RRE
Reliability & Risk Engineering

ETH zürich

Cascades

A platform for power system risk analyses and transmission expansion planning

Authors:
Blazhe Gjorgiev
Andrej Stankovski
Giovanni Sansavini

Contributors:
Bing Li
Alexander David

Contents

Abbreviations

AC	Alternating Current
BCS	Best Compromise Solution
BLX	Blending Crossover
CCDF	Complementary Cumulative Distribution Function
CDF	Cumulative Distribution Function
CFS	Cascading Failure Simulation
DC	Direct Current
DNS	Demand Not Served
ENTSO-E	European Network of Transmission System Operators for Electricity
FCR	Frequency Containment Reserve
FD	Fast Decoupled
FDXB	Fast Decoupled
FRR	Frequency Restoration Reserve
FRRa	Automatic Frequency Restoration Reserve
FRRm	Manuel Frequency Restoration Reserve
GA	Genetic Algorithm
GS	Gauss-Seidel
GS	Gauss-Seidel
IEEE	Institute Of Electrical and Electronics Engineers
IEEE	Institute Of Electrical and Electronics Engineers
LPAC	Linear-Programming AC
MEPSO	Multi-Objective Evolutionary Particle Swarm Optimization
MILP	Mixed-Integer Linear Programming
NERC	North American Electrical Reliability Council
NR	Newton-Raphson
NR	Newton-Raphson
NSGAII	Non-Dominant Sorting Genetic Algorithm
OPA	Ornl-Pserc-Alaska
OPF	Optimal Power Flow
PF	Power Flow
RRE	Risk And Reliability Engineering
TADS	Transmission Availability Data System
TSO	Transmission System Operator
UCTE	Union For the Co-Ordination of Transmission of Electricity
UFLS	Under-Frequency Load Shedding
UVLS	Under-Voltage Load Shedding
WECC	Western Electricity Coordinating Council

List of Figures

List of Tables

Summary

Cascades provides a platform for power system risk analyses and transmission system expansion planning. The objective of the Cascades module is to: (I) assess the ability of a power system to withstand sudden changes, i.e., component failures; and (II) to find contingencies critical to power system security; (III) to provide a risk-informed/based transmission system expansion planning. The Cascades module comprises of four models/methods: 1) A flow-based cascading failure simulation (CFS) model; 2) Transmission system expansion planning method; 3) Vulnerability analyses method; 4) Method for calibration of the CFS model.

1. Motivations

Power systems as a critical infrastructure are an integral part of human society and are therefore of paramount importance to modern life. They are one of the most complex engineered systems ever build, aiming to provide a reliable power supply [1]. Despite multiple layers of protective schemes, power systems are still experiencing blackouts that are mainly caused by cascading failures [2, 3]. Furthermore, an analyses of blackout data from Europe and Italy show that cascading failures are responsible for most of the large blackouts. Cascading failures in power systems are defined as the progressive failures of system components after an initial contingency.

In November 2006, a 380 kV line in northern Germany was manually disconnected due to a request from a shipyard [4]. The disconnection resulted in a violation of the N-1 security criteria in the E.ON Netz and in some of the tie-lines with the neighboring TSOs. In the same time, a 380 kV tie-line between E.ON Netz and RWE TSO was operating near its thermal limit. A small increase in the power flow in this line triggered an alarm, demanding an action from the operators. A rushed remedial action, without the power flow calculations, resulted in additional loading of the line. This in turn caused a line trip that promptly propagated in a cascade of line tripping, which split the Continental Europe synchronous system into three areas. At the time of splitting there were significant East-West power flows because of international power trade and the obligatory exchange of wind feed-in inside Germany. The grid split caused large frequency deviations in all the three areas. In the West area the under frequency (49 Hz) was resolved with large reserves deployment and load shedding. In total 17 000 MW of consumption and 1 600 MW of pumps where shed. In the North-East the over-frequency (51.4 Hz) was resolved with automatic actions, including frequency control and automatic tripping of the units sensitive to high frequency. In total, 6200 MW of wind generation were disconnected from the grid. In the South-East area, there were no significant frequency deviations, thus no load shedding occurred.

Similarly, in January 2021, disconnection of a busbar coupler at a substation in Croatia, triggered by overcurrent protection, initiated a cascading failure event, which split the Continental Europe synchronous system into two areas [5]. Before the event, the western countries in the system ware importing large amounts of power (5800 MW) from the Balkans. Therefore, when the grid splitting happened, frequency deviations occurred in both areas, i.e., under-frequency (49.7 Hz) in the North-West area and over-frequency (50.6 Hz) in the South-East area. Load shedding, i.e., disconnection of industrial loads as regulated by dedicated national contract agreements were performed in France (1300 MW) and in Italy (400 MW) to stabilize the frequency in the North-West area. In total, 5.2 GW of generating units were disconnected to stabilize the frequency in the South-East area. Just months after this event, on July 24, a forest fire caused the disconnection of a tie-line between France and Spain, which propagated in to a cascading event resulting in grid split that separated Iberian peninsula from the rest of Europe [6]. The under-frequency (48.68 Hz) that occurred in the Spain and Portugal was resolved by frequency control actions (automatic and manual) as well as load shedding (4872 MW loads and 2302 MW of pumps). Furthermore, the frequency deviations caused the disconnection of 2673 MW of generating units.

In continental Europe, we have also observed large country-level blackout events interrupting the power supply to millions of customers. In September 2003, a line trip in Switzerland caused by a tree flashover resulted in cascade of failures that eventually progressed into separation of Italian power system from the rest of continental Europe [7]. The splitting has resulted in loss of 6400 MW power imports to Italy. Consequently, the system experienced large frequency drop, which caused the disconnection of multiple generators, causing even higher frequency drop. Eventually, what started as a cascading event resulted in total blackout of the Italian power grid. In March 2015, an overload of a 400 kV power line unfolded into a cascade of line overloads in the Turkish power grid [8]. Consequently, the Eastern and Western Turkish subsystems were separated in two grids. After the separation from the Continental Europe synchronous system, due to large under-frequency the

Western subsystem collapsed in around 10 seconds. The Eastern subsystem, with significant over-generation (4700 MW) also collapsed in few seconds despite the disconnection of several generators.

In all the events, we observe that the grid splitting and the interruption of supply to costumers are a result of component failures that unfold into cascading events. We have collected over 500 events where costumers have been affected because of failures in the extra-high and high voltage grids in Europe. This data shows that although events with country-level blackouts are rear, there are significant number of events that have caused the interruption of the power supply to consumers. The blackout events are initiated by a single or multiple failures of transmission grid components and power plants. Figure 1 shows the rate of failure of power grid assets and their impact on demand not served (DNS) in the Italian power system. We observe that transmission lines and switchyard components are the dominating failures with the highest impact to the system security.

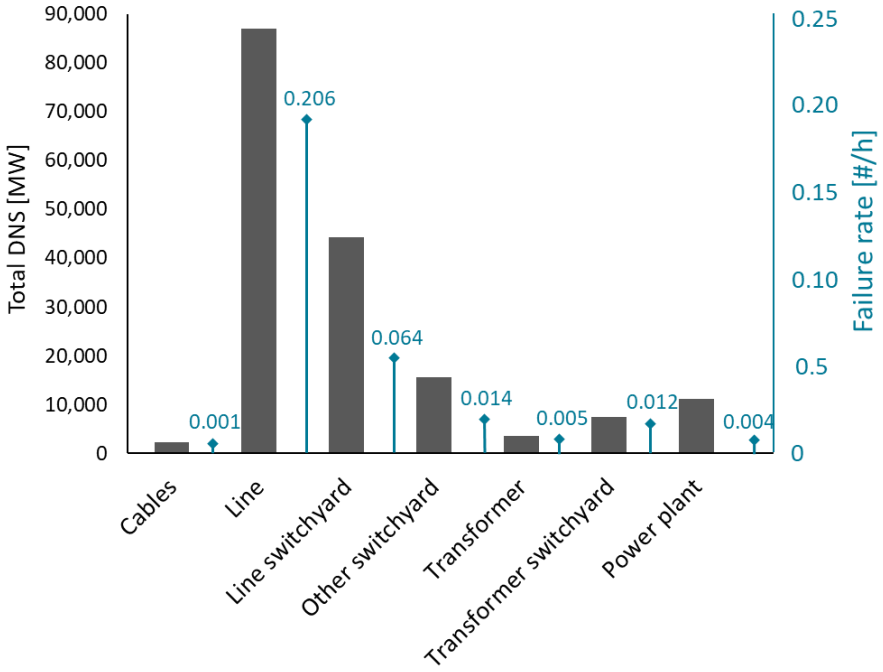


Figure 1. The rate of failure of power grid assets and their impact on demand not served (DNS) in the Italian power system.

