_g} \left(w_{gf} \times \frac{I_{SC,gf}}{D_{gf}} \right) \quad (3.22)$$

Where N_g is the total number of generators in the system, $I_{SC,gf}$ represents the short-circuit contribution of generator g to the fault at location f , D_{gf} denotes the electrical distance between generator g and the fault location f and $w_{gf} = \text{LODF}_{gf}$ is the weight assigned to each generator's contribution based on its sensitivity. The short-circuit currents are computed using standard fault analysis techniques, considering the network topology, impedance, and pre-fault operating conditions. The electrical distance D_{gf} is typically determined based on impedance or Z-bus matrices, providing a measure of the strength of the electrical connection between the generator and the fault. Normalization by N_g ensures that \mathcal{I}_f remains independent of system size, enabling fair comparisons across different power networks. This is a significant advantage over conventional approaches that may inherently favor larger systems due to their greater number of generators and overall short-circuit capacity. By incorporating both sensitivity and electrical distance, our proposed metric provides a more comprehensive and physically meaningful ranking of contingencies, allowing system operators to prioritize faults that have the most severe impact on stability and reliability.

3.3.2. Generator short circuit contributions ($I_{SC,gf}$)

A critical aspect of fault impact assessment in power systems is quantifying each generator's short-circuit (SC) contribution to a faulted location. Larger SC contributions lead to higher fault currents and increased electrical stress on equipment. The magnitude of the SC current contributed by a generator depends on its internal impedance, its electrical distance from the fault, and the network topology. For a generator connected to bus g and a fault at location f , the SC current contribution is given by:

$$I_{SC,gf} = \frac{E_g}{Z_{eq,gf}} \quad (3.23)$$

where E_g is the internal generator voltage considering transient reactance, and $Z_{eq,gf}$ is the equivalent

impedance between generator g and the fault location f . The total fault current at f is the sum of contributions from all N_g generators:

$$I_{SC,total,f} = \sum_{g=1}^{N_g} \frac{E_g}{Z_{eq,gf}} \quad (3.24)$$

To assess \mathcal{I}_f , the absolute SC current values are combined with sensitivity factors (e.g., PTDF/LODF) and electrical distance metrics derived from the Z-bus matrix introduced in further subsections. This is performed over the list of branches and each line which can participate in the cascade and provides an overview of how short-circuit faults affect the system, including the contributions from individual generators and their relative impact based on their location in the network.

3.3.3. Weighting factors, LODF sensitivities (w_{gf})

To accurately estimate the overall \mathcal{I}_f of generators on an outage, it is essential to account for varying contributions. These can be captured using weighting factors derived from sensitivity metrics, such as Line Outage Distribution Factors (LODFs), computed through load flows and contingency analysis. LODFs are critical sensitivity metrics used to quantify how the power flow on other lines in the system is redistributed when a specific transmission line is taken out of service. The LODF for a line l under a contingency f is defined as:

$$\text{LODF}_{l \rightarrow l'}(f) = \frac{\Delta P_{l'}(f)}{\Delta P_l(f)} \quad (3.25)$$

where $\Delta P_{l'}(f)$ is the change in power flow on line l' due to the outage of line l , and $\Delta P_l(f)$ is the change in power flow on line l due to the outage of line l . To compute the generator's contribution to the line's power flow during the contingency, we use the LODF to adjust each generator's contribution based on how much the outage of line l affects the overall system. The active power contribution from generator g flowing into line l under a contingency f , adjusted by the LODF for that line, is denoted as:

$$P_{g \rightarrow l}(f) = P_g(f) \cdot \text{LODF}_{l \rightarrow l'}(f) \quad (3.26)$$

where $P_g(f)$ is the active power output of generator g under the fault scenario f .

The weighting factor w_{gf} for a generator g when a fault occurs at point f on a specific line l reflects its relative contribution to the power redistribution during that contingency. This is calculated by dividing the generator's adjusted contribution by the total active power contribution from all generators to line l .

$$w_{gf} = \frac{P_{g \rightarrow l}(f)}{\sum_{g=1}^{N_g} P_{g \rightarrow l}(f)} \quad (3.27)$$

This ensures that the weights are normalized:

$$\sum_{g=1}^{N_g} w_{gf} = 1 \quad (3.28)$$

To calculate the \mathcal{I}_f of all generators in the network for a specific fault scenario f , we integrate the contributions of all generators in the system. The total short-circuit current contribution from all generators, adjusted by the LODF weighting factors, can be expressed as:

$$I_{\text{total}}(f) = \sum_{g=1}^{N_g} w_{gf} \cdot I_{SC,gf} \quad (3.29)$$

where $I_{SC,gf}$ is the short-circuit current contribution from generator g under the fault scenario f . This method ensures that the operational dynamics of all generators are reflected accurately in \mathcal{I}_f during the contingency. Further, the electrical distance using the Z_{bus} matrix is introduced.

3.3.4. Iteratively building the Z_{bus} matrix

The bus impedance matrix (Z_{bus}) represents impedances between nodes and is essential for understanding power network behaviour during faults, assessing stability, and modeling the network efficiently. The Z_{bus} represents the relationship between the voltage at each bus (node) and the currents injected into the system. Each element (z_{ij}) in this matrix shows the equivalent impedance between two buses. The Z_{bus} matrix is essential for fault analysis, as it helps determine how faults propagate through the system, and for stability analysis, as it reveals the system's response to disturbances [42]. The Z_{bus} matrix is obtained using prior information about the power system, such as the network layout and the transmission line characteristics.

The complex impedance of each transmission line is extracted as $Z = R + jX$ from the network. Each bus is represented in a vector $E = [e_1, e_2, \dots, e_n]$ to facilitate matrix computations. Given sorted line data, a network graph $G = (J, E)$ is constructed, with impedance weights Z_{ij} . The Z_{bus} matrix is iteratively built by processing edges in the order they appear in E , ensuring correctness in incremental impedance updates. Each row corresponds to a bus index, ensuring unique references without explicit numbering. The edge sequence is maintained in an ordered list to preserve consistency in updates.

Z-Bus Matrix Construction using Kron's reduction

The iterative construction of the Z-Bus matrix is done by progressively adding branches (transmission lines) and updating impedance values. This iterative process is broken down into the following cases, where each case represents a unique addition of an impedance element into the matrix.

Algorithm 1 Z-Bus Matrix Construction using Kron's Reduction

```

1: Input:  $G = (J, E)$ ,  $E = [e_1, e_2, \dots, e_n]$ 
2: Output:  $Z_{bus}$ 
3: Initialize:  $Z_{bus} \leftarrow \emptyset$ 
4: while  $\exists(i, j) \in E$  do
5:   Select  $(i, j) \rightarrow Z_{ij}$ 
6:   if  $j = 0$  then ▷ Case 1
7:      $Z_{bus} \leftarrow \begin{bmatrix} Z_{bus} & 0 \\ 0 & Z_{ij} \end{bmatrix}$ 
8:   else if  $j \in Z_{bus}$  then ▷ Case 2
9:      $Z_{kk} \leftarrow Z_b + Z_{jj}$ 
10:     $Z_{bus} \leftarrow \begin{bmatrix} Z_{bus} & Z_{:,k} \\ Z_{k,:} & Z_{kk} \end{bmatrix}$ 
11:   else if  $k \in N_0$  then ▷ Case 3
12:     $Z_{bus} \leftarrow \begin{bmatrix} Z_{bus} & Z_{:,k} \\ Z_{k,:} & Z_{kk} + Z_b \end{bmatrix}$ 
13:    Remove temporary bus  $p$ :  $Z_{bus} \leftarrow Z_{bus} - \frac{Z_{:,p}Z_{p,:}}{Z_{pp}}$  ▷ Kron's Reduction for  $p$ 
14:   else ▷ Case 4
15:     $Z_{bus} \leftarrow \begin{bmatrix} Z_{bus} & Z_{:,l} - Z_{:,k} \\ Z_{l,:} - Z_{k,:} & Z_{ll} + Z_{kk} - 2Z_{lk} + Z_b \end{bmatrix}$ 
16:    Remove temporary bus  $q$ :  $Z_{bus} \leftarrow Z_{bus} - \frac{Z_{:,q}Z_{q,:}}{Z_{qq}}$  ▷ Kron's Reduction for  $q$ 
17:   end if
18: end while

```

1. **Case 1 - New Bus Connected to the Reference Node:** The first line is introduced with two starting nodes. It forms a 1×1 matrix containing only its impedance value. This serves as the foundation for expanding the matrix as new lines are added. When a new bus is connected to the reference node through an impedance Z_b , a new row and column are added to the existing impedance matrix Z_{old} .

$$Z_{new} = \begin{bmatrix} Z_{old} & 0 \\ 0 & Z_b \end{bmatrix}$$

This maintains sparsity in the impedance matrix while initializing the connection of the new bus.

2. **Case 2 - New Bus Connected to an Existing Bus:** If a new bus is connected to an existing bus k (which is not the reference), the impedance matrix is updated by adding a new row and column at the end. The corresponding row and column values are extracted from the impedance matrix for the existing bus k . With Z_{kk} being the self-impedance of bus k before adding the new bus, the self-impedance is updated as:

$$Z_{kk}^{new} = Z_b + Z_{kk}$$

The updated impedance matrix takes the form.

$$Z_{\text{new}} = \begin{bmatrix} Z_{\text{old}} & Z_{:,k} \\ Z_{k,:} & Z_b + Z_{kk} \end{bmatrix}$$

where $Z_{:,k}$ and $Z_{k,:}$ represent the respective full column and row of bus k in Z_{old} .

3. **Case 3 - Existing Bus Connected to the Reference Node:** When an existing bus k is connected to the reference node through Z_b , a temporary new bus is added, which is later removed using Kron reduction.

$$Z_{\text{new}} = \begin{bmatrix} Z_{\text{old}} & Z_{:,k} \\ Z_{k,:} & Z_{kk} + Z_b \end{bmatrix}$$

Then, Kron's reduction given in [43] is applied by eliminating the newly introduced row and column, modifying the impedance matrix accordingly and bringing it to the previous configuration as the new node is not added. With Z_{pp} as the impedance at the newly added node before elimination.

$$Z' = Z_{\text{new}} - \frac{Z_{:,p}Z_{p,:}}{Z_{pp}}$$

4. **Case 4 - Existing Bus Connected to Another Existing Bus:** If an existing bus k is connected to another existing bus l via Z_b , a temporary bus q is introduced to represent the connection. The impedance matrix is first expanded as

$$Z_{\text{new}} = \begin{bmatrix} Z_{\text{old}} & Z_{:,l} - Z_{:,k} \\ Z_{l,:} - Z_{k,:} & Z_{ll} + Z_{kk} - 2Z_{lk} + Z_b \end{bmatrix}$$

where Z_{ll} , Z_{kk} , Z_{lk} represent the self and mutual impedances of buses l and k . Following this, Kron reduction as explained in Case 3 is applied to eliminate the temporary bus q , refining the impedance matrix.

Example of Kron's Reduction:

Assuming the following initial impedance matrix Z_{old} and the new connection where an existing bus k is connected to the reference node:

$$Z_{\text{old}} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

After adding bus k , the impedance matrix Z_{new} becomes:

$$Z_{\text{new}} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{1k} \\ Z_{21} & Z_{22} & Z_{2k} \\ Z_{k1} & Z_{k2} & Z_{kk} + Z_b \end{bmatrix}$$

Where Z_{1k} , Z_{2k} , Z_{k1} , Z_{k2} represent the impedance elements between the existing buses and the new

bus k . When this bus k is removed using Kron's reduction, the elements $Z_{1k}, Z_{2k}, Z_{k1}, Z_{k2}$ are eliminated, and the values of the remaining impedance elements are adjusted. If we let Z_{kk} and Z_b be the self-impedance and the impedance of the new connection, the final reduced matrix is obtained, and the updated values represent the network's impedance without the reference.

$$Z' = \begin{bmatrix} Z_{11} + \frac{Z_{1k}Z_{k2}}{Z_{kk}} & Z_{12} + \frac{Z_{1k}Z_{k2}}{Z_{kk}} \\ Z_{21} + \frac{Z_{2k}Z_{k1}}{Z_{kk}} & Z_{22} + \frac{Z_{2k}Z_{k1}}{Z_{kk}} - \frac{Z_{1k}Z_{k2}}{Z_{kk}} + Z_b \end{bmatrix}$$

In large-scale power systems, there may be numerous buses that are connected to a reference or "dummy" node in Case 2 of the algorithm. Kron's reduction allows us to eliminate such temporary buses and simplify the impedance matrix.

The algorithm iterates through a predefined order of edges, classifying them into the above cases based on bus connectivity conditions. It initializes the Z_{bus} matrix with the first encountered impedance and incrementally updates it by applying the corresponding case. The nodes are tracked in a list to determine connectivity and decide the case for each new edge. The values from the matrix are used to calculate the electrical distance used in \mathcal{I}_f . The workflow in Python is in Appendix B.

3.3.5. Electrical distance between fault location and generator (D_{gf})

A key differentiator in this approach is the use of D_{gf} in \mathcal{I}_f . By dividing the SC contribution ($w_{gf} \times I_{SC_{gf}}$) by D_{gf} , \mathcal{I}_f shows how the faults are present in-depth of the grid. This reduces the chance of overestimating the influence of a faraway generator and yields a more realistic sense of which generators or lines truly drive cascading failures. Higher D_{gf} reduces the fault current contribution. Mathematically, this distance can be found by examining entries of the system impedance matrix (Z_{bus}) and through line-by-line calculations of self and mutual impedance using matrix operations. The electrical distance (D_{gf}) of a generator bus g relative to a particular fault location f can be determined using Thevenin's equivalent impedance:

$$D_{gf} = Z_{gg} + Z_{ff} - 2Z_{gf} \quad (3.30)$$

where Z_{gg} is the self-impedance of the generator bus g and Z_{ff} is the self-impedance of the fault location f , which is calculated from the modified Z_{bus} matrix considering f as a new bus and following the algorithm again. Both self-impedances are obtained from the diagonal elements of the new Z_{bus} matrix. Z_{gf} is the mutual impedance between generator bus g and the fault location new bus f .

The line impedance of the fault location connecting bus i and j is defined as Z_{line} , and the self and mutual impedance of these buses is Z_{ii} , Z_{jj} , and Z_{ij} obtained from the Z_{bus} matrix. If p is the fractional distance of the fault location from bus i towards bus j , the self-impedance at fault location f is

$$Z_{ff} = (1-p)^2 Z_{ii} + p^2 Z_{jj} + 2p(1-p)Z_{ij} + p(1-p)Z_{line} \quad (3.31)$$

This formula interpolates the impedance contributions from the buses and the line to the fault point based on its location. For a fault at the midpoint ($p = 0.5$), the formula simplifies to

$$Z_{ff} = \frac{1}{4}(Z_{ii} + Z_{jj}) + \frac{1}{2}Z_{ij} + \frac{Z_{line}}{4} \quad (3.32)$$

For calculating mutual impedance, we use the already existing mutual impedances between generator bus g and the buses i and j , obtained from the modified Z_{bus} matrix.

$$Z_{gf} = (1 - p)Z_{gi} + pZ_{gj} \quad (3.33)$$

For a fault at the midpoint ($p = 0.5$), the formula simplifies to

$$Z_{gf} = \frac{Z_{gi} + Z_{gj}}{2} \quad (3.34)$$

The formula arises from the concept that the fault location is effectively a new point on the line, and its impedance contribution can be interpolated between the two endpoints (buses) of the line under fault. Thus, electrical distance D_{gf} between fault location f on line l between bus i and j and generator g is:

$$D_{gf} = Z_{gg} + (1 - p)^2 Z_{ii} + p^2 Z_{jj} + 2p(1 - p)Z_{ij} + p(1 - p)Z_{line} - 2 \times \frac{Z_{gi} + Z_{gj}}{2} \quad (3.35)$$

3.4. Risk rankings and verification

The risk for fault at location f on line l defined as \mathcal{R}_f of a specific contingency is then calculated by multiplying the final multi-line probability for each specific contingency pattern ($P(s_{k,i})$) with the overall impact \mathcal{I}_f of that specific contingency scenario for fault at location f .

$$\mathcal{R}_f = P(s_{k,i}) \times \mathcal{I}_f \quad (3.36)$$

Once the risks are calculated, each set of contingencies is ranked based on the risk value to identify the most critical contingencies. This ranking is crucial in risk management as it allows TSO operators to prioritize resources and focus on the most critical issues first. High-risk contingencies often require immediate attention and substantial mitigation strategies, whereas low-risk contingencies may be monitored with minimal intervention. This methodology is widely applied in fields such as business continuity planning, scenario analysis, and performance monitoring, helping TSOs allocate resources efficiently and develop proactive strategies to reduce vulnerabilities. Resource allocation also heavily relies on this ranked list, as it allows for the efficient distribution of limited resources. By focusing on the highest-risk contingencies, TSOs can prevent potential disruptions from escalating into major issues. Conversely, lower-priority risks can be addressed as needed, or preventive measures may be

implemented to reduce their likelihood. Ultimately, ranking contingencies by their risk value supports proactive decision-making, strengthens resilience, and ensures that efforts are concentrated where they are most needed to protect the integrity and success of operations.