2. Introduction

Ensuring the safe and stable operation of electrical power grids is imperative for the normal functioning of modern society. Due to the ever-increasing demand and adoption rate of intermittent energy sources, modern systems often operate near critical loading conditions, putting into question their capability to prevent and mitigate cascading failures [2, 9]. Therefore, understanding the complex nature of these failures can greatly aid in assessing the security of large interconnected power systems and identify vulnerabilities.

A power system is considered N-1 secure if a single loss of an element does not cause operational violations [10]. N-1 contingency analyses is the traditional method for assessing system security [11, 12]. A recent NERC Transmission Planning standard TPL-001-1 requires more rigorous contingency analyses [13]. These requirements include analyses of loss of two elements simultaneously (N-2) and loss of two elements consecutively (N-1-1). More comprehensive methods for power system security analyses are needed to comply with these requirements [14].

The cascading mechanisms in power systems have been extensively researched [15] and numerous models have been developed to analyze these occurrences. A list of the most prominent of these models is presented in [16], which sorts them based on their properties, approach and simulation scenarios. One of the most common type of models are flow-based models, which consider the physical properties of the system. Dobson et al. [17] presented one of the earlier models of this type, using DC optimal power flow (OPF) calculations to redistribute the power flows in the IEEE 73-bus test system after each stage of a cascading event. The DC-based ORNL-Pserc-Alaska (OPA) [18], a staple flow-based model, analyzes the connection between cascade progression and systems operating in near-critical loading conditions. The hidden failure model [19] investigates the impact of hidden failures on large cascading events. Recent works using DC flow-based methods identify critical risks of cascading failures using a dynamic optimization problem presented in [20] and investigate the safety implications of high wind energy integration using a graph traversal algorithm [21].

Although computationally efficient, DC-based models disregard the transmission losses and reactive power in the system, thereby introducing a notable calculation error. To overcome this, more accurate and computationally expensive AC OPF models such as the Manchester model [22], have been introduced. This model investigates the impact of critical loading conditions on the size of blackouts, concluding that blackouts follow a power law probability distribution. Additionally, the model tests the system response by introducing random outages and uses predetermined sequences for restoring load-generation balance. The model presented in [23] expands on this idea by also implementing additional protection mechanisms, such as corrective under-voltage load shedding, while [24] determines the available transfer capacity between control zones with regards to cascading failure vulnerability. To bridge the gap between AC and DC methods, linear implicit AC PF cascading failure models [25] have been developed. This model greatly increases the accuracy compared to a standard DC PF, by minimally increasing the computation time. Another application for this approach is presented in [26], where a controlled islanding strategy for severe cascading events is developed using a mixed-integer linear programming (MILP) model. A multi-step corrective control method for nodal power injections, which can assist in mitigating the propagation of cascading failures, is presented in [27]. Similarly, [28] introduces a wind power plant re-dispatch model, which optimizes wind generation and provides load-frequency control, thus mitigating system overloads. A vulnerability analysis of power systems to cascading failures using a modified H-index method is presented in [29], which is derived from the method used to quantify scientific achievements, and indicates that the degree of influence of a node on its adjacent nodes is not less than H. The proposed method can effectively identify lines which are prone to propagate faults using a two-level cascading transition graph, where the first level investigates the direct impact of a line failure on the system, while the second level investigates the impact of hidden failures. Preventive strategies to mitigate extreme-weather-related contingencies are investigated in [30]. The model introduces a proactive operational strategy intended

to enhance the system resilience and mitigate cascade progression during weather-related events. A resilience-oriented transmission expansion planning model is presented in [31], where a multi-stage optimization problem is implemented which considers investment decisions, N-1 security criterion and resilience requirements against cascading failures. With the emergence of cyber-physical power systems, researchers have investigated the impact of the cyber network on the progression of cascading failures. A framework that analyses cyber-coupled power grids is presented in [32], where the physical network, as well as the control and protection functions of a coupled cyber network are modeled. The publication concludes that grid robustness can be improved by implementing a cyber network, however, deliberate cyber-attacks can drastically increase the magnitude of the cascades. A comparison of the performance of different flow-based models is presented in [33], while [34] and [35] also bring additional academia and commercial cascading failure models into the discussion.

The Risk and Reliability Engineering (RRE) at ETH Zurich lab has developed the Cascades, which is a framework for power system risk analyses and transmission expansion planning. In the core of Cascades is a model for cascading failure simulations (CFS). This is a flow-based model that that can ply DC or AC power flows. Furthermore, the CFS model in Cascades can be used for single-zone and multi-zonal power system analyses. The latter is a feature that lacking in the state-of-the art models for cascading failure analyses. Cascades also encompasses methods for power system vulnerability analyses and transmission system expansion planning. In addition, we have developed a method for to calibrate the CFS model against real blackout data.

3. Flow-based model for cascading failure simulations

A power system consists of one or more control zones. Each zone is controlled by a transmission system operator (TSO), which is responsible for operating, maintaining, and expanding the transmission grid. A single-zone CFS model can successfully capture the development of a cascade within one control zone. However, such regional approach cannot capture the operations of multi-zonal interconnected systems. In practice, these systems consist of multiple neighboring TSOs, which have the mandate to stabilize their respective control zones on a national scale, while minimizing the impact on the remainder of the system [36, 37]. This section describes the Cascades CFS model for single-zone and multi-zonal analyses.

In general, the CFS model is characterized by the following functions:

- Simulates critical scenarios that may trigger a cascading event,
- Identifies island operation within the system,
- Deploys generation reserves to restore the power balance in the system in case of a generation/consumption mismatch or island formation
- Triggers load shedding due to system frequency deviations beyond a safety threshold/ bus voltages magnitudes falling below a tolerable limit
- Identifies blackout conditions in island operation resulting from large power imbalances that cause frequency instability
- Computes the load flow changes across the transmission grid using power flow methods
- Disconnection of overloaded power lines/transformers.
- Performs single-zone and multi-zonal CFS

3.1. Single-zone CFS model

The cascading failure analysis model procedure is described in the following 12 steps (see Figure 2):

1. *Initial system state*

   The initial system state is the starting point of the model. All the components in the system are in service and loads are equal to the forecast. There are no overloads, and the units generate power according to a given optimal power flow. In case the dispatch is provided by another module/algorithm and results in overloads, Cascades continues the simulations also considering these overloads.

2. *N-k contingencies*

   We introduce single of multiple simultaneous failures (i.e., contingencies) to simulate unexpected changes in the systems. The contingencies are comprised of 1) the transmission grid assets (power lines and transformers) and 2) power generators. We utilize a random failure mechanism to initiate failures of components based on their failure probability.

   For transmission components, the procedure starts by creating a vector of uniformly generated random numbers between 0 and 1. The size of the vector is equal to the number of lines and transformers in the system. This way we ensure that each value corresponds to a physical transmission asset in the system. If an element in the generated vectors has a value equal or smaller than the failure probability of the corresponding component, this component is considered failed and added to a contingency list. This procedure is repeated until the contingency list has reached a predefined size.

   For the generating units, the procedure starts by creating two vectors of uniformly generated random numbers between 0 and 1. The size of the vectors is equal to the number of generators in the system. The first vector is used to simulate a full loss of a generator. The second vector is used to simulate a derated state, i.e., a generator can be in a state that is only partially available.

For a generator that occurs to be in a derated state, the algorithm selects a new maximum available power. For this purpose, a sampling is performed from a discrete cumulative distribution function (CDF) of generator outputs in derated state.

The failure probabilities of transmission component, failure probabilities of generators, the probability of being in a derated state, and the available power in derated state (aforementioned CDF) are obtained using data from the European Network of Transmission System Operators for Electricity (ENTSO-E) transparency platform. We preprocess (cleaning, transformation) the raw data and employ Markov process to calculate the necessary unavailability parameter for generators (per type and size) for internal lines, interconnectors, and transformers.

The selection of contingencies is executed via Monte-Carlo sampling. Therefore, the larger the contingency list the higher the chance of exploring different sets of contingencies and, thus, obtaining a more comprehensive evaluation of the system risk. For each set on the contingency list a cascading failure simulation is run. After selecting a single set, the related component(s) are removed from the system and a new system topology is obtained.

3. *Island identification*

   The model checks the new system topology to identify if islands form, in which case the buses belonging to the respective islands are determined.