3.4.1. Verification of risks and associated contingency rankings

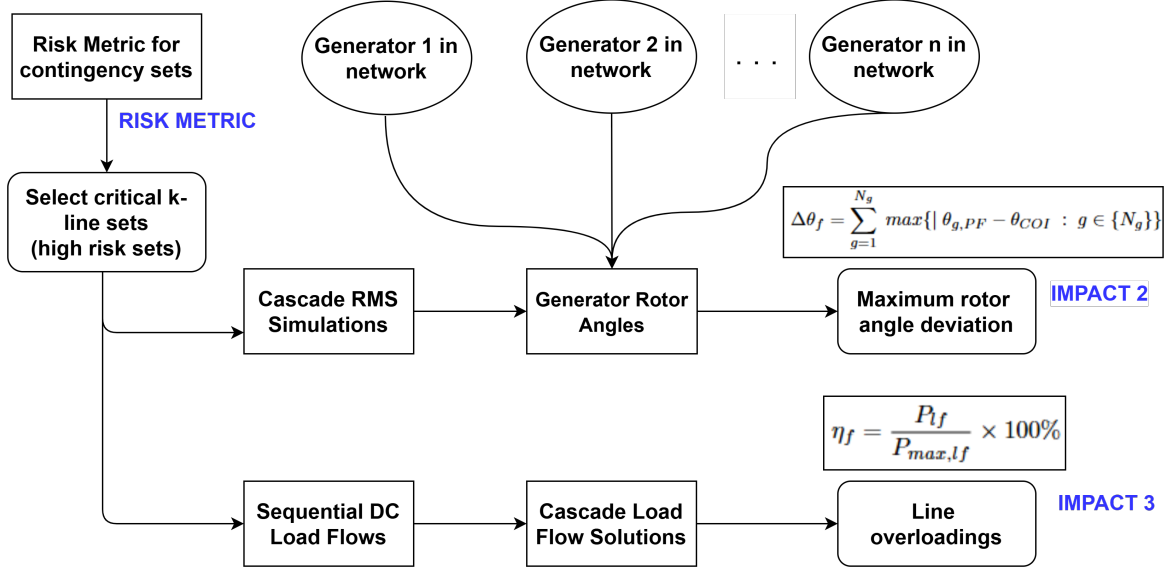


Figure 3.3: Verification after selecting critical contingency sets from risk rankings

Maximum Rotor angle deviation as a verification impact ($\Delta\theta_f$)

While $I_{SC,gf}$ used for the formulation of \mathcal{I}_f provides a steady-state, fault-current-based view of system stress, there is a need for another metric that reflects the dynamic transient stability for verification. This would allow the risk \mathcal{R}_f to be compared and the effectiveness of the model to be analyzed. A comparative analysis with an independent metric, such as maximum rotor angle deviation $\Delta\theta_f$, can be performed. To compute this, a dynamic time-domain RMS simulation is conducted for each line contingency, which models the system's dynamic behaviour under $N - k$ contingencies. The simulation is configured with fault timing, including the fault occurrence time, location and outage time of the line on which the fault occurs, and then $k - 1$ lines are sequentially switched to simulate cascading failures. During the simulation, rotor angle trajectories of all generators are recorded. The deviation of each generator's rotor angle from the system's centre of inertia (COI) is calculated, and the maximum deviation is considered.

$$\Delta\theta_{max,g} = \max\{|\theta_{g,PF} - \theta_{COI}| : g \in \{N_g\}\} \quad (3.37)$$

where $\theta_{g,PF}$ is the rotor angle of generator g at a critical Post-Fault time, and θ_{COI} is the Center of Inertia angle of the system. θ_{COI} reflects the weighted average rotor angle of all generators in the

system and is calculated as:

$$\theta_{COI} = \frac{\sum_{g=1}^{N_g} M_g \cdot \theta_g}{\sum_{g=1}^{N_g} M_g} \quad (3.38)$$

where M_g is the inertia constant of the generator g , and N_g is the total number of generators in the network.

To assess the transient stability impact with a fault at point f , a contingency-specific rotor angle impact metric $\Delta\theta_f$ is calculated by summing up the absolute values of maximum rotor angle deviations of all generators.

$$\Delta\theta_f = \sum_{g=1}^{N_g} (\theta_{max,g}) \quad (3.39)$$

$$\Delta\theta_f = \sum_{g=1}^{N_g} \max\{|\theta_{g,PF} - \theta_{COI}| : g \in \{N_g\}\} \quad (3.40)$$

Line Overloadings as a verification impact (η_f)

In cascading failure analysis, line overloads can play a crucial role in determining the propagation of outages in a power system. The overload impact on subsequent transmission lines is evaluated through sequential DC power flow simulations, where the outage of one line leads to redistribution of power flows, potentially overloading other lines. If any line exceeds its continuous rating threshold $P_{max,lf}$, it is considered overloaded, and the percentage overloading is computed as:

$$\eta_f = \frac{P_{lf}}{P_{max,lf}} \times 100\% \quad (3.41)$$

where P_{lf} is the power on line l post-contingency and $P_{max,lf}$ is the continuous thermal rating of line. The process is repeated iteratively for each newly failed line, and after each failure, power redistribution is recalculated.

\mathcal{I}_f in the model used to calculate \mathcal{R}_f and the metrics such as $\Delta\theta_f$ and η_f to verify the model effectiveness provide complementary perspectives on system behaviour. \mathcal{I}_f focuses on steady-state fault severity by identifying lines experiencing high fault currents. It is critical for evaluating thermal and mechanical stresses and for designing protection schemes. In contrast, $\Delta\theta_f$ captures the dynamic, transient behaviour of the system, providing insights into the likelihood of transient instability or synchronization issues during contingencies. In k -line cascading failure scenarios modelled using $N - k$ contingency simulations, the two metrics may align or diverge. Compared to RMS simulations, η_f offers significantly faster computation, making it feasible for large-scale system studies while still capturing key overload propagation effects. This method provides a perspective on the sequential evolution of failures, allowing for better predictive insights into cascading failures.

3.4.2. Analyzing operator decision making

The \mathcal{R} -based rankings of all contingency scenarios are obtained for different dispatch scenarios d . These dispatches include generator power outputs and load active/reactive power values. A key challenge in this analysis is the high dimensionality of generator and load variables that need to be monitored. To address this, Principal Component Analysis (PCA) is applied to reduce the dimensionality of the dispatch-related variables while preserving the most significant variations.

Given a dataset $\mathbf{X} \in \mathbb{R}^{D \times N}$, where D represents the number of dispatch scenarios and N is the number of monitored variables (e.g., generator power outputs and load values), PCA is performed as follows:

1. **Standardization:** To ensure comparability across different variables, each column of \mathbf{X} is standardized:

$$\tilde{X}_{id} = \frac{X_{id} - \mu_d}{\sigma_j} \quad (3.42)$$

where μ_d and σ_d are the mean and standard deviation of the d -th dispatch.

2. **Covariance Matrix Computation:** The covariance matrix Σ of the standardized dataset is calculated as:

$$\Sigma = \frac{1}{D} \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} \quad (3.43)$$

3. **Eigenvalue Decomposition:** The principal components are obtained by solving the eigenvalue problem:

$$\Sigma \mathbf{v}_k = \lambda_k \mathbf{v}_k \quad (3.44)$$

where λ_k are the eigenvalues and \mathbf{v}_k are the corresponding eigenvectors.

4. **Projection onto Principal Components:** The original dataset is transformed into the new lower-dimensional space using the top M eigenvectors (where $M \ll N$):

$$\mathbf{Z} = \tilde{\mathbf{X}} \mathbf{V}_M \quad (3.45)$$

where \mathbf{V}_M is the matrix containing the top M eigenvectors corresponding to the highest eigenvalues.

In this study, PCA reduces the dispatch-related variables into two primary components:

- **Principal Component 1 (PC1):** Represents the overall dispatch behavior, capturing variations in generator and load values.
- **Principal Component 2 (PC2):** Represents secondary variations in dispatch conditions.

The transformed dataset ($\mathbf{Z} \in \mathbb{R}^{D \times 2}$) is then used for contour plotting, where the total risk \mathcal{R}_{sys} is interpolated over the new reduced space (PC1, PC2).

4

Simulation and Results

DigSilent 2024 offers PowerFactory, a powerful tool for simulating and analyzing power systems. Python provides a flexible and efficient way to automate and extend these capabilities. By leveraging the `pfsim`, a Python library specifically designed for interfacing with PowerFactory, users can automate simulations and analyses efficiently. This integration allows users to perform tasks such as RMS simulation, contingency analysis, short-circuit studies, and sensitivity analysis with enhanced flexibility and repeatability. Python's ability to handle large datasets and visualize results complements PowerFactory's simulation strengths, making it an ideal choice for creating scalable methodologies.

4.1. Simulating cascades and obtaining dynamic impact

The test network used in this study is the IEEE 39-bus system, also known as the 10-machine New England Power System. This network consists of 39 buses, 10 generators, 46 transmission lines, and 19 loads, representing a realistic model of an interconnected power system. The system is widely used for dynamic stability and contingency analysis due to its well-defined parameters and established results. In this analysis, contingency scenarios are evaluated based on the number of cascading events from $N - 1$ to $N - k$ lines. The parameter k denotes the number of contingencies in a cascading failure sequence. Empirical observations in the IEEE39 bus system indicate that scenarios with $k > 4$ have negligible probability due to the low likelihood of such extensive failures occurring in practical power systems. Once $p(s_{k,i})$ of various cascade contingency sets S are obtained as outlined in the methodology, the next step is to determine the proposed impact \mathcal{I} of each set. This is carried out using PowerFactory, where each transmission line in the contingency set is subjected to a short circuit and

switching action. The corresponding LODFs (w_{df}) and short-circuit currents ($I_{SC,gf}$) are then computed to quantify the severity of each scenario. Simultaneously, D_{gf} is computed using (Z_{bus}) matrix.

The selection of fault locations can be altered based on scenarios and the impact value will reciprocate accordingly. The fault duration is determined according to standard protection system clearing times. Following the short-circuit analysis, line impedance values are extracted, and the modified Z_{bus} matrix is computed using Algorithm 1 in 3.3.4. This allows for the calculation of electrical distance d_f using equation 3.35, providing insights into how the topology of the system changes under different contingency conditions. After obtaining the proposed impact metric, the risk associated with each contingency set is calculated. These risks are then ranked, with the highest-risk scenarios selected for further validation. The verification process includes dynamic RMS simulations to assess maximum rotor angle deviations and overloading impact through sequential load-flow analysis. These steps ensure the robustness of the risk ranking and provide a comprehensive understanding of system vulnerabilities.

4.1.1. Project activation and power system dispatch

The PowerFactory project, study case, and related components are activated using the `PowerFactorySim` object. With `M` as the instance of the PowerFactory application object, the following script ensures that the IEEE 39-bus study case is activated and all relevant system components are retrieved:

```

1 # Activate the study case
2 M.study_case.Activate()
3
4 # Retrieve system components for IEEE 39-bus
5 generators = M.app.GetCalcRelevantObjects("*.ElmSym") # 10 generators
6 lines = M.app.GetCalcRelevantObjects("*.ElmLne") # 46 transmission lines
7 loads = M.app.GetCalcRelevantObjects("*.ElmLod") # 19 loads
8 buses = M.app.GetCalcRelevantObjects("*.ElmTerm") # 39 buses
9 transformers = M.app.GetCalcRelevantObjects("*.ElmTr2") # Transformers
10
11 # Retrieve short-circuit events and switching operations
12 evt_folder = M.app.GetFromStudyCase('IntEvt')
13 short_circuits = evt_folder.GetContents("*.EvtShc")
14 switches = evt_folder.GetContents("*.EvtSwitch")
15
16 MONITORED_VARIABLES = {'*.ElmSym': ['c:firot'], '*.ElmTerm': ['m:Pgen']}
```

PowerFactory dynamically organizes IEEE 39-bus system components into predefined classes:

- **Generators:** `M.app.GetCalcRelevantObjects('*.ElmSym')` retrieves all 10 generators. Their attributes, such as minimum power (`Pmin_uc`), maximum power (`Pmax_uc`), and initial power (`pgini`), are adjusted to simulate various dispatch scenarios.
- **Transmission Lines:** `M.app.GetCalcRelevantObjects('*.ElmLne')` retrieves 46 transmis-

sion lines that define power transfer paths and are key for cascading failure analysis.

- **Loads:** `M.app.GetCalcRelevantObjects('*.ElmLod')` retrieves 19 loads, with real (`plini`) and reactive (`qlini`) power dynamically adjusted to ensure consistency with generator dispatch.
- **Buses:** `M.app.GetCalcRelevantObjects('*.ElmTerm')` retrieves all 39 buses, where generators, loads, and transmission lines interconnect.
- **Transformers:** `M.app.GetCalcRelevantObjects('*.ElmTr2')` fetches all transformers regulating voltage and power transfer.

The IEEE 39-bus power flow analysis is executed using:

```
1 # Run Newton-Raphson power flow analysis
2 load_flow = M.app.GetFromStudyCase('ComLdf')
3 load_flow.iopt_net = 0 # AC load flow
4 load_flow.iopt_solver = 2 # Newton-Raphson method
5 load_flow.iopt_therm = 1 # Consider thermal limits
6 load_flow.Execute()
```

This simulation utilizes the Newton-Raphson method with a convergence tolerance of 10^{-6} per unit, ensuring accurate voltage magnitudes, angles, power flows, and branch loading across all buses connected to transmission lines. Thermal limits are enforced to monitor line capacity constraints.

4.1.2. Sensitivity Analysis

Sensitivity analysis provides insight into the redistribution of power flow across all transmission lines during cascading failures by computing Line Outage Distribution Factors (LODF). For each generator in the IEEE 39-bus system, the `ComVstab` object, obtained from `M.app.GetFromStudyCase('ComVstab')`, executes voltage stability analysis. This process assesses how bus voltages behave under different loading conditions and contingency scenarios.

```
1 # Activate voltage stability analysis module
2 vstab = M.app.GetFromStudyCase('ComVstab')
3
4 # Enable LODF calculations
5 vstab.calcPtdf = 1
6
7 # Set contingency sensitivity calculation
8 vstab.isContSens = 1
9
10 # Execute analysis for k-line cascade scenario
11 vstab.Execute()
12
13 # Retrieve results
14 sensitivity_results = vstab.GetResults()
```


The `calcPtdf` attribute is set to 1, enabling the calculation of LODF and contingency sensitivities for each transmission line (w_{gf}). The internal contingency analysis is performed by setting the `isContSens` attribute to 1, ensuring that sensitivity values are determined for the $k - 1$ lines in the k -line cascade scenario.

4.1.3. Short Circuit analysis in PowerFactory and proposed impact

For each generator in the IEEE 39-bus system, the `ComShc` object is retrieved in Python to perform short-circuit calculations in PowerFactory using `M.app.GetFromStudyCase("ComShc")`. This object executes short-circuit analysis by applying faults to selected lines and computing the resulting short-circuit currents. Predefined short-circuit fault events and switch operations used for component isolation during simulations are retrieved using:

- **Short-Circuit Events:** `evt_folder.GetContents('*.EvtShc')` fetches predefined fault scenarios in the system.
- **Switch Operations:** `evt_folder.GetContents('*.EvtSwitch')` retrieves switches used for isolating or reconnecting network components.

```

1 # Retrieve short-circuit calculation object
2 shcobj = M.app.GetFromStudyCase("ComShc")
3
4 # Select the faulted line for short-circuit analysis
5 shcobj.i_shcobj = M.app.GetCalcRelevantObjects("FaultedLine")[0]
6
7 # Set fault type and location
8 shcobj.iopt_shc = 3 # Three-phase short-circuit analysis
9 shcobj.ppro = 50 # Fault applied at 50% of the line length
10
11 # Execute short-circuit analysis
12 shcobj.Execute()
13
14 # Extract short-circuit current magnitude and angle for generators
15 generators = M.app.GetCalcRelevantObjects("*.ElmSym")
16 for gen in generators:
17     isc = gen.GetAttribute("m:I:bus1")
18     angle = gen.GetAttribute("m:phii:bus1")
19
20 # Retrieve predefined short-circuit events and switch operations
21 shc_events = evt_folder.GetContents("*.EvtShc")
22 switch_events = evt_folder.GetContents("*.EvtSwitch")

```

To set up the short-circuit analysis, several attributes are configured to define the fault conditions and extract relevant results. The faulted line is selected based on the k^{th} line in the k -line cascade scenario, which is assigned to the short-circuit analysis object (`shcobj`). The fault type is specified by setting

the attribute `3psc`, which enables three-phase short-circuit calculations. Additionally, the fault location along the transmission line is determined using the `ppro` parameter, which defines the percentage distance of the fault from the line's starting point. Once the short-circuit analysis is executed, the symmetrical short-circuit current magnitude ($I_{SC,gf}$) is extracted from the generator's attribute `m:I:bus1`, while the fault current angle ($I_{\phi,gf}$) is obtained from the attribute `m:phii:bus1`. These extracted values of $I_{SC,gf}$, w_{gf} combined with d_f is used to calculate impact \mathcal{I} given in Equation 3.22.

4.1.4. RMS Simulations in PowerFactory and maximum rotor angle impact

The primary objective of the RMS simulation is to evaluate the changes in rotor angles, which are critical for assessing the stability of the system during and after the cascade. To begin the simulation, the event folder is retrieved from the study case using the `M.app.GetFromStudyCase('IntEvt')` command. The system components, including the transmission lines involved in the cascade and the generators whose impact needs to be evaluated, are initialized using their respective system classes. Specifically, based on the line under cascade analysis, the short-circuit event and switching event are targeted by setting the `p_target` attribute of the events to the specific transmission line. Once initialized, the RMS simulation is executed using `M.run_dynamic_sim()`.

```

1 # Retrieve event folder and events
2 evt_folder = M.app.GetFromStudyCase('IntEvt')
3 shc_events = evt_folder.GetContents('*.EvtShc')
4 switch_events = evt_folder.GetContents('*.EvtSwitch')
5
6 # Initialize system components
7 generators = M.app.GetCalcRelevantObjects('*.ElmSym') # Generators
8
9 # Initialize load-flow and simulation conditions
10 M.initial_cond()
11
12 # Run RMS simulation
13 M.run_dynamic_sim()
14
15 # Get dynamic simulation results (e.g., rotor angle)
16 dynamic_results = M.get_dynamic_results()
17
18 # Extract rotor angle values
19 rotor_angles = [gen.GetAttribute('c:firot') for gen in generators]
20
21 # Calculate maximum rotor angle deviation
22 max_rotor_angle_deviation = numpy.max(numpy.abs(rotor_angles))
23
24 # Calculate overall impact
25 overall_rotor_angle_impact = sum(max_rotor_angle_deviation for _ in generators)

```

After the simulation completes, results are collected using `M.get_dynamic_results()`. These results include the rotor angle θ_f for each generator in the system, indicated by the `c:firot` attribute from PowerFactory. The absolute difference between the maximum and minimum rotor angle values for each generator is calculated to obtain the maximum rotor angle deviation using the expression `numpy.max(numpy.abs(rotor_angles))` from the numpy library in Python. The maximum rotor angle deviations are then summed up over all generators in the network to calculate the overall rotor angle impact of the cascade scenario using the formula from Equation 3.40.