4. *Check frequency deviation*

   The frequency deviation is computed using the following equation [38]:

$$\Delta f = \frac{-\Delta P_{imbalance}}{\sum_{\Omega_G} 1/R_g + \sum_{\Omega_D} D_d} \quad (1)$$

   where $\Delta P_{imbalance}$ is the active power imbalance in MW, Ω_G and Ω_D represent the set of generators and demands, R_g and D_d are the frequency characteristics of gth generator (speed droop) and dth load, in Hz/MW and MW/Hz, respectively.

5. *Emergency automation – frequency*

   Emergency under-frequency load shedding (UFLS) is the last measure to be taken to maintain power system stability after a major disturbance. In principle, whenever the frequency deviation is in the allowed range (Δf>-0.2Hz), but the available capacities of the synchronized generating units are unable to satisfy the load, the UFLS automation uniformly disconnects enough load such that a new power balance is reached and a large-scale collapse of the system is avoided [38]. The UFLS scheme used in the model is based on the Swissgrid transmission code 2013 [39]. It consists of six steps: if the frequency deviation, Δf, is lower than -0.2Hz, the pumped-storage hydro plants operating in pumping regime are disconnected; beyond -1Hz the load is uniformly disconnected proportional to the frequency deviation in four steps; finally, if the frequency deviation is -2.5Hz or +1.5Hz, the frequency blackout conditions are attained, and all generators are disconnected. Additional details on UFLS implementation in the CFS model are given in Section 3.2.3.

6. *Restore load-generation balance*

   Equation (1) shows that the frequency deviation in the system is proportional to the power imbalance. If there is a variation of load demand or island operation in the system, the balance between the power generation and load consumption in the system/island may not be maintained. The power imbalance in the system/island is measured as:

$$\Delta P_{imbalance} = \sum P_G - \left(\sum P_D + \sum P_a\right) \quad (2)$$

   where $\sum P_G$ is the total generation, $\sum P_D$ is the total demand, and $\sum P_{a_i}$ are the loses in the system/island. Restoring the load-generation-balance and, thus, the frequency in the system is performed by frequency control actions. When a frequency deviation occurs, all the generators

connected to the system, or the island will ramp up/down to increase/decrease their power output to compensate for the power imbalance. Generators change the power delivery proportional to their ramp rates and are limited by their maximum/minimum operating limits. For the extreme case, that there is no load connected to an island, all the generators are shut down. In general, generating units reach their new operating points typically in tens of seconds for primary frequency control, few minutes for secondary frequency control, and beyond 15 minutes for tertiary frequency control [40].

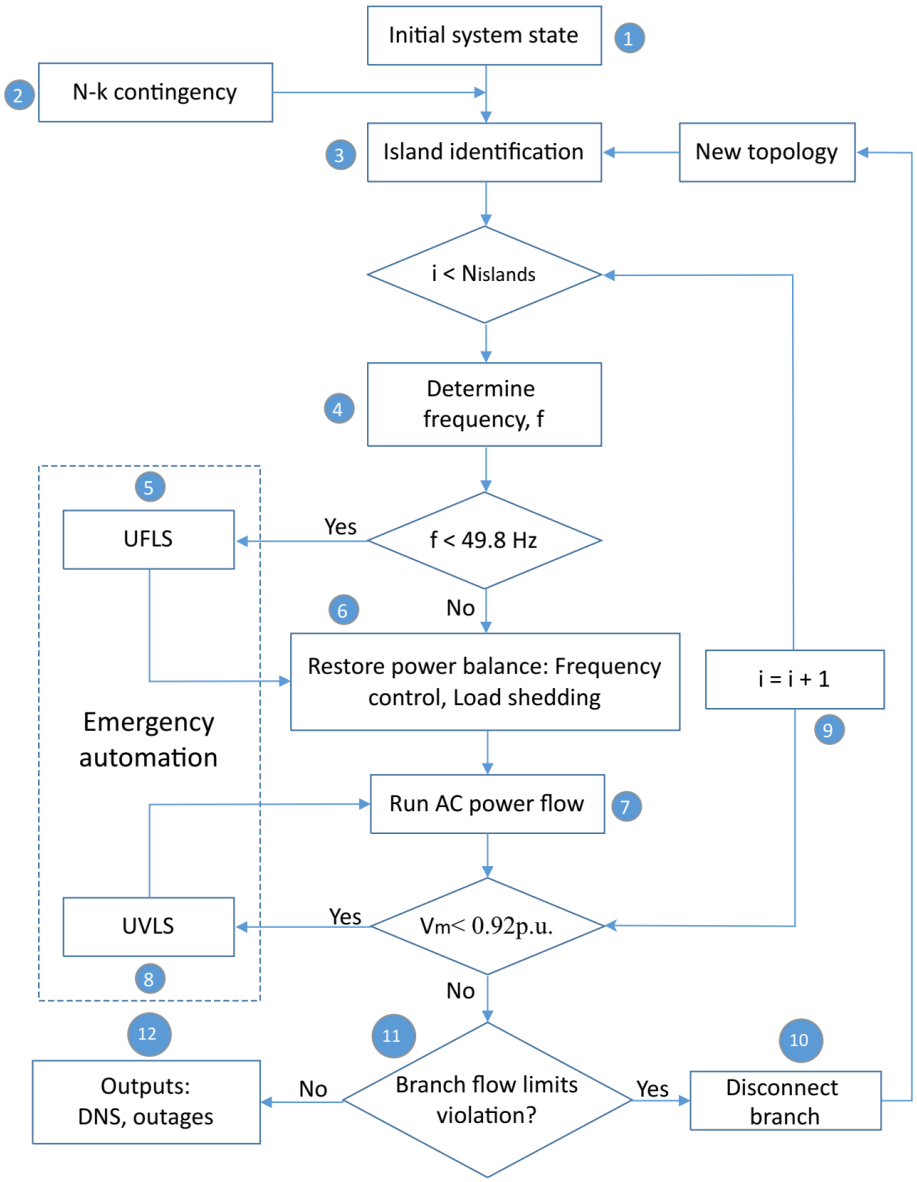


Figure 2. A simplified flow-chart of the cascading failure simulation model. UFLS stands for under-frequency load shedding; UVLS stands for under-voltage load shedding; Δf is the frequency deviation, which is a positive number when the load is smaller than the generation and negative number when the load is higher than the generation.

7. Power flow calculations

After restoring the load-generation-balance, a power flow (AC or DC) is computed using the Matpower package [41]. In case of AC power flow, a voltage stability check is performed, and the cascading failure simulation continues with step 8. In case of DC power flow, the cascading failure simulation continues directly with step 9, i.e., with an overload check of the transmission system branches.

8. *Emergency automation – voltage*

   A system may enter a state of voltage instability when it is subject to a disturbance. The steady-state threshold is a post-contingency voltage of 92% of the nominal voltage, or 0.92 p.u. [42]. One of the most economical remedial actions to voltage drop is under-voltage load shedding (UVLS) [43]. A stepwise UVLS procedure is applied to the load directly connected to the bus at which the voltage violation occurs. In each step, 25% of the initial load is shed, and the voltage magnitude is updated via the AC power flow algorithm. If the voltage violation persists, this procedure is repeated until the voltage magnitude is restored within tolerance limits. If under-voltage load shedding cannot restore the voltage, the entire load connected to the bus is lost.

9. *Island evaluation*

   Steps 4 to 8 are applied to all islands, updating the power flows and bus characteristics of each island.

10. *Branch flow limit violation*

    After the power flows in each island have been updated, the branches with power flows higher than their tripping threshold/maximum transmission capacity are identified. Only one line at a time is tripped, i.e., the line with the highest overload. The new system topology is returned to Step 3 for island identification.

11. *Termination criteria*

    The simulation is terminated when there are no branch flow violations.

12. *Output results*

    Besides the demand not served (DNS) and total number of lines/transformers tripped for each set of contingencies, the model records the impact of each line/transformer on the DNS, the frequency of line/transformer overload, the cascading stages and the corresponding DNS at each stage, and the utilization of each unit in positive and negative reserves. The quantified risk of DNS intrinsically compounds the contributions of potential frequency and voltage violations captured during the unfolding cascading failure events.

The above procedure is repeated for different loading conditions, sampled from a yearly load curve. The year is divided into four seasons and in each season, equal number of pick load and random load selections are made. For each loading condition the obtained cascading failure results are assembled, processed, and presented as an output of the model. One of the most important results of the cascading failure model is the complementary cumulative distribution function (CCDF) of the DNS. We refer to it as the cascading failure risk curve, or simply, risk curve. The risk curve gives the exceedance probability of observing DNS (blackout intensity), larger than the value reported on the x-axis. The main outputs of CFS model are summarized in the Table 1.