4.1.5. Cascade load flow and line overloading impact

The impact of line overloads is computed in a power system under contingency scenarios, particularly by using PowerFactory's load flow tools via Python.

The first step in the cascade sequential load-flow analysis involves applying specific dispatch values iteratively to the system. For each generator G , P_G and Q_G of specific dispatch d are assigned using the `pgini` attribute, which represents the generator's active power initial setting. Similarly, for each load L , P_L and Q_L values are assigned using the `plini` and `qlini` attributes, respectively. Once all the dispatch values have been applied to the generators and loads, the load flow is executed using PowerFactory's `ComLdf` command, accessed via `M.app.GetFromStudyCase("ComLdf")`. The load flow calculation ensures that the system is operating in a steady-state condition, and the results include solutions on P_{lf} , C_{lf} , and $P_{max,lf}$. Once the load flow is complete, the η_f for each line l are retrieved using `l.GetAttribute('m:loading')`. The process is performed iteratively by disconnecting all lines one by one in set S by setting the `outserv` attribute of each line in the contingency set to 1 and executing the load-flow again.

```

1 # Retrieve the load flow object
2 load_flow_obj = M.app.GetFromStudyCase("ComLdf")
3
4 # Initialize system components (generators and loads)
5 generators = M.app.GetCalcRelevantObjects('*.ElmSym') # Generators
6 loads = M.app.GetCalcRelevantObjects('*.ElmLod') # Loads
7
8 # Assign dispatch values to generators and loads
9 for gen in generators:
10     gen.pgini = specific_dispatch_value_for_generator # Set generator active power
11     gen.qgini = specific_dispatch_value_for_generator # Set generator reactive power
12
13 for load in loads:
14     load.plini = specific_dispatch_value_for_load # Set load active power
15     load.qlini = specific_dispatch_value_for_load # Set load reactive power
16
17 # Execute the load flow calculation
18 load_flow_obj.Execute()

```

```

19
20 # Retrieve the loading values for each line in the system
21 lines = M.app.GetCalcRelevantObjects('*.ElmLne') # Transmission lines
22 for line in lines:
23     line_loading = line.GetAttribute('m:loading') # Get the loading attribute
24
25 # Perform contingency analysis: disconnect each line one by one
26 for line in lines_in_contingency_set:
27     line.outserv = 1 # Disconnect the line
28     load_flow_obj.Execute() # Execute load flow again after disconnection

```

Once all the lines in a single contingency set from S are included, the η_f is considered for the system's overloading impact after that specific cascade contingency set.

4.2. Case Study Setup

4.2.1. Contingency analysis for critical line sets

To get the sets of critical lines, we perform the contingency analysis described in 2.1.3 after choosing the IEEE-39 bus test network as a first step to identify top contingency sets S which are part of contingency line set \mathcal{L} following the condition of $\delta(S) > \theta$ and of size $|S| = 1123$ scenarios for N-2 cascade scenarios.

Although all the possible sets in a network can be considered, doing so can become computationally expensive, particularly when considering higher-order cascades (N-3 or beyond) due to the number of samples rising exponentially, as shown in Figure 4.1 for the

IEEE 39-Bus system. As an example, the number of samples for different contingency orders is given in Table 4.1. Before calculating the probabilities of contingency scenarios or motifs, it is best to choose a few critical ones. The number of possible scenarios increases exponentially with network size, making an exhaustive evaluation computationally impractical.

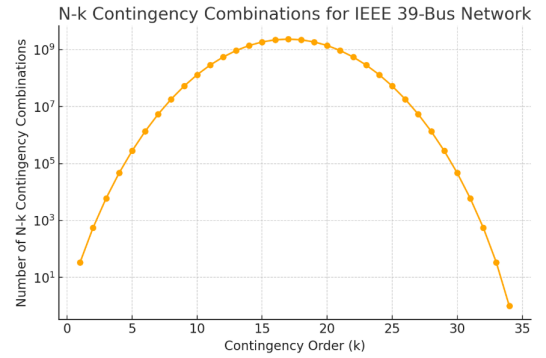


Figure 4.1: Contingency scenarios with increasing k in N- k cascades for IEEE39 bus network

Contingency Order	Number of Samples
N-2	46
N-2	1123
N-3	16902
N-4	178365
N-5	1533939

Table 4.1: CONTINGENCY SAMPLES FOR CASCADE ORDERS IN IEEE39 BUS SYSTEM

Thus, a contingency analysis, helped with the use of a threshold value Θ for overloading ($\delta(S)$), ensures that the analysis is computationally feasible and only top contingencies (over the threshold) are selected for analysis. Here, $\Theta = 80\%$ overloading threshold was chosen ($\delta(S) > 80\%$). The other method of determining these contingencies involves solving power flow equations for each overloaded line, but this is significantly slower for more extensive cascades.

Necessity of using contingency analysis

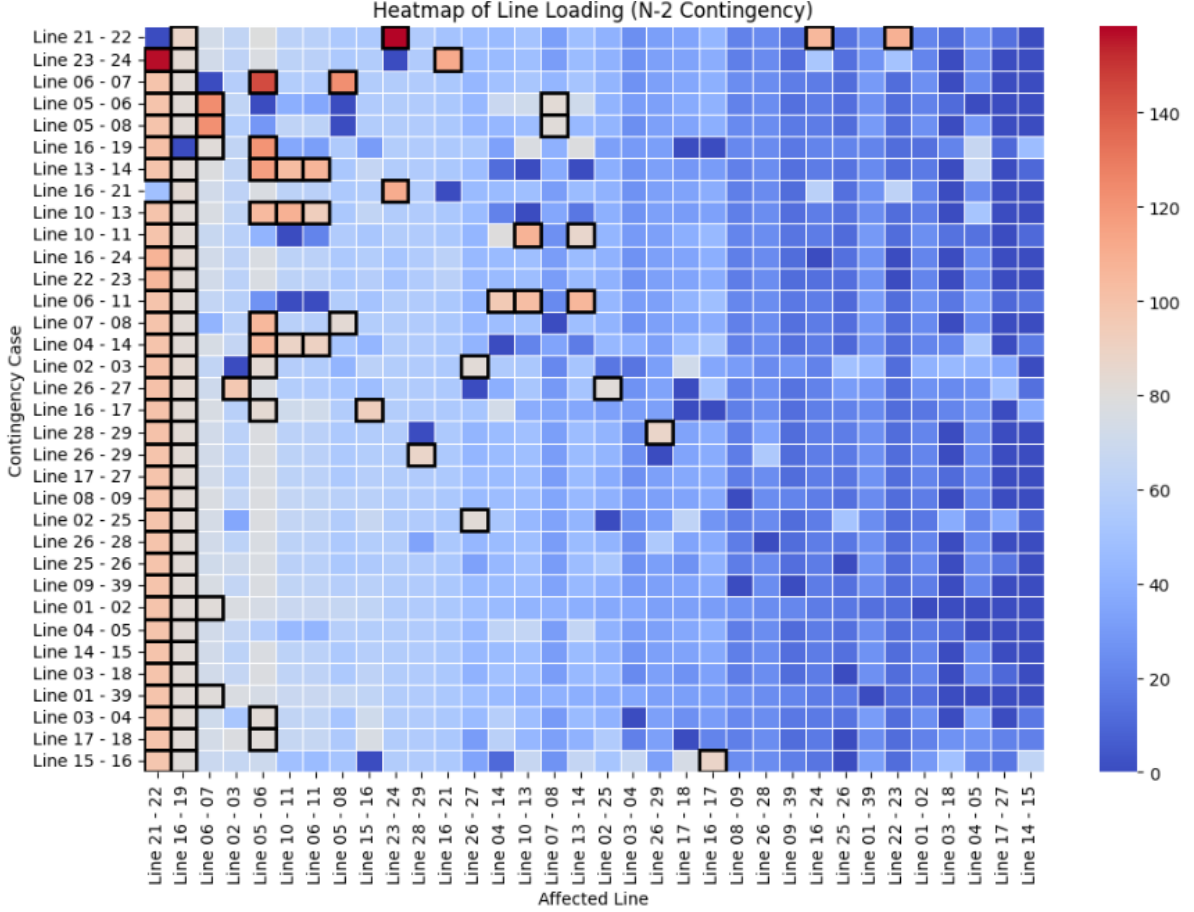


Figure 4.2: Heatmap of line loading under $N - 2$ contingency for IEEE39 network

The heatmap in Figure 4.2 represents the heatmap of $N - 2$ contingency sets under a given operating condition. The x-axis corresponds to the next affected transmission line, while the y-axis represents the initiating line failure. The color intensity indicates the degree of stress experienced by each line, with darker regions signifying higher stress levels corresponding to loadings. The values in the heatmap range from minimum to maximum overloading levels, computed for the selected operating condition. This condition includes active and reactive power loads, such as Load 03 Active Power (706.03 MW), Load 04 Active Power (312.78 MW), Load 07 Active Power (532.26 MW), Load 08 Active Power (324.45 MW), and Load 12 Reactive Power (226.47 MVAR), among others. The heatmap has more than $\theta > 80\%$ of its area marked by a black border, emphasizing highly critical lines. Additionally, it is sorted

based on the loading of affected lines, resulting in a visible gradient. Notably, a few lines, such as Lines 21-22 and Lines 16-19, appear to be the most stressed in terms of overloading.

4.2.2. Probabilities of critical line sets

Once the critical line sets are identified, a probabilistic analysis is performed to determine the motif probabilities. For this, the uniform ($P^{uni}(S_{k,i} | k)$) using Equation 2.24 and empirical ($P(S_{k,i} | k)$) probabilities using Equation 2.25 are calculated. A hypothesis test is performed, given in Equations 2.27 and 2.28, to identify a motif. For motif $S_{k,i}$ at depth k , a motif is identified based on this comparison. This is illustrated in Table 4.2.












	$S_{k,i}$	$ S_{k,i} $	$P^{uni}(S_{k,i} k)$	n_k	$n_{k,i}$	$P(S_{k,i} k)$	motif	p-value
	$S_{2,1}$	57	0.108	17	7	0.412	true	0.
	$S_{2,2}$	504	0.955	17	10	0.588	false	1.
	$S_{3,1}$	22	0.004	11	2	0.182	true	0.
	$S_{3,2}$	1589	0.291	11	7	0.636	true	0.
	$S_{3,3}$	83	0.015	11	2	0.182	true	0.
	$S_{3,4}$	1	0.0002	11	0	0	false	1.
	$S_{3,5}$	4289	0.786	11	0	0.677	false	1.
	$S_{4,1}$	6	0.0001	8	0	0	false	1.
	$S_{4,2}$	579	0.014	8	1	0.125	true	0.
	$S_{4,3}$	77	0.0018	8	1	0.125	true	0.
	$S_{4,4}$	119	0.003	8	0	0	false	1.
others	$S_{4,*}$	40350	0.981	8	6	0.75	false	1.

Table 4.2: PROBABILITIES OF PATTERNS IN IEEE39 SYSTEM FOR IDENTIFIED CONTINGENCIES

Once the contingency motifs are identified, the probability of a specific contingency subgraph is calculated. Firstly, the probability of k line outages $P(k)$ for $k - line$ cascade is calculated using Equation 2.31. For $S_{2,2}$ and $S_{3,5}$, there is a significant discrepancy between uniform and empirical probabilities. The deviation suggests that the uniform model overestimates these motifs. While hypothesis testing fails, the probability of cascades containing $S_{3,5}$ remains high, justifying its inclusion. However, excluding $S_{2,2}$ may weaken the analysis, as its empirical probability, though lower than uniform, is still notable. Secondly, the Table 4.3 shows the empirical probabilities of patterns or $P(S_{k,i} | k)$, which is the probability of a pattern given k -line outages. It was calculated earlier to identify the motif.

Thirdly, to calculate $P(d | S_{k,i})$, the diameters d of each disconnected pattern are identified, and $P(d | disconnected)$ is calculated using Equation 2.35. This is shown in Table 4.4. For connected patterns $P(s | connected)$, this value is 1, as suggested earlier in Equation 2.34.

$S_{k,i}$	$S_{2,1}$	$S_{2,2}$			
$P(S_{k,i} k)$	0.4118	0.5882			
$S_{k,i}$	$S_{3,1}$	$S_{3,2}$	$S_{3,3}$	$S_{3,4}$	$S_{3,5}$
$P(S_{k,i} k)$	0.1818	0.6364	0.1818	0	0.6770
$S_{k,i}$	$S_{4,1}$	$S_{4,2}$	$S_{4,3}$	$S_{4,4}$	$S_{4,5}$
$P(S_{k,i} k)$	0	0.1250	0.1250	0	0.75

Table 4.3: DISTRIBUTION OF PATTERNS (EMPIRICAL PROBABILITIES) $P(S_{k,i} | k)$ FOR IEEE39 NETWORK

Diameter d	$P(d S_{k,i})$
2	0.0038
3	0.0285
4	0.1004
5	0.1999
6	0.2972
7	0.2324
8	0.1151
9	0.0225

Table 4.4: PROBABILITY $P(d | S_{k,i})$ FOR DISCONNECTED SUBGRAPHS FOR DIFFERENT DIAMETERS IN $S_{k,i}$ for IEEE39 NETWORK

Finally, the Table 4.5 shows the number of subgraphs of the network with a specific diameter. This is used in Equation 2.36 to calculate the probability of a contingency given its pattern and diameter ($P(s_{k,i} | S_{k,i}, d)$).

d	$ S_{2,2} $	$ S_{3,2} $	$ S_{3,5} $	$ S_{4,2} $
2	79	82	2	20
3	102	329	189	115
4	106	379	648	156
5	93	328	1043	127
6	73	276	1216	103
7	36	134	792	42
8	13	52	340	14
9	2	9	59	2
10	0	0	0	0

Table 4.5: NUMBER OF DISTINCT SUBGRAPHS WITH DIFFERENT DIAMETERS IN $S_{k,i}$ for IEEE-39 BUS

Equation 2.39 is used for the calculation of outage probability of a specific set $P(s_{k,i})$ with a diameter d and following a subgraph pattern $s_{k,i}$. It is represented in the Table 4.6. The table shows that the outage probability of motifs has a higher value compared to the outage probability of the non-motif subgraph pattern.

d	$s_{2,1}$	$s_{2,2}$	$s_{3,1}$	$s_{3,2}$	$s_{3,3}$	$s_{3,4}$	$s_{3,5}$	$s_{4,1}$	$s_{4,2}$	$s_{4,3}$	$s_{4,4}$
1	0.0034	0	0.0025	0	0	0	0	0	0	0	0
2	0	0.000013	0	0.0000091	0.00067	0	0.0004	0	0.0000053	0.00036	0
3	0	0.000078	0	0.000017	0	0	0.000031	0	0.0000069	0	0
4	0	0.00026	0	0.000052	0	0	0.000032	0	0.000018	0	0
5	0	0.0006	0	0.00012	0	0	0.00004	0	0.000044	0	0
6	0	0.0011	0	0.00021	0	0	0.000051	0	0.00008	0	0
7	0	0.0018	0	0.00034	0	0	0.000061	0	0.00015	0	0
8	0	0.0025	0	0.00043	0	0	0.00007	0	0.00023	0	0
9	0	0.0031	0	0.00049	0	0	0.000079	0	0.00031	0	0
10	0	0	0	0	0	0	0	0	0	0	0

Table 4.6: PROBABILITY OF OUTAGES ($P(s_{k,i})$) WITH DIFFERENT PATTERNS AND DIAMETERS (d)

The higher probabilities of motifs, along with the inclusion of $s_{2,2}$ and $s_{3,5}$ are highlighted in the table. The Zipf distribution of diameters, as suggested previously, can be observed as having a heavy tail, implying that large patterns are rare, but they do occur. The Complementary Cumulative Distribution Function (CCDF) [44] plot in Figure 4.3 is a graph that visually demonstrates the Zipfian nature of cascading failures. The Zipfian nature follows the relation $P(r) \propto r^{-\alpha}$, where $P(r)$ is the probability of the r -th ranked event, and α is a positive exponent. The CCDF slowly decays as diameter increases, showing that larger cascades are rare, but they do occur and exhibit a characteristic heavy-tailed distribution.

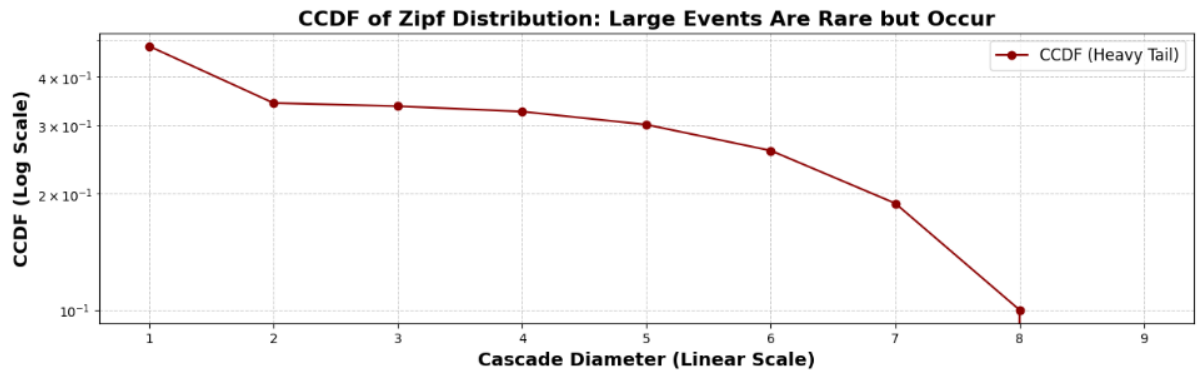


Figure 4.3: Complementary Cumulative Distribution Function (CCDF) Plot showing diameter distribution of cascade sets

An example of calculating the probability for a specific contingency line set is illustrated in Figure 4.4. The process begins with **STEP 1**, where contingency analysis is conducted using continuous loading as a threshold ($\theta > 80\%$) to determine whether the contingency set is critical. If deemed critical, **STEP 2** involves checking whether the contingency subgraph qualifies as a motif. In **STEP 3**, the subgraph's diameter is computed, followed by **STEP 4**, where the pattern type $s_{k,i}$ is identified. Finally, in **STEP 5**, the probability of the corresponding cascade outage $P(s_{k,i})$ is determined for a motif.