Table 1: List of CFS model output data.

Variable	**Resolution**	**Unit**	**Description**
DNS	-	-	A complementary cumulative distribution function, i.e., risk curve
Outage occurrence	-	-	Transmission system component outage occurrence
Cascading stages	-	MW	the lines/transformers tripped at each stage and the corresponding DNS
Importance	-	-	Importance/criticality of the transmission system components
Reserves utilization	-	MW	To what extend each unit is used to provide reserves

3.2. Multi-zonal CFS model

The multi-zonal model is an extension of the single zone approach. Both options are available to the user of Cascades. Figure 3 shows the flowchart of the Cascades algorithm for multi-zonal cascading failure simulations. The multi-zonal approach is detailed in [44].

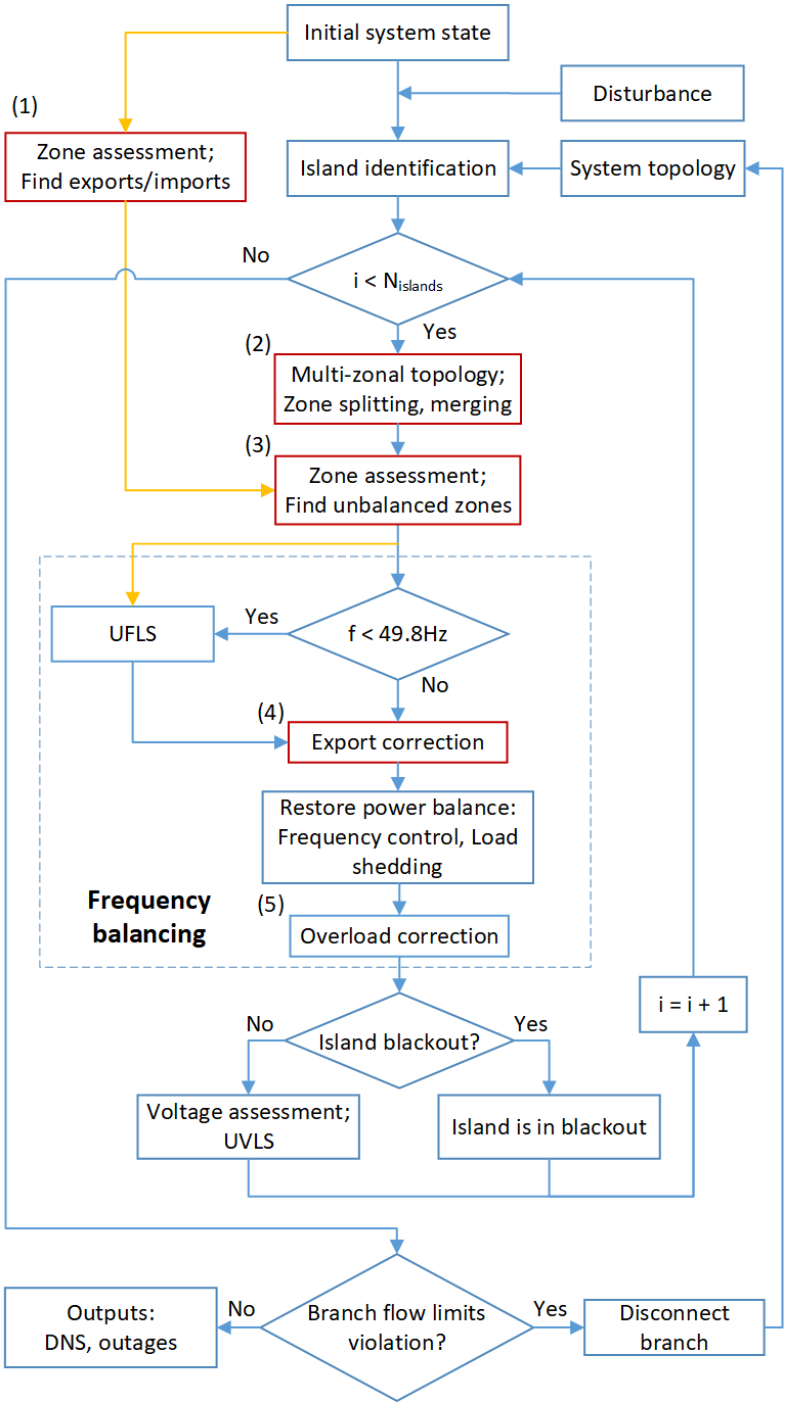


Figure 3. Algorithm of the Cascades framework. Red blocks are used exclusively for a multi-zonal simulation; blue blocks are used in both multi-zonal and single-zone simulation methods; blue lines represent the execution order of the algorithm; yellow lines represent information flow between blocks.

The model first establishes the initial state of the system, where the generating units are dispatched using an AC optimal power flow (OPF) solver [41] to match the demand in the system determined by a yearly load curve, while respecting the transmission line limits. At this stage, the system is stable, and no overloads are observed. The control zones in the initial state are identified and the power exports/imports between zones are calculated. Subsequently, disturbances (contingencies) are introduced in the system, which simulate transmission line/transformer failures, via Monte-Carlo simulation based on their failure probabilities. The topology of the system is examined in order to identify new islands and assess their stability. Multi-zonal islands are further assessed for the occurrence of splitting or merging of the control zones and their topology is updated accordingly. Following this, the power flow between the control zones in the island is calculated and unbalanced zones are subjected to corrective measures (e.g., under-frequency load shedding - UFLS), if the frequency deviation in the island is larger than $\pm$200 mHz. After bringing the frequency to acceptable levels, the reserves of every zone are reassessed. If an exporting zone experiences a demand increase which exceeds its reserves, a reduction of the exports is performed until the zone is stabilized, and the imbalance is transferred to the importing zones. The power flows are calculated, and unbalanced zones are identified for frequency control, until the island frequency is brought to nominal value. After stabilizing the island, the transmission lines are analyzed for overloads stemming from the actions of the frequency control. If that is the case, overload correction is performed on the control zones that contributed to the overloads. For islands in which blackout conditions have occurred, the demand and power flows are set to zero. For the remaining islands, the final stability evaluation is performed by executing an AC power flow analysis and identifying buses with voltages below the threshold of 0.92 pu. In these buses, a stepwise under-voltage load-shedding (UVLS) is performed, shedding 25% of the demand with each step until the voltage level is restored or all demand is disconnected. After analyzing the islands, all lines in the system are examined and the exceedance of the maximum transfer capacity is identified. Overloaded lines are disconnected one at a time and the topology of the system is reassessed. The island identification loop is executed again until there are no overloaded lines, or in the worst-case scenario, the system has collapsed.

The term “island” in the model refers to the whole system as well as small, isolated clusters of buses. The procedures developed for the multi-zonal approach, enumerated from 1 to 5 in Figure 3 are organized into three functional groups, namely, control zone assessment, system topology update, and frequency balancing.

3.2.1. Control zone assessment

The frequency deviations in a synchronous area are first tackled by the frequency containment (primary) reserves, which are shared by all operating generators, and therefore, are simultaneously activated in all control zones. Following the primary reserves, ENTSO-E specifies that the unbalanced TSOs should apply corrective measures (secondary and tertiary reserves, and load shedding) in their control zones [36, 45], while minimizing the impact on the rest of the system. ENTSO-E also specifies that neighboring zones can assist the unbalanced TSOs, if their safety is not compromised. However, this option will not be considered in the model, as it is usually discussed on a bilateral level and is not mandatory. Another deviation from ENTSO-E specifications is introduced regarding the usage of reserves. If the reserves in the system are pre-determined, they are differentiated in the model as automatic (primary + secondary) and manual (tertiary) reserves. If the reserves in the system are not pre-determined, they are calculated for every dispatchable generator as the difference between the maximal capacity and current generation levels.

The procedures for assessing the control zones are:

- Control zone identification - identifies all buses and branches belonging to the individual control zones in the system. Their connections with neighboring zones are determined by identifying all tie lines, i.e., branches which connect two buses from different zones.

- Power flow assessment – calculates transmission losses in the system, assigns them to control zones and analyzes the power flow between the control zones. Transmission losses are first calculated for each control zone individually, by considering every zone to be an isolated island, followed by the transmission losses over the tie lines. Transmission losses over tie lines are calculated by subtracting the power flows in the starting and ending nodes of the respective line and are always assigned to the importing zone. With all losses known, the power deviation in each control zone is calculated as:

$$\Delta P_i = \sum P_{G_i} - \left(\sum P_{D_i} + \sum P_{a_i} + \sum P_{L_i}\right), \qquad i = 1, 2, \dots, n \tag{3}$$

where n is the number of control zones in the island, while for every individual zone i, $\sum P_{G_i}$ is the total generation, $\sum P_{D_i}$ is the total demand, $\sum P_{a_i}$ are the transmission losses within the zone, $\sum P_{L_i}$ are transmission losses over tie lines importing power, and ΔP_i is the power deviation.