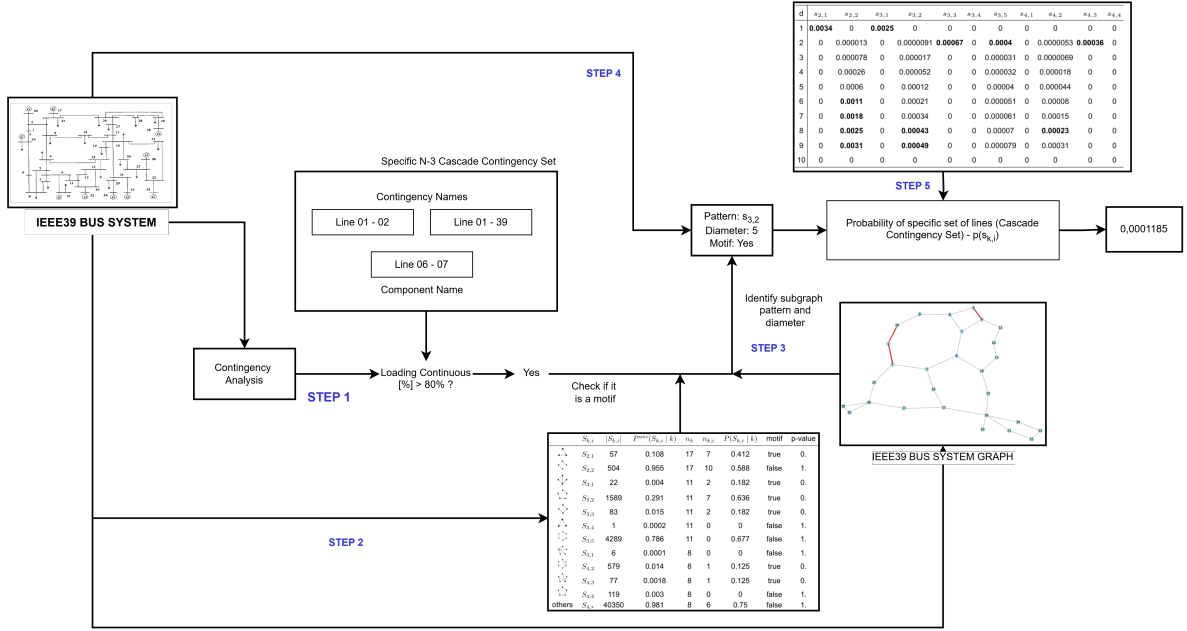


Figure 4.4: An example of probability of a specific line set from the cascade contingency list

4.3. Case Studies, simulations and results

4.3.1. CS1: Impact Metric

Once the probabilities for critical contingency scenarios in \mathcal{S} are computed, the next focus will be on computing the impact metric \mathcal{I} for each scenario using the equation outlined in Equation 3.22. First, the short circuit analysis is performed for each set in \mathcal{S} to get $I_{SC,gf}$ using Equation 3.23, then the sensitivity weights w_{gf} using sensitivity analysis in PowerFactory is calculated using Equation 3.27 and the electrical distance D_{df} is calculated using Equation 3.35.

CS1.1: Correlation of \mathcal{I} with verification-impact $\Delta\theta_f$

This case study outlines the measure of the similarity of \mathcal{I} with the verification-metric $\Delta\theta_f$ calculated using Equation 3.40 for each scenario in \mathcal{S} . The scatter plot compares the \mathcal{I} (X-axis) against $\Delta\theta_f$ (Y-axis) to assess the similarity of the proposed metric with dynamic simulation results for $|\mathcal{S}|$ scenarios on specific operating conditions described in 4.2.1. Each point in the scatter plot is the combination of normalized $\Delta\theta_f$ and the mean \mathcal{I} .

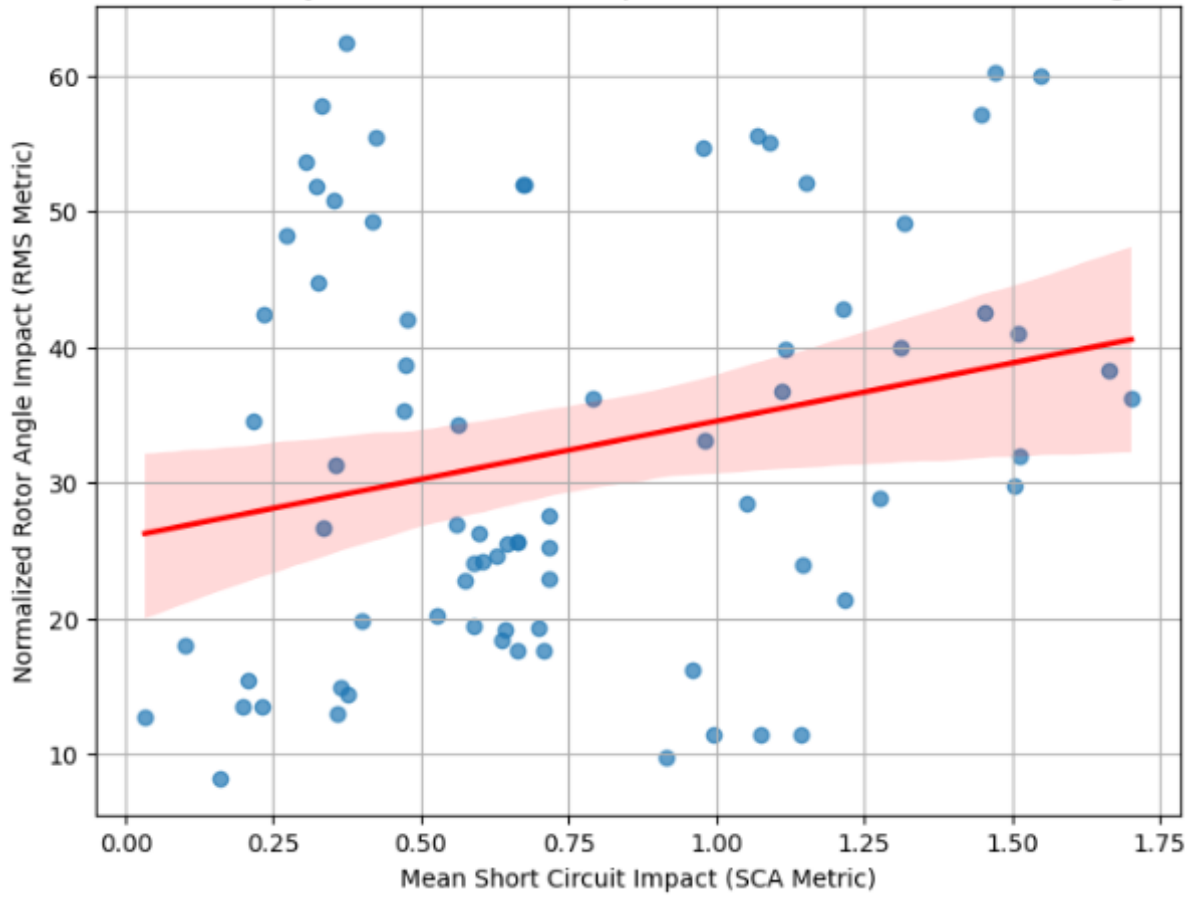


Figure 4.5: Correlation plot between proposed impact \mathcal{I} versus verification impact $\Delta\theta_f$

The scatter plot presented in Figure 4.5 plots the correlation coefficient \mathcal{CF} between the short circuit impact metric \mathcal{I} , and the normalized maximum rotor angle impact, $\Delta\theta_f$, obtained through RMS simulation. The correlation coefficient is computed as:

$$\mathcal{CF} = \frac{\sum(\mathcal{I}_i - \bar{\mathcal{I}})(\Delta\theta_{f,i} - \bar{\Delta\theta_f})}{\sqrt{\sum(\mathcal{I}_i - \bar{\mathcal{I}})^2} \sqrt{\sum(\Delta\theta_{f,i} - \bar{\Delta\theta_f})^2}} \quad (4.1)$$

Quantifies the linear relationship between these two variables. A regression line has been fitted to highlight the general trend. The graph, along with iterations of other dispatch cases, suggests that the observed correlation ranges from negligible to very slight, depending on the case.

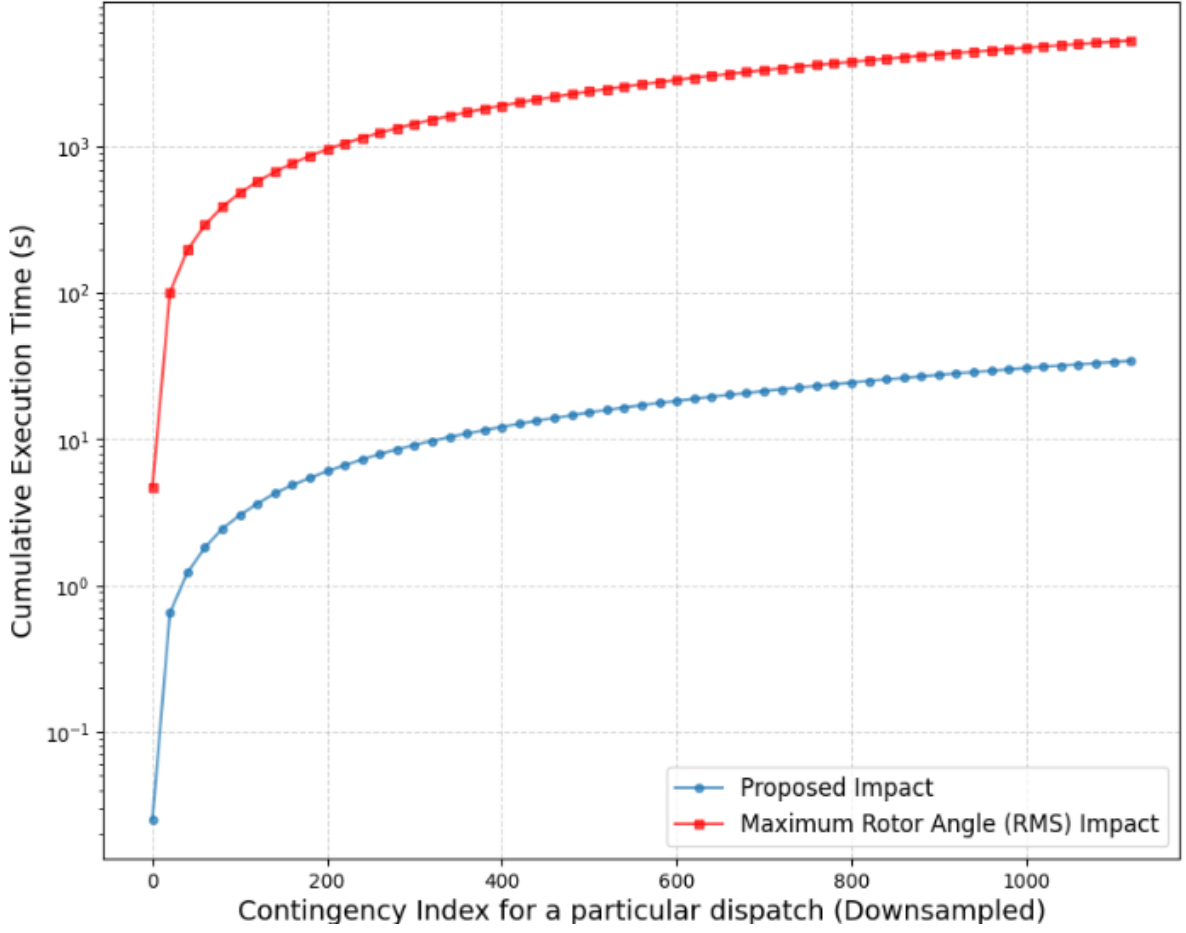
CS1.2: Computational efficiency of \mathcal{I} 

Figure 4.6: Computational time comparison between \mathcal{I} versus $\Delta\theta_f$

The cumulative execution time plot presented in Figure 4.6 illustrates the comparison of computational time required for the proposed SCA-based impact metric \mathcal{I} versus the RMS-based maximum rotor angle deviation verification impact metric $\Delta\theta_f$ across $|\mathcal{S}|$ contingency scenarios for a single dispatch case. The x-axis represents the contingency index (downsampled for clarity), while the y-axis represents the cumulative execution time (in seconds) on a logarithmic scale. The following observations and results can be decided from the plot.

By comparing the average computation times, we find that calculating the proposed metric \mathcal{I} takes approximately 0.03066548 *seconds*, seconds on average, which is significantly faster than RMS simulations for obtaining $\Delta\theta_f$, which have a mean computation time of 4.768355872 *seconds* per contingency.

4.3.2. CS2: Risk calculation and analysis

A total of $|\mathcal{S}|$ contingency scenarios are analyzed, and \mathcal{I} and $P(s_{k,i})$ are calculated for each contingency scenario across D dispatch cases each represented by D_d where $d = 1$ to 450. Therefore, a total of

$D \times |\mathcal{S}|$ short-circuit simulations were executed, with \mathcal{I} calculated for each simulation. Similarly, $P(s_{k,i})$ is obtained for each scenario from Table 4.6, depending on the subgraph pattern and diameter of that scenario. The risk \mathcal{R} is calculated for each scenario using Equation 3.36.

CS2.1: Comparing the effectiveness of risk metric

The verification impact metric $\Delta\theta_f$ outlined in Section 3.4.1 is used as a standard test. A comparison is made between the results obtained from the proposed risk \mathcal{R} -based method, the verification impact metric η_f , and random selection. The focus is on assessing whether the \mathcal{R} -based metric offers a better estimate compared to the overload-based η_f and the random selection method.

Each method identifies the top 10 contingencies \mathcal{S}_{10} per D_d based on different selection criteria:

1. **Risk-Based Selection (\mathcal{R} -metric):** For each dispatch D_d , the set of contingencies \mathcal{S}_d

$$\mathcal{S}_d = \{s_1, s_2, \dots, s_{|\mathcal{S}_d|}\}$$

is defined where $|\mathcal{S}_d|$ is the total number of contingencies in dispatch D_d . The risk for each contingency $s_i \in \mathcal{S}_d$ is calculated using the risk \mathcal{R} -method, yielding the set of risk values \mathcal{RF}_d :

$$\mathcal{RF}_d = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_{|\mathcal{S}_d|}\}$$

Where \mathcal{R}_i is the risk associated with contingency s_i , calculated using Equation 3.36 with the given input parameters. The risks are then arranged based on their values, yielding the ordered set $\mathcal{R}_d^{\text{ordered}}$:

$$\mathcal{RF}_d^{\text{ordered}} = \{\mathcal{R}_1^{\text{ordered}}, \mathcal{R}_2^{\text{ordered}}, \dots, \mathcal{R}_{|\mathcal{S}_d|}^{\text{ordered}}\}$$

The top 10 contingencies, based on the order of the risks \mathcal{R} , are selected as the most hazardous. The set of the top 10 contingencies $\mathcal{S}_{\mathcal{R}10}^{(d)}$ is given by:

$$\mathcal{S}_{\mathcal{R}10}^{(d)} = \{s_{\mathcal{R}1}^{(d)}, s_{\mathcal{R}2}^{(d)}, \dots, s_{\mathcal{R}10}^{(d)}\} \subset \mathcal{S}_d$$

Where each $s_{\mathcal{R}z}^{(d)} \in \mathcal{S}_{\mathcal{R}10}^{(d)}$ represents a selected contingency, and the corresponding risk value is $\mathcal{R}_z^{\text{ordered}}$ for each dispatch D_d . For each selected contingency $s_{\mathcal{R}z}^{(d)} \in \mathcal{S}_{\mathcal{R}10}^{(d)}$, the verification impact $\Delta\theta_{f_z}^{(d)}$ is calculated using Equation 3.40, and the true impact for each selected contingency is represented by the set:

$$\mathcal{T}_{\mathcal{R}10}^{(d)} = \{\Delta\theta_{\mathcal{R}f_1}^{(d)}, \Delta\theta_{\mathcal{R}f_2}^{(d)}, \dots, \Delta\theta_{\mathcal{R}f_{10}}^{(d)}\}$$

where each $\Delta\theta_{\mathcal{R}f_z}^{(d)}$ represents the verification impact for the selected contingency s_z for each dispatch D_d

2. **Line Overloading-Based Selection:** In Section 3.4.1, the line overloading-based selection method identifies the top 10 contingencies for a given dispatch D_d through sequential load-flow analysis.

For each line $l \in \mathcal{L}$, the overloading $\eta_l^{(d)}$ is calculated using Equation 3.41 as

$$\eta_l^{(d)} = \frac{P_l^{(d)}}{P_{\max,l}^{(d)}} \times 100\%.$$

The first contingency $s_1^{(d)}$ is selected based on the highest overloading, and its corresponding impact is given by

$$\eta_{f_1}^{(d)} = \eta_l^{(d)}.$$

After simulating the fault and outage on the overloaded line, a new load-flow analysis is conducted to identify the next most overloaded line. This process is repeated iteratively until the top 10 contingencies $\mathcal{S}_{\eta 10}^{(d)}$ are identified:

$$\mathcal{S}_{\eta 10}^{(d)} = \{s_{\eta 1}^{(d)}, s_{\eta 2}^{(d)}, \dots, s_{\eta 10}^{(d)}\} \subset \mathcal{S}_d.$$

The overload impact for each selected contingency $s_{\eta z}^{(d)} \in \mathcal{S}_{\eta 10}^{(d)}$ is calculated as

$$\eta_{f_z}^{(d)} = \frac{P_{l_z}^{(d)}}{P_{\max,l_z}^{(d)}} \times 100\%.$$

The line overloading-impact set for the top 10 contingencies is represented by

$$\eta\mathcal{F}_{10}^{(d)} = \{\eta_{f_1}^{(d)}, \eta_{f_2}^{(d)}, \dots, \eta_{f_{10}}^{(d)}\},$$

The corresponding verification impacts for these top 10 contingencies are calculated and given as

$$\mathcal{T}_{\eta 10}^{(d)} = \{\Delta\theta_{\eta f_1}^{(d)}, \Delta\theta_{\eta f_2}^{(d)}, \dots, \Delta\theta_{\eta f_{10}}^{(d)}\}$$

where each $\Delta\theta_{\eta f_z}^{(d)}$ represents the verification impact for the selected overloading-contingency s_z for each dispatch D_d

3. **Random Selection:** In this approach, 10 contingencies are randomly selected from \mathcal{S}_d . These contingencies are considered to be the most hazardous without any specific order or prioritization. The process of selection is based on a random distribution, ensuring that every contingency has an equal probability of being chosen. Let the set of all contingencies in dispatch D_d be $\mathcal{S}_d = \{s_1, s_2, \dots, s_{|\mathcal{S}_d|}\}$. The selection is done by choosing 10 contingencies randomly from \mathcal{S}_d using a uniform distribution. Mathematically, the randomly selected set $\mathcal{S}_{\text{random}}^{(d)} \subset \mathcal{S}_d$ is given by:

$$\mathcal{S}_{\text{random}}^{(d)} = \{s_{r_1}^{(d)}, s_{r_2}^{(d)}, \dots, s_{r_{10}}^{(d)}\} \quad \text{where each } r_z \in \{1, 2, \dots, |\mathcal{S}_d|\}.$$

The selection is based on a uniform distribution, ensuring that every contingency has an equal chance of being chosen. Each contingency $s_{r_z}^{(d)}$ is randomly chosen without any preference, and

their corresponding $\Delta\theta_f$ are calculated using RMS simulation. These $\Delta\theta_f$ for each selected contingency $s_{r_z}^{(d)}$ are denoted $\Delta\theta_{r_z}^{(d)}$, and the set for the top 10 contingencies is represented by:

$$\mathcal{T}_{\text{random}}^{(d)} = \{\Delta\theta_{r_1}^{(d)}, \Delta\theta_{r_2}^{(d)}, \dots, \Delta\theta_{r_{10}}^{(d)}\}.$$

Where each $\Delta\theta_{r_z}^{(d)}$ for the selected contingency $s_{r_z}^{(d)}$ is represented.

The goal for this case study is to evaluate and compare these three methods (risk-based (\mathcal{R}), line overloading-based (η_f) and random selection (r)) based on key performance indicators such as Selectivity, Coverage, Precision, Recall, and F1-score using $\Delta\theta_f$ as the reference benchmark. The performance metrics were calculated for each method against the true top 10 high-impact cases $\mathcal{T}_{\text{true}}^{(d)}$, which are determined by ordering all contingencies obtained from the three selection methods based on $\Delta\theta_f$. The combined set is given by:

$$\mathcal{T}_{\text{combined}}^{(d)} = \mathcal{T}_{\mathcal{R}10}^{(d)} \cup \mathcal{T}_{\eta10}^{(d)} \cup \mathcal{T}_{\text{random}}^{(d)} = \{\Delta\theta_1^{(d)}, \Delta\theta_2^{(d)}, \dots, \Delta\theta_{30}^{(d)}\}$$

From this combined set, the top 10 contingencies with the highest $\Delta\theta_f$ values are selected as the true high-impact cases:

$$\mathcal{T}_{\text{true}}^{(d)} = \{\Delta\theta_{t_1}^{(d)}, \Delta\theta_{t_2}^{(d)}, \dots, \Delta\theta_{t_{10}}^{(d)}\}, \quad \text{where } \Delta\theta_{t_z}^{(d)} \text{ are the } z^{\text{th}} \text{ highest-ordered values.}$$

And the corresponding true order is:

$$\mathcal{S}_{\text{true}}^{(d)} = \{s_{t_1}^{(d)}, s_{t_2}^{(d)}, \dots, s_{t_{10}}^{(d)}\}, \quad \text{where } s_{t_z}^{(d)} \text{ are the } z^{\text{th}} \text{ highest-ordered contingency.}$$

Method	Precision (%)	Recall (%)	F1-score (%)	Coverage (%)	Selectivity (%)
Risk Model (\mathcal{R})	71.8960	71.4576	71.6761	71.4576	71.8960
Overloading-Based (η_f)	89.3417	41.3763	56.5588	41.3763	89.3417
Random Selection (r)	23.9336	26.3937	25.1036	26.3937	23.9336

Table 4.7: PERFORMANCE METRICS OF CONTINGENCY SELECTION

The Table 4.7 presents the percentage selectivity, coverage, precision, recall, and F1-score for each method. It is important to note that, although coverage and recall share the same numerical values in this case, they are not inherently the same metric. Recall ($R^{\mathcal{R}}$) and Coverage ($C^{\mathcal{R}}$), in case of \mathcal{R} are given as

$$R^{\mathcal{R}} = \frac{|\mathcal{S}_{\mathcal{R}10}^{(d)} \cap \mathcal{S}_{\text{true}}^{(d)}|}{|\mathcal{S}_{\text{true}}^{(d)}|} \quad C^{\mathcal{R}} = \frac{|\mathcal{S}_{\mathcal{R}10}^{(d)} \cap \mathcal{S}_{\text{true}}^{(d)}|}{|\mathcal{S}_{\mathcal{R}10}^{(d)} \cup \mathcal{S}_{\text{true}}^{(d)}|}$$

Similarly, Precision ($P^{\mathcal{R}}$) and Selectivity ($S^{\mathcal{R}}$), in case of \mathcal{R} are given as

$$P^{\mathcal{R}} = \frac{|S_{\mathcal{R}10}^{(d)} \cap S_{\text{true}}^{(d)}|}{|S_{\mathcal{R}10}^{(d)}|} \quad S^{\mathcal{R}} = \frac{|S_{\mathcal{R}10}^{(d)} \cap S_{\text{true}}^{(d)}|}{|S_{\mathcal{R}10}^{(d)} - S_{\text{true}}^{(d)}|}$$

Though they match in the $N - 2$ case case, as k increases for $N - k$ cascade cases, recall remains stable, but coverage tends to vary. Similarly, selectivity and precision also match in value but differ conceptually. Precision diverges from selectivity due to increased false positives.

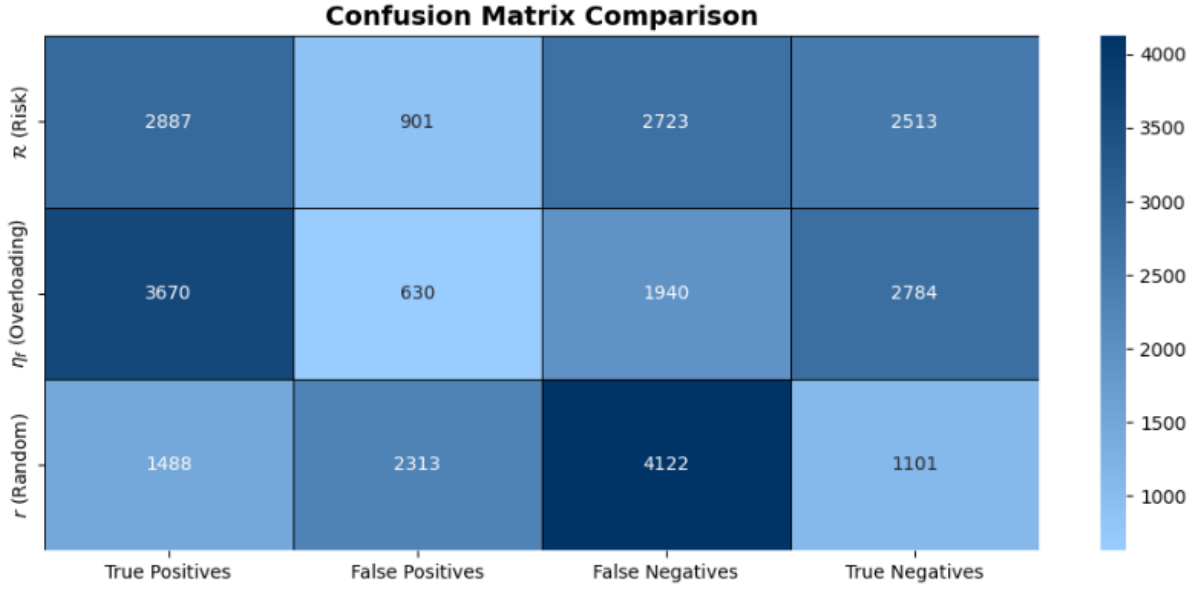


Figure 4.7: Confusion matrix comparison of Contingency Selection Methods

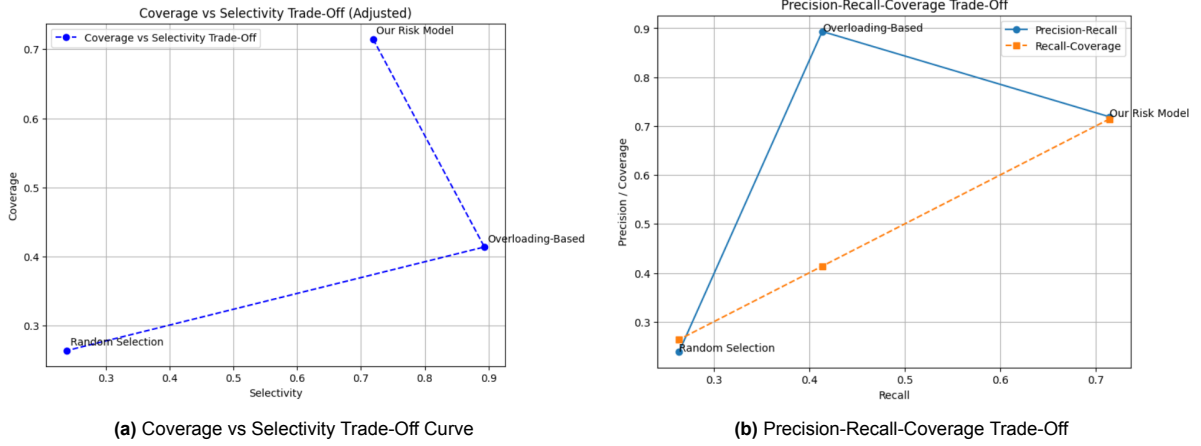
Figure 4.7 shows the confusion matrix comparing the positives and negatives of the ordered contingency sets obtained from the three methods as $S_{\eta10}^{(d)}$, $S_{\mathcal{R}10}^{(d)}$, $S_{\text{random}}^{(d)}$ and the true-order of contingencies in the set $S_{\text{true}}^{(d)}$ for a specific dispatch scenario d . The Overloading-Based Selection method achieves the highest selectivity (89.34%), meaning it is highly confident about its chosen contingencies. However, it has low coverage (41.37%), implying that many high-impact contingencies are missed. Furthermore, it has the highest precision (89.34%) but a very low recall (41.37%), meaning it fails to capture many actual high-impact contingencies. The proposed risk approach balances selectivity (71.89%) and coverage (71.45%), indicating that it effectively captures more true high-impact contingencies while maintaining good selectivity. It achieves a strong recall (71.45%) with good precision (71.89%), proving to be the most reliable method in predicting high-risk cases. As expected, the random method performs worst in both coverage (26.29%) and selectivity (23.933%), demonstrating that it is not a reliable approach.

Table 4.8 summarizes the computational efficiencies of the three methods. The overloading-based approach to get $S_{\eta10}^{(d)}$ is computationally expensive, requiring sequential power flow simulations to determine overload propagation. In contrast, the risk model to get $S_{\mathcal{R}10}^{(d)}$ calculates all risks \mathcal{R} at once,

Method	Set Computation Time (s)
Risk Model ($\mathcal{S}_{\mathcal{R}10}^{(d)}$)	0.031
Overloading-Based ($\mathcal{S}_{\eta10}^{(d)}$)	7.997
Random Selection ($\mathcal{S}_{\text{random}}^{(d)}$)	≈ 0

Table 4.8: COMPUTATION TIME PER SCENARIO ($N = 2$) FOR DIFFERENT METHODS

making it significantly more efficient. Random selection is the fastest, with negligible computation time, as it does not rely on complex calculations.

**Figure 4.8:** Comparative Performance of Selection Methods for High-Impact Contingencies

The Coverage vs Selectivity trade-off curve in Figure 4.8a shows the relationship between two critical metrics for each selection method. The precision-recall coverage Trade-Off graph in Figure 4.8b highlights how well each method balances precision, recall, and coverage.

CS2.2: Comparing the effectiveness of risk rankings

This study compares the rankings obtained using the risk-based metric $\mathcal{S}_{\mathcal{R}10}^{(d)}$ coming from the impact \mathcal{I} and contingency probabilities $p(s_{k,i})$ with the rankings derived from true-impact ($\mathcal{T}_{\mathcal{R}fz}^{(d)}$), particularly from $\Delta\theta_{\mathcal{R}10}^{(d)}$. By employing Spearman correlation coefficient (ρ), we evaluate how well the risk-based ranking aligns with the chosen stability metric.

To validate the effectiveness of \mathcal{R} -based ranking, two comparative approaches were undertaken:

1. **Random 10 Contingencies:** 10 randomly selected contingencies were analyzed using $(\Delta\theta_{rs}^{(d)})$ to get the ranking $\mathcal{S}_{\text{random}}^{(d)}$.
2. **Top 10 Risk-Based Contingencies:** RMS dynamic simulations were performed on the top 10 ranked contingencies, and ranking pertaining to $\mathcal{S}_{\mathcal{R}\infty}^{(d)}$ is recorded in $\mathcal{T}_{\mathcal{R}10}^{(d)}$.

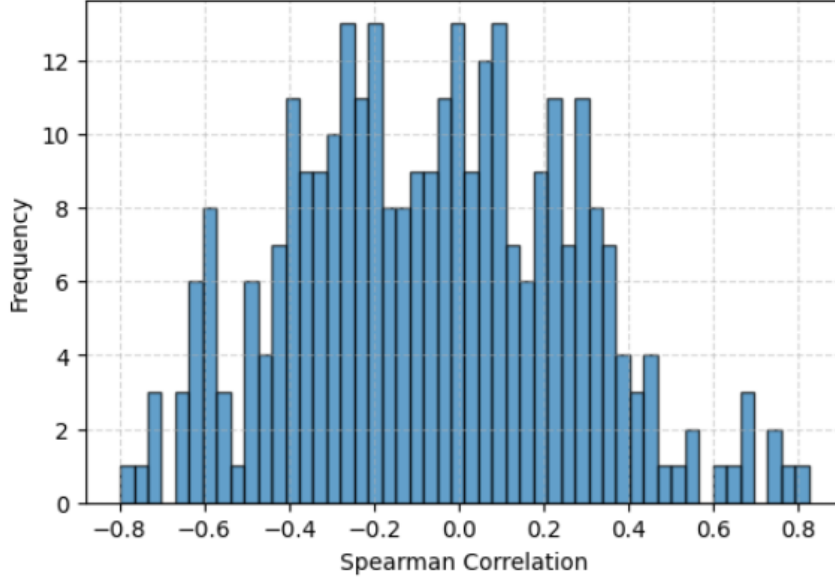
To assess the consistency between risk rankings ($\mathcal{S}_{\mathcal{R}10}^{(d)}$) and true-impact rankings ($\mathcal{S}_{\Delta\theta_{f10}^{(d)}}$) which is ranked contingencies of $\mathcal{T}_{\mathcal{R}10}^{(d)}$, spearman correlation coefficients ($\rho_{10}^{(d)}$) were computed. Higher correlations indicate that risk-based rankings align well with actual system behavior. Spearman's rank

correlation coefficient (ρ) measures the strength and direction of the association between two ranked variables.

$$\rho_{10}^{(d)} = 1 - \frac{6 \sum_{i=1}^{10} rd_i^2}{n(n^2 - 1)} \quad (4.2)$$

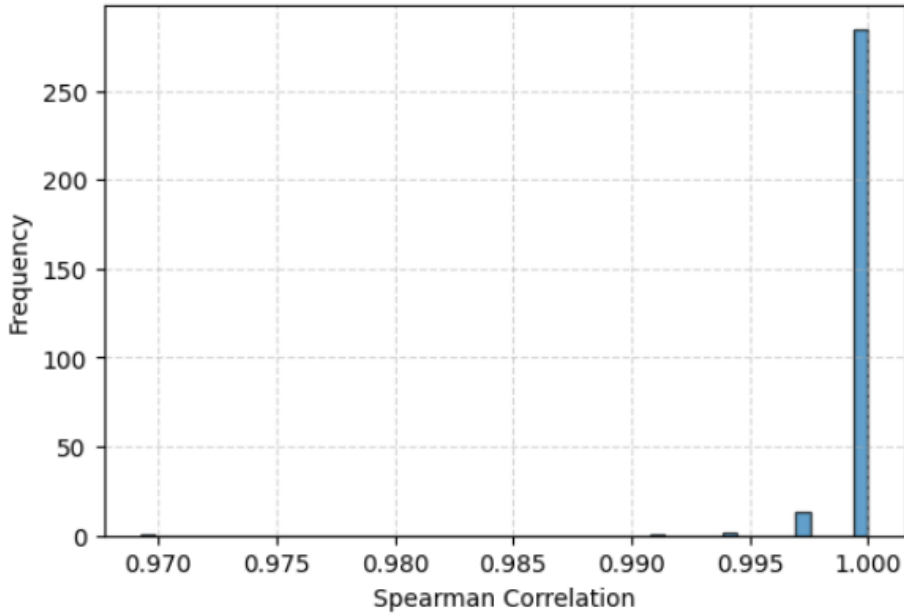
where rd_i is the difference of rankings between each pair for a total of n observations. Non-linear relationships are also captured, making it effective for ranking analysis.