The following procedure is used for identifying unbalanced zones in the system. Initially, the power flows in the system are determined by calculating the power deviation of each control zone before a contingency. Zones which export power will have a positive deviation (generation larger than demand in the zone), while importing zones will have a negative deviation (demand larger than generation in the zone). As specified by ENTSO-E, power flows across zones should be maintained to prevent system destabilization when a TSO suffers imbalance. Following a contingency, the new state of the system is determined by updating the power deviations in each control zone. The difference between the post-contingency and initial power deviations represents the real power imbalance that should be corrected by the affected TSO:

$$\Delta P_i = \Delta P_i^{PC} - \Delta P_i^{Init}, \qquad i = 1, 2, \dots, n \tag{4}$$

where ΔP_i^{Init} is the initial power deviation representing the exports/imports of control zone i, ΔP_i^{PC} is the post-contingency power deviation of the same zone, and ΔP_i is the real power deviation that needs to be corrected by the TSO of zone i.

3.2.2. System topology update

Updating the post-contingency topology is essential for cascading failure analysis because disconnected lines form islands and remove loads and generators from the grid resulting in power imbalance. Islands must be individually stabilized, as they have no connection to the grid. The Cascades model employs an island extraction procedure [41], which can accurately identify all buses belonging to islands. However, in multi-zonal islands complex scenarios can occur that drastically affect the topology, such as zone merging or splitting. The following procedures serve to correctly update the topology of the system and do not represent optimal splitting strategies that can be employed to stop the progression of cascades, which can depend on the practice of the particular TSO.

3.2.2.1. Control zone merging

This procedure analyzes multi-zonal islands and performs zone merging if predefined conditions are met. Contingencies can isolate small islands from their native control zone, which remain connected to large neighboring zones via tie lines (Figure 4a). In most cases, the large control zone and the small island connected to it will coordinate balancing measures to ensure stability, with the larger zone effectively absorbing the much smaller island. This occurrence was observed during the French-Spanish split in 2021 [46], and the South Australian islanding of 31st of January, 2020 [47].

Such scenarios are simulated in the model by assessing if control zone pairs exist in the observed island for which merging is possible, i.e., there are tie lines that connect them (Figure 4b). After the potential pairs are identified, the number of buses within each zone are compared. If one zone is smaller than a selected threshold (10% in our analysis) compared to the other, it will be merged into the larger zone. Subsequently, the topology of the island is updated in which the merged zone ceases to exist, and its buses are delegated to the absorbing zone (Figure 4c).

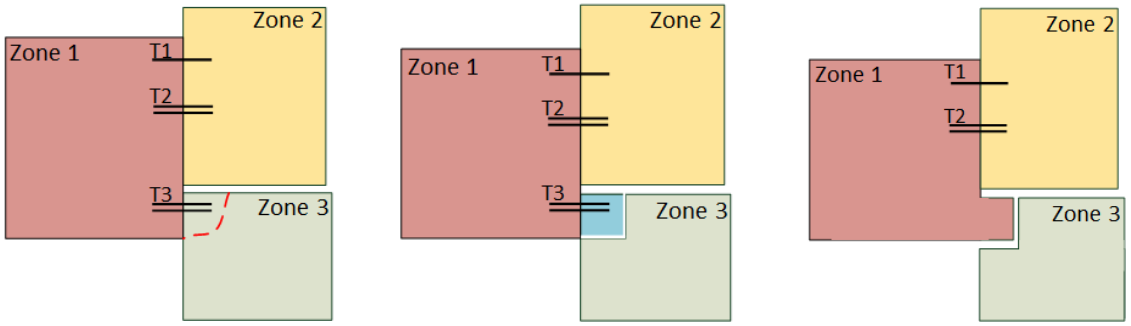


Figure 4. Control zone merging scenario in a three-zonal system simulated by the Cascades model. Lines labeled with T1-3 represent tie lines. a) cascading failure occurs in control zone 3, isolating part of the zone; b) merging conditions are evaluated as the isolated island (marked in blue) remains connected to zone 1; c) the island is merged with control zone 1, zone 3 is isolated from the system as there are no remaining tie lines.

3.2.2.2. Control zone splitting

This procedure takes place when control zones split in two or more parts. This scenario is in large part inspired by the recent splitting of the European synchronous area on 8th of January, 2021 [48]. To assess the power imbalance after such disturbance, the islands are analyzed based on their size and number of control zones:

- Single zone islands – islands which do not have active trade with other control zones, therefore the power mismatch between generation and demand is equal to the post-contingency deviation:

$$\Delta P^{Init} = 0 \rightarrow \Delta P = \Delta P^{PC} \tag{5}$$

- Multi-zonal islands – power imbalance is determined based on the size of the island. The parent island from which the new island separated after the contingency is identified by comparing the buses belonging to the new island with the buses of all pre-contingency islands.

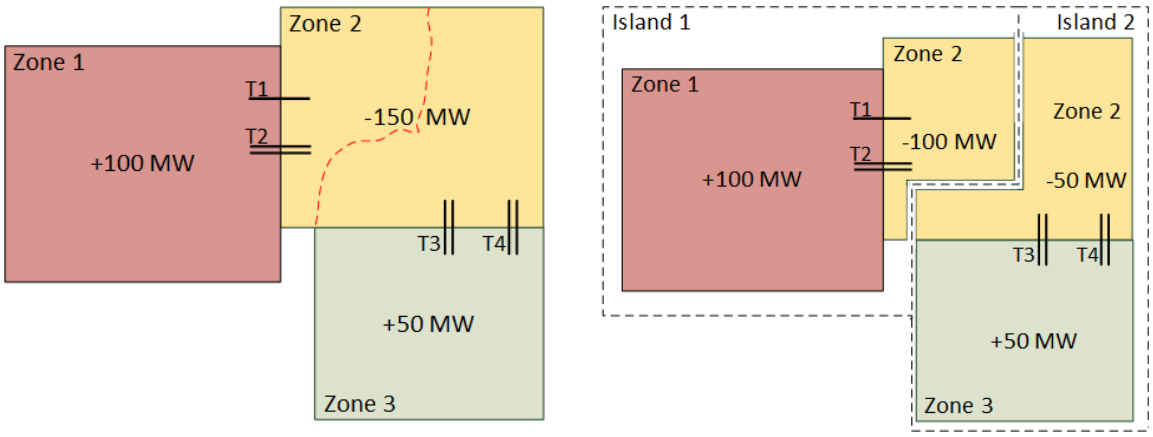
In multi-zonal islands which are smaller than a certain threshold (10% in our analysis) compared to their parent island, the island is considered to be too small to maintain active trade between the control zones, and every power imbalance is corrected within each individual control zone:

$$\Delta {P_i}^{Init} = 0 \rightarrow \Delta P_i = \Delta {P_i}^{PC}, \qquad i = 1, 2, \dots, n \tag{6}$$

For islands that are larger than the set threshold, the agreed power exports and imports between the control zones should be respected. This is depicted in Figure 5, where a cascading event in a three-zonal system results in a zone split. The topologies of the observed islands (Figure 5b) are compared to their parent system (Figure 5a) to identify which tie lines remain connected after the contingency. Subsequently, the pre-contingency power flows over the remaining tie lines are analyzed for every zone to find the original exports and imports. These power flows are maintained after the contingency, otherwise the fault from control zone 2 propagates to zones 1 and 3. Finally, the real power imbalance in every control zone, also accounting for the split zones, are calculated as the difference between the post-contingency power deviation and the pre-contingency exports/imports:

$$\Delta {P_i}^{Init} = \sum_{\substack{j=1 \\ j \neq i}}^{n} {P_{ij}}^{tie} \rightarrow \Delta P_i = \Delta {P_i}^{PC} - \Delta {P_i}^{Init}, \qquad i = 1, 2, \dots, n \tag{7}$$

where ${P_{ij}}^{tie}$ is the sum of power transferred over tie lines between control zones i and j, ΔP_i is the real power deviation of zone i.



a) b)

Figure 5. Control zone splitting scenario in a three-zonal system simulated by the Cascades model. Lines labeled with T1 – 4 represent tie lines. a) Cascading failure occurs in control zone 2, which imports power from zones 1 and 3; b) two new islands are created after zone 2 splits. The model recalculates the power exports/imports in each island that need to be maintained.