Distribution of Spearman Correlations (Rotor Angle Impact) (Random 10 lines)



(a) Correlation with Randomly selected 10 contingencies

Distribution of Spearman Correlations (Rotor Angle Impact) (Top 10 lines)



(b) Correlation with top 10 risk-based contingencies

Figure 4.9: Distribution of Spearman Rank Correlation coefficient between \mathcal{R} -based metric and $\Delta\theta$ impact based metric for random and top-10 ranked contingencies

1. **Random 10 N-2 cascade contingencies:** A histogram of Spearman correlations for randomly

selected contingencies in Figure 4.9a demonstrates a wide distribution centered around zero. The mean correlation coefficient for these cases is -0.08, with a standard deviation of 0.32. A one-sample t-test against zero yielded a p-value of 0.187, indicating that there is no significant correlation between risk and impact rankings in randomly chosen cases.

2. **Top-10 \mathcal{R} -Based N-2 Cascade Contingencies:** A stark contrast is observed when analyzing the correlation coefficient for the top 10 \mathcal{R} -based contingencies in Figure 4.9b. The distribution is highly skewed toward values near 1.0, with a mean coefficient of 0.998 and a standard deviation of 0.004. A t-test confirms statistical significance ($p < 0.001$), strongly supporting that risk-based rankings provide a much stronger predictive power for $\Delta\theta_f$.

4.3.3. CS3: Effects on dispatch on system risk

This case study explores how different dispatch scenarios (d) impact system risk \mathcal{R}_{sys}^d . The analysis is based on dispatch data, where each scenario has a varied generator powers (P_G), and active (P_L) and reactive (Q_L) powers of all the loads in the IEEE39 bus system is associated with a calculated \mathcal{R}_{sys}^d .

The calculations performed for various dispatches(d) for N-2 contingency scenarios in the IEEE39 bus system yield the risk of each contingency pair. From this, the combined system risk $\mathcal{R}_{sys}^{(d)}$ for each dispatch d is determined which is a comprehensive measure of grid vulnerability, computed by aggregating individual risk $\mathcal{R}^{(d)}$ components from two-line ($N - 2$) cascade contingencies. For each dispatch scenario (d),

$$\mathcal{R}_{sys}^{(d)} = \sum_{c=1}^{|S|} \mathcal{R}^{(d)} = \sum_{c=1}^{|S|} p(s_{k,i})_c \times \mathcal{I}_c^{(d)} \quad (4.3)$$

where $p(s_{k,i})_c$ is the probability of contingency scenario c for a dispatch d and $\mathcal{I}_c^{(d)}$ is impact of contingency c for dispatch d when $|S|$ contingency scenarios are considered per dispatch.

CS3.1: Generator Dispatch and Variability in Risk

A parallel coordinate plot is used to study the variability of different generators with respect to total system risk $\mathcal{R}_{sys}^{(d)}$. The plot connects multiple dispatch values (G01, G02, ... G10) with corresponding risk levels. Each line represents a single dispatch scenario (d), and colors are used to categorize low-risk, medium-risk, and high-risk cases based on a Quantile-based discretization function. This function partitions the dataset into three distinct risk categories using quantiles. Given a set of total system risk $\mathcal{R}_{sys}^{(d)}$ values across D different dispatch scenarios:

$$\mathcal{R}_{sys} = \{\mathcal{R}_{sys}^{(1)}, \mathcal{R}_{sys}^{(2)}, \dots, \mathcal{R}_{sys}^{(D)}\}$$

Where D is the total number of dispatch scenarios, quantile thresholds are defined as:

$$Q_1 = Q(1/3), \quad Q_2 = Q(2/3)$$

$Q(p)$ represents the empirical quantile function, which returns the value below which a proportion p of the data falls. The risk levels are then assigned as follows:

$$\text{Low Risk: } \mathcal{R}_{sys}^{(d)} \leq Q_1$$

$$\text{Medium Risk: } Q_1 < \mathcal{R}_{sys}^{(d)} \leq Q_2$$

$$\text{High Risk: } \mathcal{R}_{sys}^{(d)} > Q_2$$

Where each dispatch scenario d is categorized into one of these three bins. By applying this discretization, we effectively divide the dataset into three equal-sized groups, ensuring a balanced classification of risk levels. This approach helps in visualizing the distribution of system risk while preserving the relative structure of the data.

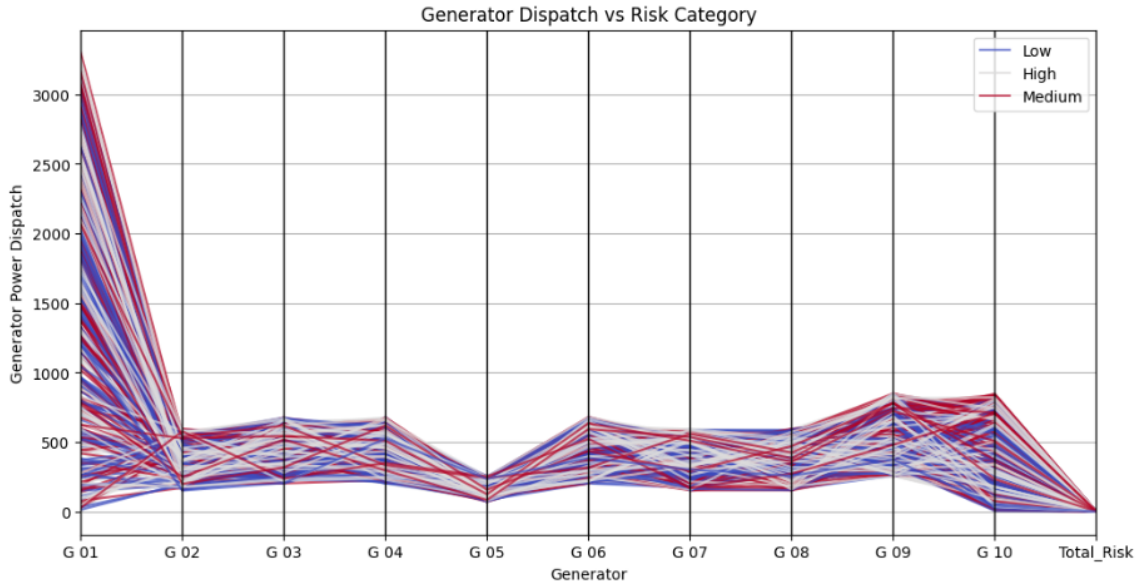


Figure 4.10: Parallel Coordinate Plot: Generator Dispatch vs Risk Category

This visualization in Figure 4.10 highlights that Generator G01 exhibits the highest variability across different risk levels and that Generator G05 shows the least variability. This is likely due to its larger capacity of G01, which allows for a wider range of dispatch values, leading to greater observed variation and vice versa. To further investigate, an additional analysis can be conducted to explore how system risk $\mathcal{R}_{sys}^{(d)}$ varies with respect to system generation or by isolating the effect of a single generator.

CS3.2: PCA-Based Contour Plot for Quick Operator Decision-Making

In this study, as described in 3.4.2, PCA condenses all generator and load dispatch variables into two principal components ($PC1$ and $PC2$), which represent a transformed version of the dispatch strategy. By plotting system risk on a contour map in PCA space, operators can quickly check if a dispatch scenario lies in a high-risk region (red zones) or low-risk region (blue zones). Notably, the contour

cannot be accurately plotted using a convex optimization function, highlighting the necessity of this methodology for effectively capturing and representing system risk.

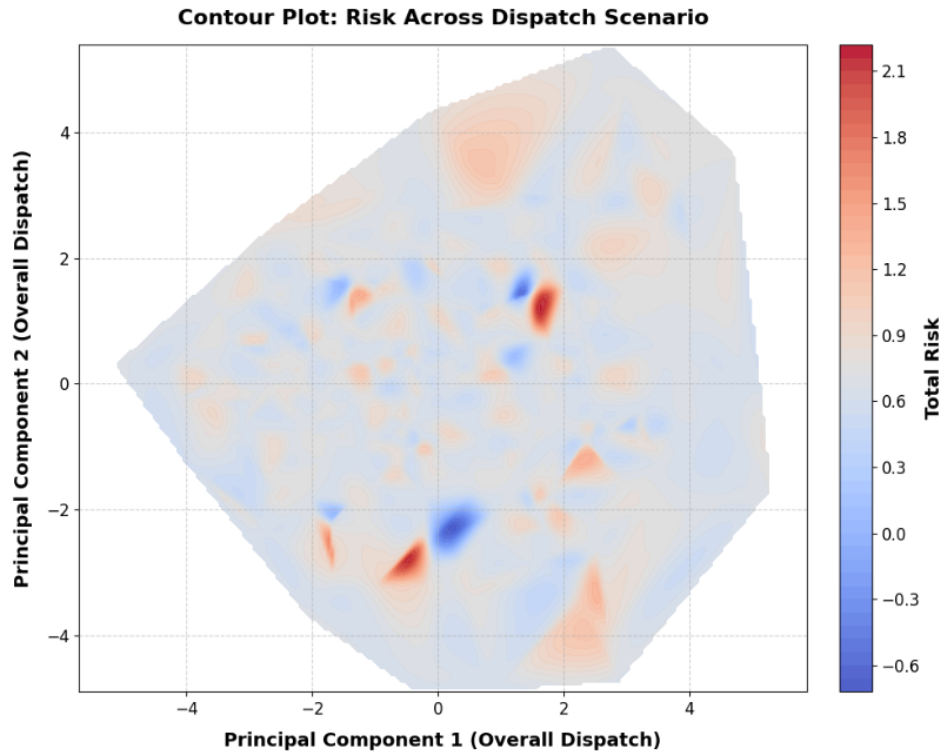


Figure 4.11: PCA-Based Contour Plot: Risk Across Dispatch Scenarios

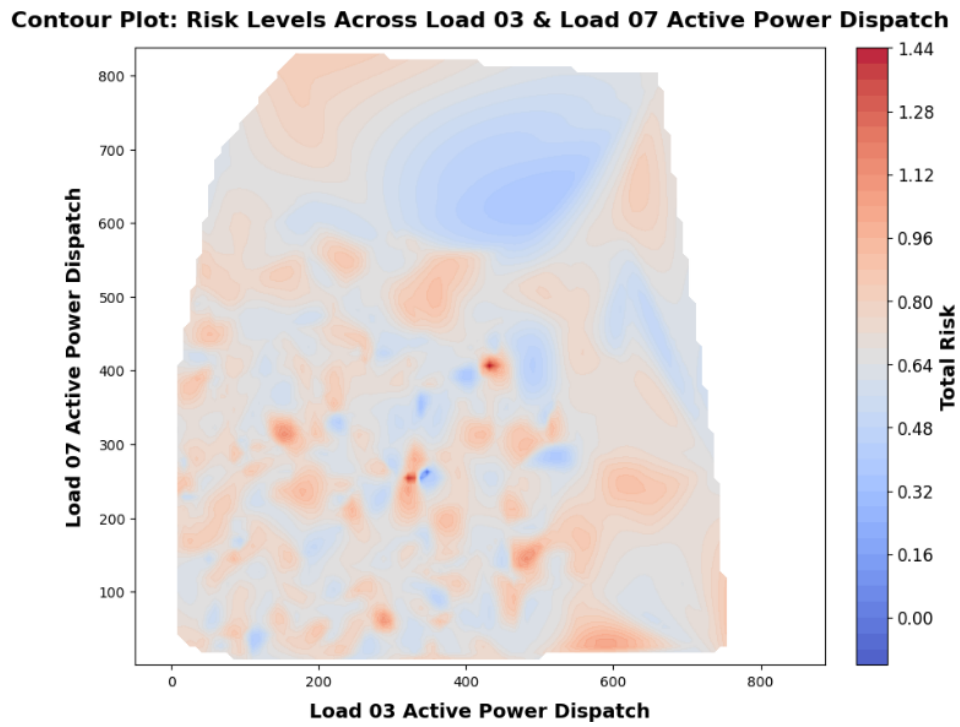


Figure 4.12: Contour Plot of Load 03 and 07 in IEEE39 bus network: Risk Across Dispatch Scenarios

The contour plot in Figure 4.11 visualizes $\mathcal{R}_{sys}^{(d)}$ across different dispatch scenarios using Principal Component Analysis (PCA). The x-axis represents $PC1$, while the y-axis represents $PC2$, both derived from dispatch data. The contour levels indicate varying risk levels, interpolated onto a balanced grid using cubic interpolation. The color bar provides a reference for risk intensity.

The contour plot in Figure 4.12 represents the total risk levels based on active power dispatch (P_L) at two specific load points (Load 03 and Load 07). Similar to the previous plot, a cubic interpolation method is applied to estimate risk values across the grid. The contour levels depict changes in system risk $\mathcal{R}_{sys}^{(d)}$ intensity, offering a structured visualization for assessing risk variations across different dispatch scenarios.

CS3.3: Identifying Key Dispatch Factors Influencing Risk

To further quantify the impact of different dispatch variables on risk, a machine learning-based feature importance ranking can be utilized. A Random Forest Regressor is trained on dispatch variables to predict total risk, and the importance of each variable is extracted from this regressor.

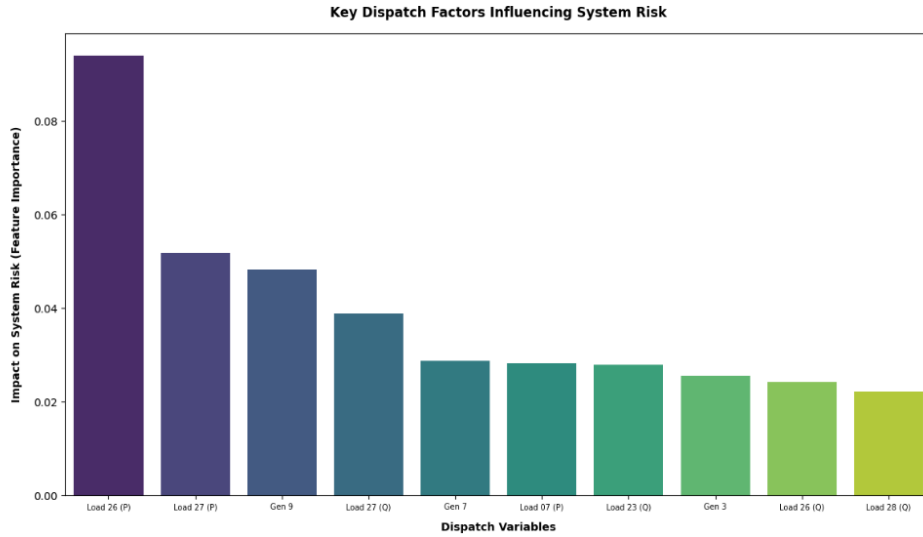


Figure 4.13: Feature Importance Plot: Key Dispatch Factors Influencing Risk

The feature importance plot in Figure 4.13 visualizes the relative effect of various dispatch variables on $\mathcal{R}_{sys}^{(d)}$, derived from feature importance indices. The variable names represent specific dispatch factors. It raises the question of whether sensitivities in these variables could lead to larger steps or changes in the risk levels, as observed in the contour plots. This highlights the importance of understanding how changes in individual dispatch variables, as indicated by feature importance, may influence the broader risk landscape depicted in the contour plots. The feature importance plot, therefore, offers insight into which dispatch factors might warrant further investigation or closer monitoring to mitigate system risk effectively.

5

Discussion and Limitation

In this chapter, the results gathered in the previous chapter are analysed and provided with some interpretations. These interpretations can allow us to answer the thesis's research questions, mention the thesis's limitations, and provide recommendations.

5.1. Interpretation of Results

5.1.1. Probability of a cascade contingency scenario

Given this exponential complexity, an initial contingency analysis is necessary to identify overloaded components. By setting a threshold on the overloading levels, only the most critical lines are selected for further analysis, significantly reducing computational effort while maintaining accuracy in cascade modeling.

Motifs exhibit a higher probability of participating in a cascade compared to other subgraphs. This trend is evident in the empirical probability results, where patterns with a higher presence in the network correspond to higher empirical probabilities. Additionally, the heavy tail of the subgraph diameter distribution reveals that as the diameter increases, especially from $d = 6$, the probability of disconnected subgraphs decreases quickly, yet remains nonzero. Notably, for the IEEE39 bus network, the probability peaks at a diameter of 6, subtly indicating that cascades can propagate across distant lines up to six edges apart rather than being restricted to directly connected elements. Cascades are not solely localized events but can also involve spatially separated lines due to their higher prevalence in the network. Table 4.5 further supports this by illustrating that disconnected subgraphs with diameters

of 5 and 6 are the most frequently occurring in the network, thereby increasing their likelihood of being involved in cascading failures.

The key findings in Table 4.6, which analyzes the probability of cascade propagation based on subgraph patterns and their diameters, highlight an important trend: certain patterns have a high probability of occurrence, while others are almost nonexistent. This supports the initial hypothesis that motifs are expected to be higher in probability. Specifically, motifs play a dominant role in cascade propagation and should be identified for further analysis or mitigation. Furthermore, as the cascade size increases, patterns with larger diameters demonstrate a higher probability of occurrence. This can be attributed to the network's structural properties, where larger diameter subgraphs tend to appear more frequently due to the sheer number of possible connections within a dense network. Moreover, as cascading failures progress, they are more likely to involve multiple paths and bypass direct connectivity constraints, enabling failures to propagate along alternate routes. This aligns with the statistical distribution of subgraphs, where larger-diameter motifs inherently have a higher representation, thereby increasing their likelihood of participating in cascades.

5.1.2. Effectiveness of dynamic impact metric

The first method to test for effectiveness is to show a correlation between the proposed impact metric (\mathcal{I}) and RMS-based Maximum rotor angle impact ($\Delta\theta_f$) for verification, which can be considered as standard. The regression trend line in Figure 4.5 reveals a relatively very weak correlation. Despite this, the proposed impact metric appears to effectively capture the severity of dynamic disturbances under specific high-impact contingency conditions obtained through the proposed impact metric. This can be especially beneficial in real-time monitoring and decision-making processes for power system operators.