3.2.3. Frequency balancing measures

Ensuring a stable frequency close to nominal value (50Hz in the UCTE) is vital for the integrity of the grid. As the frequency deviation is proportional to the power imbalance in the system, unbalanced control zones have to be stabilized, while minimizing their effect on the grid [36, 45]. Using the previously discussed methods that can identify the unbalanced zones, several procedures are employed to enact balancing measures in a multi-zonal system.

3.2.3.1. Under-frequency load shedding (UFLS) and frequency control

UFLS is the final resort when the frequency deviation exceeds a specified range. The activation setpoints in this scheme are based on the ENTSO-E UFLS policy [45], while the load shedding steps are based on Swissgrid transmission code 2013 [39] as described in Step 5 (Single-zone approach). ENTSO-E specifies emergency actions for frequency deviations $\Delta f < -0.2Hz$ from nominal, therefore the measures are taken in the following order:

- $-1\ Hz \leq \Delta f < -0.2\ Hz \rightarrow$ pumped-storage hydro plants operating in pumping regime are disconnected in the unbalanced control zones. This is performed in a stepwise manner by shutting down the pumps one by one until the frequency deviation is brought within acceptable levels. ENTSO-E specifies that all pumps should be automatically disconnected if the frequency falls below 49.2 Hz, while in our model that threshold is set at 49 Hz.
- $-2.5\ Hz \leq \Delta f < -1\ Hz \rightarrow$ load shedding is uniformly performed in the unbalanced control zones in four steps, proportional to the frequency deviation and shedding up to 85% of the demand.
- $\Delta f < -2.5\ Hz$ or $\Delta f > 1.5\ Hz \rightarrow$ generators in all control zones are disconnected and the system effectively enters a blackout.

The frequency deviation in the system is computed using the Eq. (1). To identify the unbalanced control zones so they can be subjected to UFLS and limit the progression of the disturbance, the contribution of every zone in the frequency deviation of the system is calculated using the following approximation:

$$C_i = \frac{\Delta P_i}{\sum_{j=1}^{n} \Delta P_j} \rightarrow \Delta f_i = \Delta f \cdot C_i\,, \qquad i = 1, 2, \dots, n \tag{8}$$

where C_i is the contribution of control zone i to the total power imbalance in the island and Δf_i is the approximate frequency deviation originating from that zone. The control zones that are responsible for the frequency deviation are subjected to UFLS.

If UFLS is successful, frequency control is used to restore the frequency to nominal. This measure is enforced on all islands whose frequency deviation is higher than ±10mHz. The power deviation in the unbalanced zones is corrected by employing the generation reserves, performed by ramping up/down the generators proportional to their ramp rates until they match the demand, while conserving the exports and imports to adjacent zones. When the deviation is smaller than ±200 mHz, if the reserves in the control zone are pre-determined, automatic reserves (primary + secondary) are dispatched first, and, if they fail to match the demand, manual reserves (tertiary) are used. For deviations larger than ±200 mHz, ENTSO-E [45] specifies that TSOs can use all available resources to stabilize the zone, therefore, automatic and manual reserves are used simultaneously. This type of dispatch is also used if the reserves in the system are not pre-determined. If the available reserves are not sufficient to match the demand in the control zone, export correction and/or uniform load shedding is performed until power equilibrium is established.

3.2.3.2. Export correction

If the disturbance disables large generating units in the control zone, the demand might exceed the generation by a margin larger than the available reserves. To avoid potential destabilization of control zones that export power in normal conditions, a correction of the export to the adjacent zones is performed after the disturbance. This procedure simulates both the physical properties of the control zone which will not have enough generation to fulfil its exporting contracts, as well as the decision making of the TSO. The procedure first employs all remaining generation reserves in the zone, followed by calculating the surplus demand:

$$\Delta P_{S_i} = \sum P_{D_i} - \left(\sum P_{G_i} + \sum P_{R_i}\right), \qquad i = 1, 2, \dots, n \tag{9}$$

where ΔP_{S_i} is the surplus demand in zone i which needs to be corrected, $\sum P_{G_i}$ is the pre-contingency generation and $\sum P_{R_i}$ are the total generation reserves in the zone.

Subsequently, the zones that import power from zone i are identified and their import ratio is determined. The surplus demand from zone i is then transferred proportionally to the importing zones j, based on their import ratio and added to their power imbalance:

$$R_{ij} = \frac{I_{ij}}{\sum_{\substack{k=1 \\ k \neq i}}^{n} I_{ik}} \quad \rightarrow \quad \Delta P_j{}' = \Delta P_j + R_{ij} \cdot \Delta P_{S_i}, \qquad j = 1, 2, \dots, n \text{ and } j \neq i \tag{10}$$

where I_{ij} is the power imported to zone j from zone i, R_{ij} is the import ratio of zone j to the total imports from zone i, and $\Delta P_j{}'$ is the new power imbalance in zone j. The updated power imbalance in the importing zones is corrected using frequency balancing measures. If the reserves are not sufficient to match the increased demand in the zone, uniform load shedding is performed across the zone until power equilibrium is established.

3.2.3.3. Overload correction

TSOs should take into account the grid constraints when applying frequency control actions as specified by ENTSO-E [45]. Therefore, the experience and expertise of the operating staff is crucial in identifying potential congestion points and manipulating the dispatch of reserves accordingly. This decision-making process is implemented in the Cascades model via an overload correction procedure to mitigate line congestion.

The frequency control ramps the dispatchable generators proportional to their ramp rate across the affected control zone. Even though this method will effectively restore the nominal frequency in the system, it carries the inherent possibility that line overloads can occur. Therefore, a method is implemented which consists of identifying the problematic zones and correcting the dispatched generating reserves. The identification procedure finds the overloaded lines immediately after a contingency, then scans for new overloaded lines after frequency control is applied to every unbalanced zone. Zones that caused new overloads in the system undergo correction. To simulate an

idealized decision-making process by the TSO, the correction is implemented via AC OPF by adding the following constraints:

- The minimal values of the generators in the problematic control zones are fixed to their generating levels before applying frequency control.
- Their maximal values are set to match the available reserves for each generator. If the frequency deviation is below 200 mHz, the algorithm can use the automatic or both automatic and manual reserves. For deviations higher than 200 mHz, all reserves are used, as well as for systems where the reserves are not pre-determined.
- The generators in non-problematic control zones are fixed to the levels after frequency control is applied and are not affected by the correction.

Therefore, the correction exclusively alters the generator levels in the problematic zones, while accepting the generators set points identified by the frequency control for the remaining zones. If the AC OPF fails to converge, uniform load shedding is performed in the problematic zones in five steps (20% of demand per step) until convergence. If the solver still fails to converge and all demand is shed, the overload correction method is discarded and the original solution with applied frequency control is taken as preferred, despite the introduction of new line overloads into the system.

3.3. Algorithm features

The models described within Section 3.1 and 3.2 are integrated into a single CFS model. The user selects the simulation option. If single-zone approach is selected, the algorithm uses all the reserves available in the system, although the system may comprise multiple zones (TSO's). If multi-zonal approach is selected, the algorithm uses the reserves scheduled by each TSO to regulate the frequency deviations in its respective zone. In addition, we employ a third, hybrid, option that allows the TSOs to share reserves with a TSO that needs them. In fact, when a TSO is not able to resolve a power imbalance, the algorithm checks how much reserves are available in the neighboring TSOs. The available amount per neighboring TSO is limited by the internal needs for reserves because of simultaneous failures in these TSOs. Furthermore, the remaining reserves are constrained by the branch limits, i.e., the deployed reserves from the neighboring countries will not cause any line limit violations.

Cascades incorporates two cascading failure analysis models: the first one relies on DC power flow, while the second is AC power flow based. The default method for solving the AC power flow problem is the Newton-Raphson (NR) method. However, also the Gauss-Seidel (GS) and Fast Decoupled (FDXB and FDBX) methods are available. The main difference between the two models is that the DC-based model does not evaluate bus voltages and, therefore, UVLS cannot be performed. Furthermore, the AC model calculates the reactive power flows in the grid, which is not the case with the DC model. Consequently, it can be expected that the AC-based model will yield more load shedding events and requires more computing time. Both models use the MATPOWER package [41] for solving the optimal power flow (OPF) and the power flow (PF) problems that are part of the simulations. As OPF solvers the user can selected from MATPOWER's default MIPS solver, or MATLAB's FMINCON, the IPOPT which is made available through the OPTI Toolbox, or GUROBI for DC OPF.