The proposed impact metric does not cover the entire system dynamics range, such as inertia, damping coefficients, and pre-fault load flows, yet it can be valuable in some respects. It offers a practical way to estimate the severity of disturbances, particularly in critical scenarios that lead to cascading failures. These scenarios often form the foundation of large-scale system instability. While a more comprehensive analysis, which incorporates these additional system characteristics, would provide a fuller understanding of rotor angle deviations and dynamic stability, the proposed metric is advantageous in its ability to significantly reduce computational time. It provides a reasonable proxy for assessing the impact of faults in cases where time-sensitive decisions are needed without compromising the overall stability evaluation in these critical situations. Thus, while the proposed metric may not fully replace more detailed stability analysis, it is a useful, computationally efficient alternative for assessing dynamic impact in high-stakes fault scenarios.

The observations from the plot in Figure 4.6 highlight the computational efficiency and practicality of the proposed impact's computation time compared to traditional RMS-based simulations. Notably, the

proposed impact computation line (represented in blue) consistently remains significantly below the RMS-based maximum rotor angle impact computation line (depicted in red) across all analyzed contingencies. This indicates that the proposed method provides a highly computationally efficient alternative for assessing system impact, particularly in scenarios where rapid estimations are required. Despite the clear advantage in computation speed, RMS-based simulations remain essential for detailed dynamic stability analysis as they provide a more comprehensive assessment of system response, particularly in capturing the full extent of rotor angle deviations following disturbances. However, the proposed impact metric serves as an effective rapid estimation tool that allows system operators to efficiently prioritize critical contingencies. By leveraging this metric, high-risk contingencies can be quickly identified, and computational resources can be allocated more effectively, reserving detailed simulations only for the most critical cases that require deeper analysis. A direct comparison of mean computation times mentioned in the results further reinforces the efficiency of the proposed approach. This stark contrast in computational time highlights the practical advantage of the proposed method, particularly in real-time monitoring and contingency analysis. The ability to rapidly estimate impact using the proposed metric makes it a valuable tool for system operators who require fast decision-making capabilities, especially in large-scale power networks where running full RMS simulations for every contingency may be computationally intensive.

5.1.3. Implications of risk calculation and ranking

The results in Table 4.7 highlight the effectiveness of the risk-based method in identifying high-impact contingencies, outperforming the random selection approach. The risk-based method achieves a precision of 71.89% and recall of 71.45%, demonstrating its ability to detect critical cascades. In contrast, the random selection method, with a precision of 23.93% and recall of 26.39%, struggles to reliably identify high-risk contingencies. Its low selectivity results in frequent misclassifications. The risk-based method's structured, probabilistic framework ensures a more precise identification of critical events, reducing the likelihood of cascading failures and improving power system reliability.

Beyond outperforming random selection, the risk-based approach is also comparable to, and in certain aspects superior to, the overloading-based selection method. The overloading approach exhibits high selectivity (89.34%), meaning it is highly accurate in classifying severe cases. However, its recall is significantly lower at 41.37%, indicating that it fails to capture a substantial number of critical contingencies. This limitation makes it less suitable for real-world applications, where missing critical contingencies can have catastrophic consequences. The risk-based method, while slightly lower in selectivity, achieves a much higher recall of 71.45%, ensuring a more comprehensive identification of contingencies. The risk-based approach strikes a better balance by prioritizing the identification of high-risk contingencies, reducing the likelihood of missed alarms. In practical scenarios, where avoiding missed alarms is critical, this method can ensure more reliable detection of critical cascades, enhancing the overall system's ability to mitigate potential cascades. Another key advantage of the risk-based method is its compu-

tational efficiency as shown using the average computation times in results. The drastic reduction in computational time makes the risk-based model particularly advantageous for time-sensitive grid operations, where rapid assessment and response are critical. The ability to provide fast and reliable results without compromising accuracy makes the risk-based method not only a competitive alternative but a preferred choice over the overloading-based approach.

The analysis of Spearman correlations between risk rankings and actual impact rankings in Figure 4.9 reveals a stark difference between randomly selected contingencies and those identified using the risk-based method. The graph confirms that the random selection approach does not provide any reliable predictive power in determining critical contingencies. The lack of correlation underscores the inherent limitation of arbitrary selection, which fails to capture the underlying system vulnerabilities and does not prioritize contingencies based on their potential impact. In stark contrast, a near-perfect correlation suggests that the risk-based rankings are highly aligned with the dynamic rotor angle impact, making them a far superior method for identifying critical contingencies. The ability to accurately rank contingencies based on their potential impact provides grid operators with a powerful decision-making framework, allowing for targeted mitigation strategies and enhanced resilience of the power system.

Operators should closely monitor the dispatches to ensure it does not exceed critical thresholds that contribute to high-risk conditions. The contour plot in Figure 4.11 represents the risk associated with different dispatch strategies. The dispatch scenario's principal components can be projected onto this plot, and if it falls into a red region, it indicates high system risk, whereas a dispatch falling into a blue region signifies low system risk. This is particularly useful for real-time monitoring and quick operator decision-making.

5.2. Answers to the Research Questions

How can contingency motif subgraphs be identified and their probabilities be used to model and predict cascading failures?

Probabilistic Contingency Motifs (PCMs) are identified based on their shape and connection between nodes. [29] discusses the identification of motifs using historical outage data. However, the unavailability of such data in certain scenarios necessitates alternative approaches for modeling outage patterns, with contingency analysis being a key method. Once contingency analysis provides critical cascade scenarios for an $N - k$ cascade, the empirical probabilities of subgraphs are compared to uniform probabilities to check the motif. The probability of this motif $P(S_{k,i})$ is the probability that those transmission lines participate in a specific cascade contingency scenario.

For an $N - k$ cascade scenario with $k - 1$ lines already outaged, the probability of the k^{th} line is determined using the motif probability of k -line outages, probability of that cascade pattern given k -line outages, probability of the subgraph diameter given for that cascade pattern and the probability

of contingency for the specific pattern with the specific diameter. This probability of the k^{th} line is the overall cascade probability of the specific PCM with k^{th} line as the next failure in the cascade.

How does the proposed impact metric correlate with traditional impact metrics when ranking contingencies?

The proposed impact metric for cascades, derived from I_{SC} of generators, using short-circuit analysis, LODFs and electrical distance calculations, shows almost no correlation with impact metric based on dynamic RMS simulations such as rotor angle deviations. This suggests the influence of additional system parameters such as inertia, damping, and pre-fault load conditions while calculating the dynamic impact. Despite this, the proposed metric effectively captures severe cascade contingencies when risk rankings are analysed and is computationally extremely cheap. Thus, the proposed metric can serve as a reasonable proxy.

The risk ranking comparison suggests that while traditional RMS-based simulations remain crucial for detailed stability analysis, the proposed metric enables rapid pre-screening of contingencies, significantly improving the efficiency of risk assessment. The computational efficiency of the metric further strengthens its applicability.

How well does this risk-based probabilistic model's ranking of cascading events align with the rankings obtained by RMS simulation results in evaluating system security and resilience?

The risk-based probabilistic model demonstrates a strong alignment with the rankings derived from RMS simulation results in evaluating system security and resilience. By leveraging a structured probabilistic framework that integrates PCM probabilities with short-circuit impact, the model effectively identifies high-risk contingencies. The comparison with overloading-based selection and random selection highlights its superior balance between the precision of identification and sensitivity to high-risk cascade scenarios. Unlike the random selection approach, which lacks systematic prioritization and misclassifies contingencies, the risk model consistently captures critical contingencies with high reliability.

The comparison of risk rankings with rankings obtained from maximum rotor angle impact further supports the model's robustness. The Spearman correlation analysis reveals a strong association between the rankings produced by the former risk-based method and actual system behavior, as measured by the latter. This high correlation indicates that the probabilistic ranking method successfully reflects real-world cascading effects for critical scenarios. The probabilistic model's ability to prioritize contingencies based on risk ensures that operators can effectively focus on the most hazardous scenarios, improving decision-making in preventive and corrective actions.

Additionally, the risk-based method's computational efficiency enhances its practical applicability for security assessments. Unlike overloading-based selection, which relies on iterative load-flow simulations

and is computationally intensive as k increases for an $N - k$ cascade, the risk-based approach computes contingency probabilities in a single step. The impact calculations are also highly computable, and this advantage allows for faster risk assessments. The model's ability to balance computational efficiency with accuracy in ranking cascading events makes it a valuable tool for improving system resilience and risk-informed operational strategies.

5.3. Limitations

The risk-based model can outperform randomized selection and is comparable to overloading-based impact (η_f) and RMS simulation-based $\Delta\theta_f$ in terms of providing a ranking of critical cascade contingency scenarios. Although it is highly computationally efficient compared to both methods, several limitations need to be overcome to effectively use the risk-based rankings.

Inability to calibrate motif frequencies for higher order cascades

The methodology is well-supported for $N - 2$, $N - 3$, and potentially $N - 4$ outages, as these events that occur frequently generate meaningful probability distributions. A central premise of motif-based outage probability estimation is that one has access to a dataset of multiple-line outages from which to extract frequent subgraph patterns. The $n_{k,i}$ for empirical probability ($P(S_{k,i} | k)$) calculation in Equation 2.25 for determining motifs is the number of observed k -line outages matching subgraph $S_{k,i}$. For larger k (e.g., $N-5$ and beyond), the available data becomes sparse. If no historical cascade outage data exists, then $P(S_{k,i} | k)$ is effectively not defined, and it is difficult to estimate probabilities in an empirical, data-driven way. This is why all outages with $k > 4$ are assumed to be of the same subgraph $S_{4,*}$ and this pattern is not considered a motif.

Assumption of stationarity in outage patterns

The approach assumes that the historical outage patterns are representative of future outages. An assumption that a stationary probability distribution of contingency motifs exists for the motifs is made. However, power system configurations evolve due to grid upgrades, changing protection schemes, and new operational systems. These changes may render past cascade probabilities, which have been calculated based on outage data on the previous network configurations, inaccurate for risk estimation.

Static fault model for short-circuit contributions

We use generator short-circuit currents during the outage ($I_{SC,gf}$). However, these calculations assume a static Thevenin equivalent of the network state. Real faults occurring as a part of cascades evolve dynamically, with the tripping of protective relays, activation of reclosures and multiple simultaneous faults. In practice, the time-domain progression of short-circuit currents and the transient response of generators may alter the actual fault currents significantly as $Z_{eq,gf}$ varies throughout the fault due to system reconfiguration. Relying on a single instance I_{SC} may, therefore, underestimate or overestimate the severity of a multi-line event, especially deeper into the cascade when topology has varied.

Furthermore, the model treats sub-transient reactance X'' among generators as homogeneous. Large discrepancies in these reactances can cause I_{SC} to concentrate disproportionately at certain fault points, challenging the simplified notion in Equation 3.23 with a nominal value of E_g . In reality, generator saturation levels, saliency effects, and exciter limitations, which tend to vary X'' , could alter I_{SC} .

Underrepresenting rare contingencies and preemptive actions

When ranking cascading events solely via a single measure of risk $\mathcal{R}_f = P(s_{k,i}) \times \mathcal{I}_f$, low-probability but contingencies with moderate impact can overshadow high-impact cascade scenarios, which can prove catastrophic but very high probabilities. Furthermore, without the uncertainty quantification of risks, such as standard deviations or confidence bounds for both $P(s_{k,i})$ and \mathcal{I}_f , the resulting ranked list can encourage over-reliance on point estimates. This can discourage consideration of scenarios that have higher variance or occur in the tail and undermine the risk-informed decision-making process. The ranking also does not account for the adaptive operator interventions or advanced digital protection, which can isolate faults rapidly.

6

Conclusion

This thesis, titled "Probabilistic Framework for Assessing Cascading Failures in Power Systems", established a probabilistic contingency motif (PCM) framework for modeling and predicting the probabilities of critical cascade contingencies in transmission networks. It uses existing research on the identification of contingency motifs by comparing empirical and uniform probabilities of subgraph patterns and then calculating their probabilities using its subgraph type and its diameter. This probability is extended to cascade probability with an assumption that for $N - k$ cascade contingency and $k - 1$ outages already occurred, the multi-line motif probability or the probability of k^{th} line outage is the probability $P(s_{k,i})$ for the cascade subgraph $s_{k,i}$. Concurrently, an impact metric \mathcal{I}_f for the fault at f on the k^{th} line in cascade calculated using the generator short-circuit currents ($I_{SC,gf}$), sensitivities using LODFs (w_{gf}), and electrical distance (D_{gf}) calculated using the modified Z_{bus} matrix after fault occurrence is proposed. It offers a fast-yet-robust measure of how severely each potential cascade may undermine network stability. The impact and the probability are used to calculate the risk \mathcal{R}_f of each cascade contingency. Although this static \mathcal{R}_f metric shows minimal correlation with maximum rotor angle deviation $\Delta\theta_f$, a fully dynamic metric calculated using RMS simulations, it proves adequate as a computationally cheap proxy due to its high correlation in rankings of critical cascade contingencies with the same metric.

While RMS simulations remain essential for in-depth transient and voltage stability assessment, the proposed method can be utilized as a pre-screening mechanism, drastically reducing computational expense. Consequently, the PCM approach offers a new technique for risk-informed operation, hazardous cascade identifications and dispatch planning.

6.1. Recommendations

A key area of future research involves improving the accuracy of cascade probability modeling by incorporating dynamic and time-varying system conditions into the probability calculations. Rather than assuming stationarity, one can develop parametric expressions for calculating probabilities which depend on parameter ω , $P(s_{k,i} | \omega)$ where ω is an umbrella variable corresponding to variations in weather, load, equipment condition and others. There can also be an option to explore time-series probabilistic models, $P(s_{k,i} | t)$ for calculating probabilities of critical cascade contingencies dependent on schedules. Additionally, methods that explicitly capture multi-line correlations, such as partial discharge, circuit breaker failures, or hidden failures, could yield more realistic joint distributions of outages. Mathematically, this may be approached through a Bayesian framework to obtain the overall marginal probability

$$P(s_{k,i}) = \int P(s_{k,i} | \omega) f(\omega) d\omega \quad (6.1)$$

ω represents one or more conditions or variables upon which the outage $s_{k,i}$ might depend, and $f(\omega)$ represents the PDF describing the frequency of each state. Incorporating such conditions in near real-time could significantly refine how probabilities scale across different dispatch scenarios.

Beyond probabilistic modeling, enhanced impact assessment can be pursued by merging short-circuit methodologies shown in the thesis along with time-domain simulations. For example, $I_{SC,gf}$ could be integrated into a transient stability method, allowing each outage scenario to trigger relay operations. This can be supplemented with $LODF$ with both linear and non-linear studies, iterative power-flow calculations measuring line overloading and tracking how each subsequent tripping reconfigures the network. Doing so would update the power flows and generator redispatch in real time, giving a more precise representation of \mathcal{I}_f . Furthermore, stochastic load variations or changes in sub-transient reactances could allow for confidence intervals around both the probability and impact values, improving the robustness of risk \mathcal{R}_f .

Finally, ranking and \mathcal{R} -based decision-making can benefit from multi-objective and tail-sensitive analysis. Future methods could incorporate risk-aversion methods or distributional ranking using priority metrics. This can ensure that the scenarios residing in the tail still receive enough attention in system planning and show up higher in risk rankings. Operator interventions such as remedial action schemes or redispatch can be integrated directly into the ranking schemes so that the system can dynamically mitigate the highly ranked contingency scenarios before they propagate deeper into the network. Together, the refinements would enable a more holistic approach allowing an improved outage probability modelling, refined impact calculation and sophisticated risk ranking and critical cascade prioritization for diverse dispatch conditions.

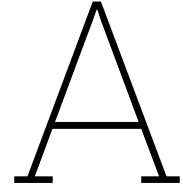
References

- [1] US Force and Roger Anderson. "Final Report on the Blackout in the United States and Canada: Causes and Recommendations Ch 1-3". In: (Aug. 2004).
- [2] MZ Zakariya and J Teh. "A Systematic Review on Cascading Failures Models in Renewable Power Systems with Dynamics Perspective and Protections Modeling". In: *Electric Power Systems Research* 214 (2023), p. 108928. DOI: 10.1016/j.epsr.2022.108928.
- [3] Yakup Koç et al. "A topological investigation of phase transitions of cascading failures in power grids". In: *Physica A: Statistical Mechanics and its Applications* 415 (2014), pp. 273–284. ISSN: 0378-4371. DOI: <https://doi.org/10.1016/j.physa.2014.07.083>. URL: <https://www.sciencedirect.com/science/article/pii/S0378437114006694>.
- [4] Hengdao Guo et al. "A critical review of cascading failure analysis and modeling of power system". In: *Renewable and Sustainable Energy Reviews* 80 (2017), pp. 9–22. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2017.05.206>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032117308432>.
- [5] S. Kalyani and K. Shanti Swarup. "Binary SVM Approach for Security Assessment and Classification in Power Systems". In: (2009), pp. 1–4. DOI: 10.1109/INDCON.2009.5409433.
- [6] Xi Zhang, Choujun Zhan, and Chi Tse. "Modeling the Dynamics of Cascading Failures in Power Systems". In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 7 (Mar. 2017), pp. 1–13. DOI: 10.1109/JETCAS.2017.2671354.
- [7] Jaime De La Ree et al. "Catastrophic Failures in Power Systems: Causes, Analyses, and Countermeasures". In: *Proceedings of the IEEE* 93 (June 2005), pp. 956–964. DOI: 10.1109/JPROC.2005.847246.
- [8] Mohammad Alam et al. "Effects of VSC Based HVDC System on Distance Protection of Transmission Lines". In: *International Journal of Electrical Power & Energy Systems* 92 (May 2017), pp. 245–260. DOI: 10.1016/j.ijepes.2017.04.012.
- [9] JS Thorp et al. "Anatomy of power system disturbances: importance sampling". In: *International Journal of Electrical Power & Energy Systems* 20.2 (1998), pp. 147–152. ISSN: 0142-0615. DOI: [https://doi.org/10.1016/S0142-0615\(97\)00034-3](https://doi.org/10.1016/S0142-0615(97)00034-3). URL: <https://www.sciencedirect.com/science/article/pii/S0142061597000343>.