The cascading failure model utilizes two methods for automatic frequency control: (1) the default method, and (2) the procured reserves-based method. When the default method is applied, every available unit is ramped up or down based on the current generation level, as described in Section 3.1 (Point 6). The procured reserves-based method only uses the reserves schedule provided by a market/dispatch model for frequency deviation lower than 200 mHz and the same procedure as the default method for frequency deviation higher than 200 mHz (see Section 3.2.3.1). Here, the primary and secondary reserves are combined and treated as one; this applies to both the positive and the negative reserves. The tertiary reserves are activated in case the primary and secondary reserves are not sufficient to balance the load demand. In both methods generators are ramped up according to

their individual ramp up rates. For the negative reserves, besides using the ramp down rates, an additional option is offered where the ramp down is performed proportionally to the current output of the units. The model records the utilization of the reserves from each unit, both positive and negative. This option can be used to perform a reserves procurement that improves system security, i.e., prioritize units that will be scheduled for reserves provision based on the utilization statistics obtained by the cascading failure simulations.

AC power flow convergence is not always guaranteed and there is a possibility that the algorithm will fail to obtain solutions under certain conditions. There are various reasons for non-convergence, from an inadequate initial guess of the bus voltages to voltage collapse. Before any measures are taken, we run simulations using the four available AC PF methods, i.e., NR, FDXB, FDBX, and the GS PF method. Each of these methods have different convergence properties, thus, the one showing the least convergence issues is selected. When a cascading failure investigation is carried out, it is likely that for some system configurations the AC power flow will not converge due to the system operating outside the steady state stability limits. This can be the result of a 1) reduced capacity of the system to transfer the power from one part to another and 2) lack of reactive power to supply load demand. Both can cause a voltage collapse in some parts of the system. We run a continuation PF to check for system stability limit violations. To mitigate the convergence problem, a built-in mechanism that can resolve convergence issues is implemented. The mechanism can act in four different ways and, depending on the order in which they are applied, four strategies exist. These actions are: (1) decrease system loading, i.e., decrease the load and generation simultaneously step-by-step until the AC PF converges; (2) perform full AC OPF; (3) accept the non-converged solution or consider system blackout; (4) perform reactive power AC OPF, i.e., fix the active power and run the AC OPF, thus perform reactive power re-dispatch. The preferred (default) order is (1)-(2)-(3), such that if (1) is not successful (2) is used and so forth. Current tests show that in the majority of non-convergence cases action (1) is sufficient; action (3) is only used in extremely rare situations. When action (3) is used, we accept the non-converged solution (the default option) and simulate how the cascade progresses. Current tests show that if none of the above actions is leading to convergence, the non-converged solution shows large active/reactive power flows in some parts of the system, resulting in overloads. This means that if this solution is accepted, it is very likely that the cascading will progress further, causing blackouts. However, the option exists that when action (3) is used a total system collapse is acknowledged, i.e., the algorithm sheds all load and disconnects all generation. The default value of the load/generator decrement in action (1) is 5%. Smaller values are not recommended because it may result in longer calculation times, while higher values may result in unnecessary load shedding.

In case the minimum and maximum operational constraints of the slack bus are violated in the AC PF solution, a correction is performed. Namely, the power output is 1) reduced to the maximum allowed value if the maximum power limit constraint is violated and 2) increased to the minimum allowed value if the minimum power limit is violated. The power difference is covered by the reserves or, alternatively, load shedding is performed. AC power flow computations can result in unrealistic reactive power flows that, in turn, cause unrealistic power losses. This issue is resolved by ignoring the losses and treating the power flow as DC. It is worth mentioning that these events are very rare and occur in some systems more often than in others. The algorithm also has the option to cancel this step.

The AC-based cascading failure model is adapted to receive generation schedules from a DC-based dispatch (obtained by a unit commitment / generation scheduling algorithm). The difference between the AC-based dispatch and the DC-based dispatch is in the power losses occurring in the transmission lines, which are neglected in the DC approximation. Thus, when a DC dispatch is fed into an AC model, there will a power loss mismatch that must be compensated by the slack bus, which may not be the most economical solution or may even be infeasible as the compensation would violate the maximum power limit constraint of the slack bus. Therefore, a procedure is created that performs a re-dispatch such that the losses are distributed to the cheapest generators with available capacity. In the process, all generators keep the value provided by the DC dispatch expect the ones that increase their power

output to compensate for the losses. In fact, the generators can only increase their output, while their minimum is temporarily fixed to the values of the given dispatch. This mechanism restricts complete re-schedule of this dispatch. The generators will increase the power output based on the available capacity such that, if they are scheduled to provide reserves, the reserves are not depleted. However, in case the total amount of additionally available generator power is smaller than the losses, the algorithm will use some of the scheduled reserves.

The failure probabilities of lines, transformers, and generators are calculated using large unavailability data from ENTSO-E and Italian TERNA. The algorithm has the option to use: 1) single failure probability for all branches and single failure probability for all generators, 2) failure probability for each branch type (i.e., internal line, interconnector, transformer), 3) separate failure probability for each branch in the system, and 4) failure probability per generator type and size. The branch contingencies are generated probabilistically as described in Section 3.1 (Point 2). By default, when multiple simultaneous failures are produced, the algorithm does not consider the geographical closeness of the failed branches. However, the algorithm provides the options that when two or more simultaneous failures occur, the failed branches are connected (i.e., geographically close). In addition, we can simulate hidden failures with a given probability. This means that, during the cascading simulations when a branch is tripped because of an overload, there is a likelihood that another branch (bellow maximum loading) is tripped in addition, i.e., hidden failure.

The algorithm also provides the option to use single branch rating coefficients (branch tripping thresholds) for all the branches or multiple ratings for different classes of branches. The purpose of these coefficients is for use as tuning parameters when model calibration is performed. When a single line rating coefficient is used a single value is set for all lines and transformers in the system. When multiple line rating coefficients are used, a unique value is assigned to each component based on the voltage level, i.e., the number of coefficients is equal to the number of unique voltage levels. A value of 1.3, corresponding to 130% the nominal line rating should be assigned as the default. The reason for the extra capacity is because the branches do not automatically trip when the nominal line rating is reached. In fact, a TSO is likely to allow some tolerance (often 30%) for a limited period. This tolerance gives the time to the TSO to reduce the branch loading. The line failure probability and line rating coefficients are decision variables in a validation and model calibration procedure we have developed [49]. The procedure is based on meta-heuristic with the objective of finding the optimal values for the line failure probability and rating, which will enable the simulation results to have the best possible agreement with historical data. We have performed this type of model calibration for the Western Electricity Coordinating Council (WECC) network, for which historical data on DNS and line outages is available [50, 51]. The entire calibration methodology is described in detail in Section 6.

3.4. Model limitations

Cascades is a steady-state model for power system security analysis. The following list provides context on some of the main limitations of Cascades:

- The stochastic behavior of the power systems is limited to the initial contingencies and hidden failures.
- Transients, e.g., generator swing equations, are neglected.
- Voltage control is not considered (under-voltage emergency protection is implemented).

3.5. Model validation

The cascading failure analysis model is validated on the reduced 240-bus WECC network [52]. The reduced 240-bus WECC system data include the hourly load profiles and RES generation for 2004. In total, 17 coal-fired and 4 nuclear power plants are modeled as base load units. The hourly set points of the 50 aggregated dispatchable gas fired generators are obtained via optimal power flow based on cost information. In the WECC network, the main source of blackouts data is the frequency distribution

of the blackouts size from 1984 to 2006 [53], which is provided by the North American Electrical Reliability Council (NERC). Additionally, the transmission line outage data of the Transmission Availability Data System (TADS) are recorded for 8864 outages by a WECC utility over the decade 1999-2008 [54]. The line outages are grouped into different cascade events [55]. Therefore, the probability distribution of the total number of outages in each cascade can be computed. The calibration of the CFS model (DC version) is performed using the Calibration methodology described in Section 6. The validation/calibration results are shown in Figure 6. The figure shows a good agreement between the historical data and the simulation results after calibration.