- [10] Naeem Md Sami and Mia Naeini. "Machine learning applications in cascading failure analysis in power systems: A review". In: *Electric Power Systems Research* 232 (2024), p. 110415. ISSN: 0378-7796. DOI: <https://doi.org/10.1016/j.epsr.2024.110415>.
- [11] M. R. Aghamohammadi, S. Fazel Mahdavizadeh, and Z. Rafiee. "Controlled Islanding Based on the Coherency of Generators and Minimum Electrical Distance". In: *IEEE Access* 9 (2021), pp. 146830–146840. DOI: 10.1109/ACCESS.2021.3119529.
- [12] Faycal Znidi et al. "Multi-layer spectral clustering approach to intentional islanding in bulk power systems". In: *Journal of Modern Power Systems and Clean Energy* 7.5 (2019), pp. 1044–1055. DOI: 10.1007/s40565-019-0554-1.
- [13] Rezoan A. Shuvro et al. "Predicting Cascading Failures in Power Grids using Machine Learning Algorithms". In: (2019), pp. 1–6. DOI: 10.1109/NAPS46351.2019.9000379.
- [14] R. Kinney et al. "Modeling cascading failures in the North American power grid". In: *The European Physical Journal B* 46.1 (July 2005), pp. 101–107. ISSN: 1434-6036. DOI: 10.1140/epjb/e2005-00237-9.
- [15] Andrej Stankovski et al. "Power blackouts in Europe: Analyses, key insights, and recommendations from empirical evidence". In: *Joule* 7.11 (2023), pp. 2468–2484. ISSN: 2542-4351. DOI: <https://doi.org/10.1016/j.joule.2023.09.005>.
- [16] M. Kalantar Neyestanaki and A. M. Ranjbar. "An Adaptive PMU-Based Wide Area Backup Protection Scheme for Power Transmission Lines". In: *IEEE Transactions on Smart Grid* 6.3 (2015), pp. 1550–1559. DOI: 10.1109/TSG.2014.2387392.
- [17] O. Alizadeh Mousavi, R. Cherkaoui, and M. Bozorg. "Blackouts risk evaluation by Monte Carlo Simulation regarding cascading outages and system frequency deviation". In: *Electric Power Systems Research* 89 (2012), pp. 157–164. ISSN: 0378-7796. DOI: <https://doi.org/10.1016/j.epsr.2012.03.004>.
- [18] Meng-Yun Lee et al. "Transmission Line Fault Classification Using Conformer Convolution-Augmented Transformer Model". In: *Applied Sciences* 14 (May 2024), p. 4031. DOI: 10.3390/app14104031.
- [19] Woo-Hyun Kim et al. "LSTM-Based Fault Direction Estimation and Protection Coordination for Networked Distribution System". In: *IEEE Access* 10 (2022), pp. 40348–40357. DOI: 10.1109/ACCESS.2022.3166836.
- [20] Anjan Sahoo and Sudhansu Samal. "Online fault detection and classification of 3-phase long transmission line using machine learning model". In: *Multiscale and Multidisciplinary Modeling, Experiments and Design* 6 (Dec. 2022), pp. 1–12. DOI: 10.1007/s41939-022-00132-x.
- [21] Masoud Najafzadeh et al. "Fault Detection, Classification and Localization Along the Power Grid Line Using Optimized Machine Learning Algorithms". In: *International Journal of Computational Intelligence Systems* 17 (Mar. 2024). DOI: 10.1007/s44196-024-00434-7.

- [22] Abderrahmane El Rhatrif, Bouchra Bouihi, and Mohammed Mestari. "Challenges and Limitations of Artificial Intelligence Implementation in Modern Power Grid". In: *Procedia Computer Science* 236 (2024). International Symposium on Green Technologies and Applications (ISGTA'2023), pp. 83–92. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2024.05.008>.
- [23] Mahshid Rahnamay-Naeini et al. "A probabilistic model for the dynamics of cascading failures and blackouts in power grids". In: (2012), pp. 1–8. DOI: 10.1109/PESGM.2012.6345574.
- [24] Sudha Gupta et al. "Support-Vector-Machine-Based Proactive Cascade Prediction in Smart Grid Using Probabilistic Framework". In: *IEEE Transactions on Industrial Electronics* 62.4 (2015), pp. 2478–2486. DOI: 10.1109/TIE.2014.2361493.
- [25] David L. Pepyne. "Topology and cascading line outages in power grids". In: *Journal of Systems Science and Systems Engineering* 16 (2007), pp. 202–221. DOI: 10.1007/s11518-007-5044-8.
- [26] Géza Ódor et al. "Improving power-grid systems via topological changes or how self-organized criticality can help power grids". In: *Physical Review Research* 6.1 (2024). DOI: 10.1103/physrevresearch.6.013194.
- [27] Shenhao Yang et al. "Blocking cascading failures with optimal corrective transmission switching considering available correction time". In: *International Journal of Electrical Power & Energy Systems* 141 (2022), p. 108248. ISSN: 0142-0615. DOI: <https://doi.org/10.1016/j.ijepes.2022.108248>.
- [28] Yumeng Liu, Tao Wang, and X.P. Gu. "A risk-based multi-step corrective control method for mitigation of cascading failures". In: *IET Generation, Transmission & Distribution* 16 (Oct. 2021). DOI: 10.1049/gtd2.12327.
- [29] Kai Zhou, Ian Dobson, and Zhaoyu Wang. "The Most Frequent N-k Line Outages Occur in Motifs That Can Improve Contingency Selection". In: *IEEE Transactions on Power Systems* 39.1 (2024), pp. 1785–1796. DOI: 10.1109/TPWRS.2023.3249825.
- [30] Vaiman et al. "Risk Assessment of Cascading Outages: Methodologies and Challenges". In: *IEEE Transactions on Power Systems* 27.2 (2012), pp. 631–641. DOI: 10.1109/TPWRS.2011.2177868.
- [31] Chengdi Lai and Steven H. Low. "The redistribution of power flow in cascading failures". In: *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 2013, pp. 1037–1044. DOI: 10.1109/Allerton.2013.6736639.
- [32] Pierre Henneaux, Pierre-Etienne Labeau, and Jean-Claude Maun. "A level-1 probabilistic risk assessment to blackout hazard in transmission power systems". In: *Reliability Engineering & System Safety* 102 (2012), pp. 41–52. ISSN: 0951-8320. DOI: <https://doi.org/10.1016/j.res.2012.02.007>.
- [33] B.W. Tuinema et al. *Probabilistic Reliability Analysis of Power Systems: A Student's Introduction*. English. Springer, 2020. ISBN: 978-3-030-43497-7. DOI: 10.1007/978-3-030-43498-4.

- [34] Contingency Analysis in PowerFactory. *DigSILENT Power System Solutions*. <https://www.digsilent.de/en/contingency-analysis.html>.
- [35] Chu Donatus Iweh et al. "Analysis of contingency scenarios towards a suitable transmission pathway in the southern interconnected grid (SIG) of Cameroon". In: *e-Prime - Advances in Electrical Engineering, Electronics and Energy* 7 (2024), p. 100486. ISSN: 2772-6711. DOI: <https://doi.org/10.1016/j.prime.2024.100486>.
- [36] *Contingency Analysis in power systems — allumiax.com*. <https://www.allumiax.com/blog/contingency-analysis-using-power-system-studies>. [Accessed 13-01-2025].
- [37] Prabha Kundur, Neal J. Balu, and Mark G. Lauby. *Power System Stability and Control*. New York, NY: McGraw-Hill Education, 1994. ISBN: 978-0070359581.
- [38] B. Stott and O. Alsac. "Fast Decoupled Load Flow". In: *IEEE Transactions on Power Apparatus and Systems* PAS-93.3 (1974), pp. 859–869. DOI: 10.1109/TPAS.1974.293985.
- [39] Ian Dobson et al. "Obtaining Statistics of Cascading Line Outages Spreading in an Electric Transmission Network From Standard Utility Data". In: *IEEE Transactions on Power Systems* 31.6 (2016), pp. 4831–4841. DOI: 10.1109/TPWRS.2016.2523884.
- [40] T. Niimura et al. "Machine learning approach to power system dynamic security analysis". In: *IEEE PES Power Systems Conference and Exposition, 2004*. 2004. DOI: 10.1109/PSCE.2004.1397549.
- [41] Contingency Analysis. *DigSILENT | POWER SYSTEM SOLUTIONS*. <https://www.digsilent.de/en/contingency-analysis.html>.
- [42] Jae-Won Chang, Gyu-Sub Lee, and Seaseung Oh. "Bus Impedance Matrix Estimation Using Non-PMU Measurements". In: *IEEE Transactions on Power Systems* 38.2 (2023), pp. 1743–1746. DOI: 10.1109/TPWRS.2022.3233772.
- [43] Chapter 3: Network Admittance and Impedance Matrices - NPTEL Archives. *Node Elimination by Kron Reduction*. https://archive.nptel.ac.in/content/storage2/courses/108104051/chapter_3/3_6.html.
- [44] Rohde & Schwarz | YouTube. *Understanding CCDF*. <https://www.youtube.com/watch?v=-7U82ZzIbUs>.
- [45] OpenAI. *ChatGPT: Language Model for Text and Code Assistance*. <https://openai.com/chatgpt>. Accessed: 2025-03-18. 2025.
- [46] Gonzalo Seguel. *Zbus from Cases*. https://github.com/Gseguelg/casesZbus/blob/master/Zbus_from_cases.py. 2024.



Electrical Distance D_{gf} from Z_{bus}

```
1 import numpy as np
2 from scipy import sparse
3 import networkx as nx
4 import pandas as pd
5
6 def Case_1(Z_old, Zb):
7     """Case 1: Add impedance Zb from a new bus to reference."""
8     dim = Z_old.shape[0]
9     new_row = sparse.coo_matrix((1, dim), dtype=np.complex128)
10    new_col = sparse.coo_matrix((dim, 1), dtype=np.complex128)
11    return sparse.vstack([sparse.hstack([Z_old, new_col]), sparse.hstack([new_row, Zb])])
12
13 def Case_2(Z_old, Zb, node_idx):
14     """Case 2: Add impedance Zb from a new bus to an existing bus."""
15     new_row, new_col = Z_old.getrow(node_idx), Z_old.getcol(node_idx)
16     Zkk = Z_old.tolil()[node_idx, node_idx]
17     return sparse.vstack([sparse.hstack([Z_old, new_col]), sparse.hstack([new_row, Zkk + Zb])
18                             ])
19
20 def Case_3(Z_old, Zb, node_idx):
21     """Case 3: Add impedance Zb from a bus to reference."""
22     return kron_reduction(Case_2(Z_old, Zb, node_idx), Z_old.shape[0])
23
24 def Case_4(Z_old, Zb, node_l_idx, node_k_idx):
25     """Case 4: Add impedance Zb between two existing buses."""
26     Z_old_lil = Z_old.tolil()
```

```

26     new_row, new_col = Z_old.getrow(node_l_idx) - Z_old.getrow(node_k_idx), Z_old.getcol(
        node_l_idx) - Z_old.getcol(node_k_idx)
27     Zkk, Zll, Zlk = Z_old_lil[node_l_idx, node_l_idx], Z_old_lil[node_k_idx, node_k_idx],
        Z_old_lil[node_l_idx, node_k_idx]
28     return kron_reduction(sparse.vstack([sparse.hstack([Z_old, new_col]), sparse.hstack([
        new_row, Zkk + Zll - 2 * Zlk + Zb]))), Z_old.shape[0])
29
30 def kron_reduction(matrix, pivot_idx):
31     """Perform Kron's reduction to eliminate a row and column."""
32     matrix = matrix.tolil()
33     pivot_value = matrix[pivot_idx, pivot_idx]
34     dim = matrix.shape[0] - 1
35     reduced_matrix = sparse.lil_matrix((dim, dim), dtype=np.complex128)
36
37     for row in range(matrix.shape[0]):
38         if row == pivot_idx:
39             continue
40         reduced_matrix[row - (row > pivot_idx), :] = matrix[row, :] - (matrix[row, pivot_idx]
            / pivot_value) * matrix[pivot_idx, :]
41
42     return reduced_matrix
43
44 def construct_Zbus(graph, edge_order):
45     """Construct Zbus incrementally based on network connections."""
46     ref_node, nodes_used, first = -1, [], True
47     for node_a, node_b in edge_order:
48         weight = graph[node_a][node_b][0]['weight']
49         if first:
50             Zbus = sparse.coo_matrix((weight, (0, 0)), dtype=np.complex128)
51             nodes_used += [n for n in (node_a, node_b) if n != ref_node]
52             first = False
53             continue
54
55         if node_a == ref_node and node_b not in nodes_used or node_b == ref_node and node_a
            not in nodes_used:
56             Zbus = Case_1(Zbus, weight)
57         elif node_a not in nodes_used and node_b in nodes_used:
58             Zbus = Case_2(Zbus, weight, nodes_used.index(node_b))
59         elif node_b not in nodes_used and node_a in nodes_used:
60             Zbus = Case_2(Zbus, weight, nodes_used.index(node_a))
61         elif node_a in nodes_used and node_b in nodes_used:
62             Zbus = Case_3(Zbus, weight, nodes_used.index(node_a))
63
64         nodes_used += [n for n in (node_a, node_b) if n not in nodes_used and n != ref_node]
65
66     return Zbus, nodes_used
67

```

```

68 def create_zbus_line_data(lines):
69     """Convert line data to Zbus matrix representation."""
70     bus_index = {bus: idx for idx, bus in enumerate(set(branch[key] for branch in lines for
71                                                         key in ['bus1', 'bus2']))}
72     return pd.DataFrame([
73         'line': branch['loc_name'],
74         'sending_node': bus_index[branch['bus1']],
75         'receiving_node': bus_index[branch['bus2']],
76         'line_impedance': branch['R1'] + 1j * branch['X1']
77     ] for branch in lines])
78
79 def process_zbus_data(lines):
80     """Compute Zbus matrix and return impedance values."""
81     impedance_data = create_zbus_line_data(lines)
82     impedance_data.replace({'sending_node': {0: -1}, 'receiving_node': {0: -1}}, inplace=True)
83
84     sorted_data = impedance_data[impedance_data['sending_node'] == -1].iloc[:1]
85     remaining_data = impedance_data.drop(sorted_data.index)
86     visited_nodes = set(sorted_data[['sending_node', 'receiving_node']].values.flatten())
87
88     while not remaining_data.empty:
89         matching_rows = remaining_data[(remaining_data['sending_node'].isin(visited_nodes)) |
90                                         (remaining_data['receiving_node'].isin(visited_nodes))]
91         if matching_rows.empty:
92             break
93         sorted_data = pd.concat([sorted_data, matching_rows])
94         remaining_data = remaining_data.drop(matching_rows.index)
95         visited_nodes.update(matching_rows[['sending_node', 'receiving_node']].values.flatten())
96
97     sorted_data['line_impedance'] = sorted_data['line_impedance'].apply(complex)
98     G_zbus = nx.MultiGraph()
99     edge_order = [(row['sending_node'], row['receiving_node']) for _, row in sorted_data.
100                  iterrows()]
101
102     for _, row in sorted_data.iterrows():
103         G_zbus.add_edge(row['sending_node'], row['receiving_node'], weight=row['
104                         line_impedance'])
105
106     Zbus, nodes_used = construct_Zbus(G_zbus, edge_order)
107     Zbus_array = Zbus.toarray()
108
109     bus_pairs = {int(bus): row['sending_node'] for _, row in impedance_data.iterrows() for
110                  bus in row['line'].replace("Line_", "").split("_-")}
111
112     return pd.DataFrame([

```

```

108     [bus_a, bus_b, Zbus_array[bus_pairs[bus_a] - 1, bus_pairs[bus_b] - 1]]
109     for bus_a in range(1, max(bus_pairs) + 1)
110     for bus_b in range(1, max(bus_pairs) + 1)
111     if bus_a in bus_pairs and bus_b in bus_pairs
112 ], columns=["Bus_A", "Bus_B", "Zbus_impedance"])
113
114 def find_distance(line, bus_name, Z_Bus_Values, percentage_distance=50):
115     """Compute electrical distance based on Zbus matrix."""
116     bus1, bus2 = map(int, line.loc_name.replace("Line_", "").split("_-"))
117     Z_line = line.GetAttribute('R1') + 1j * line.GetAttribute('X1')
118
119     def get_Z(bus_a, bus_b):
120         return Z_Bus_Values.loc[(Z_Bus_Values["Bus_A"] == bus_a) & (Z_Bus_Values["Bus_B"] ==
121             bus_b), 'Zbus_impedance'].iloc[0]
122
123     Z11, Z22, Z12, Z21 = get_Z(bus1, bus1), get_Z(bus2, bus2), get_Z(bus1, bus2), get_Z(bus2,
124         bus1)
125     Zb1, Zb2, Zbb = get_Z(bus_name, bus1), get_Z(bus_name, bus2), get_Z(bus_name, bus_name)
126
127     p = percentage_distance / 100
128     Zff = (1 - p)**2 * Z11 + p**2 * Z22 + 2*p*(1 - p) * Z12 + p*(1 - p) * Z_line
129     Zbf = (1 - p) * Zb1 + p * Zb2
130
131     return Zbb + Zff - 2 * Zbf

```

B

Impact weight w_{gf} from $LODFs$

```
1 import os
2 import pandas as pd
3
4 def weights_from_LODF(M, Branch_list, n_loc):
5     """Find sensitivities using LODF (Line Outage Distribution Factors)."""
6
7     Branch_list = [branch.loc_name for branch in Branch_list]
8     sensitivities = pd.DataFrame(columns=["Generator"] + Branch_list)
9
10    for gen in M.app.GetCalcRelevantObjects('*.ElmSym'):
11        new_row = {"Generator": gen.loc_name}
12
13        # Run LODF calculation
14        com_vstab = M.app.GetFromStudyCase('ComVstab')
15        com_vstab.SetAttribute('iopt_method', 0)
16        com_vstab.SetAttribute('calcLODF', 1)
17        com_vstab.SetAttribute('p_bus', gen)
18        com_vstab.SetAttribute('isContSens', 1)
19        com_vstab.Execute()
20
21        # Get results
22        current_elmres = M.app.GetCalcRelevantObjects('*.ElmRes')[0]
23
24        if os.path.isfile("trial.csv"):
25            os.remove("trial.csv")
26
27        res_com = M.app.GetFromStudyCase('*.ComRes')
```