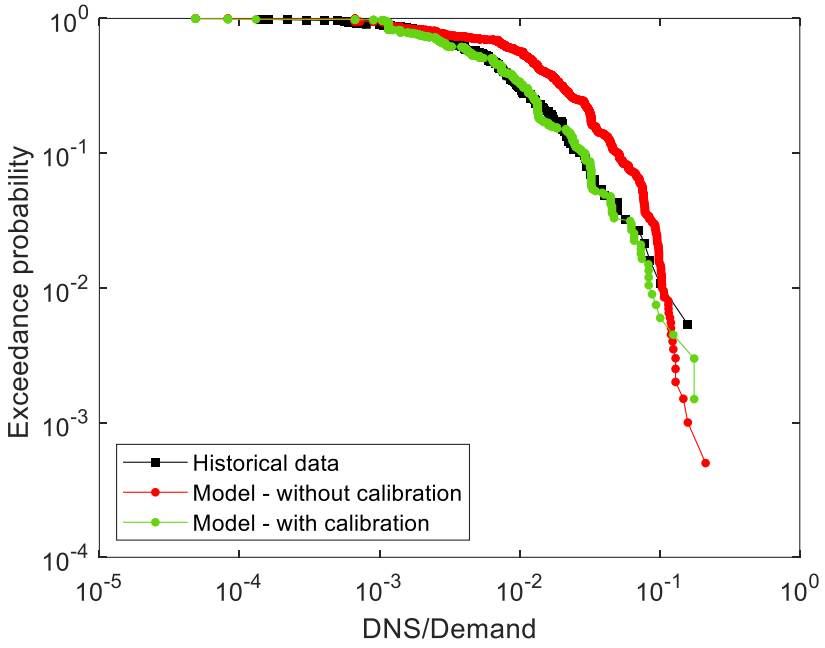


Figure 6. Risk curves for the 1) historical data, 2) model results, 3) model results with calibration.

4 Transmission system expansion planning method

Transmission system operators (TSOs) are constantly working towards strengthening the power grid by building new and increasing the capacity of the assets. To this aim, transmission expansion planning (TEP) identifies the most effective yet economical network upgrades. The decisions to install new infrastructure involves multiple factors such as investment costs and system security. The latter often implies the utilization of N-1 security criterion to ensure secure operations and identify weaknesses in the power grid. The latest NERC standards for transmission system planning require investigation of the performance during the simultaneous loss of two elements and an assessment of the risk of cascading outages [13]. Identifying TEP solution where the capital costs and the desired system security are the most favorable is, however, a challenging task and requires the multi-objective decision-making among alternative solutions [56].

TEP approaches can encompass deterministic, probabilistic, or epistemic-uncertainty conditions. Deterministic models are based on a set of specific parameters with predicted values, such as demand, generation, or market behavior, which are considered exactly known [57-60]. This approach usually provides a unique expansion solution. Probabilistic models acknowledge randomness in the parameter forecasts, thereby considering different scenarios. Probabilistic models can account for the uncertainties in the electricity demand level [61, 62], in solar power generation [63, 64], in wind power generation [65, 66], and forced outage rates of system components [63]. Epistemic uncertainties are also considered recognizing that part of the uncertainty originates from an inherent lack of knowledge. The probabilistic and uncertain models are often solved via stochastic [67] and robust [68] optimization. An extensive overview of available literature on TEP is provided in [69]. Another important factor affecting TEP is the model used in the power flow investigations, namely “AC” and “DC”. The DC power flow (PF) problem represents a linearized version of the AC PF equations [70] and presents computational advantages over AC PF [71]. However, DC simulations also come with an overall reduction of the result accuracy that can sometimes be considerable [72]. Before 2016, the vast majority of all TEP studies were carried out using the DC approach, while AC PF based studies have started gaining ground only recently [69]. Bent et al. [73] find that DC-based TEP solutions often underestimate the expansion costs and cause severe constraint violations when evaluated in an AC PF simulation. To overcome the shortfalls of DC-based approaches while avoiding the computational cost of AC models, the use of LPAC (linear-programming AC [74]) has been investigated for transmission planning. Contrary to DC, the LPAC model allows capturing both voltage magnitudes and reactive power and produce TEP solutions, which are mostly compliant to a regular AC model. In addition, the estimated cost of expansion is more realistic.

The TEP problem is tackled as single objective or multi-objective optimization problem using mathematical programming and metaheuristic techniques. Wu et al. [75] describe a linear programing (LP) representation of the TEP problem, which is solved as single objective optimization using the CPLEX Optimizer. Similarly, Zhang et al. [60] present a mixed-integer linear programming (MILP) formulation of the TEP problem based on the linearized power flow model, which is solved as single objective optimization using the Gurobi Optimizer. MILP is utilized by Zhang et al. [76] to perform combined expansion planning of transmission and natural gas systems considering N-1 contingency in both grids. The computational complexity is reduced by disjunctive model, which decreases the number of variables and constraints. Rahmani et al. [77] develop strategies to reduce the number of variables and the search space of a multistage TEP problem. The obtained results demonstrate the computational gain by the proposed technique. Furthermore, Braga et al. [78] use metaheuristics, i.e., a Simulate Annealing algorithm to solve the MILP representation of the TEP problem. The methodology uses the ε-constrained method and an interactive decision-making to explore the Pareto front. Gomes et al. [56] solve an multi-objective, non-linear, and non-convex representation of the TEP problem using a Multi-Objective Evolutionary Particle Swarm Optimization (MEPSO-II) algorithm. Even though MEPSO-II is more time-consuming, it improves the Pareto front and gives transmission network planners more insight compared to established algorithms.

TEP methods should integrate risk considerations to yield comprehensive solutions. However, risk has received a limited attention in TEP studies [69]. Qiu et al. [79] develop a multi-stage expansion planning model that combines both DC and AC network models and takes risk into account via a novel probability index. This index quantifies the expected load curtailment, allowing system planners to balance risk, reliability and cost by defining a "required capped load curtailment probability"-parameter as part of the TEP model input. Karimi et al. [80] present an expansion planning procedure that incorporates the risk of cascading line outages. It relies on a modified version of the DC PF based ORNL-Pserc-Alaska (OPA) model to compile a list of candidate lines for an expansion and compare different TEP scenarios derived from that list using cost-benefit-analysis. Zheng et al. [81] introduce a probabilistic multi-objective TEP model designed to enhance the ability of the transmission grid to protect against overloads when wind power is integrated. The model includes three risk-indicators, namely the non-overload probability of each branch, the non-overload probability of the entire power system, and the margin of non-overload probability of the system, based on which an overall risk index is determined. Qorbani et al. [31] present a single objective TEP method that uses DC based cascading failure simulation model to account for potential cascading events in the expansion.

The literature lacks a methodology that embraces AC PF based TEP and simultaneously minimizes capital costs and the risk of cascading outages. To address this research gap, we introduce a state-of-the-art risk-informed TEP tools that relies on an AC PF based cascading failure simulation model, which assesses the capabilities of a system to withstand outages. We have developed two methods for risk-informed transmission expansion planning (TEP): (1) optimization based TEP (Section 4.1); (2) importance list base TEP (Section 4.2). Both methods utilize the cascading failure model outlined in Section 3.

4.1. Optimization based transmission expansion planning method

The objective of the optimization based TEP method is to provide an adequate expansion plan with respect to the cost of expansion and risk of systemic failures. The minimization of the cost of expansion and risk of cascading outages are conflicting objectives, where the latter is calculated by simulating the cascading process. Therefore, the TEP problem is formulated as a simulation-based, multi-objective optimization problem that provides a set of Pareto optimal solutions, none of which outperforms the others with respect to all objectives. This set of solutions provides multiple expansion trade-offs to decision makers with respect to risk-improvement and expansion costs. Testing on the IEEE 118-bus system demonstrates the capability of the tool to identify multiple superior solutions with respect to literature benchmarks. The contributions of this method are twofold: 1) we integrate an AC-based cascading failure simulation model into a TEP formulation, thus expanding the current state-of-the-art; 2) the developed approach allows for system security assessment that goes beyond the N-1 security condition and can perform analyses on the progression of systemic failures into the grid. These assessments are aligned with the latest grid standards for transmission system planning [13]. The description of the TEP model presented here is detailed in [82].

4.1.1. Formulation of the TEP

In the TEP, we consider a list of branches with size B that consists of candidate lines and transformers. A candidate branch can be any existing branch or a new branch in a corridor where no branches exist. The decision variables in the TEP optimization problem are the statuses b_j of the candidate branches. The TEP optimization comprises two objective functions, i.e., the cost of TEP and the cascade risk improvement. The cost objective is defined as:

$$c = \sum_{j=1}^{B} b_j C_j , \quad j = 1,2, \dots, B \tag{12}$$

where c is the total cost of TEP, C_j is the cost of building branch j.

The risk improvement objective r is defined as the Euclidean distance [83] between the risk curve of the system before transmission expansion (reference system) and the risk curve of the system after transmission expansion:

$$r = \left(\sum |p\{d_R > x\} - p\{d > x\}|^2\right)^{\frac{1}{2}} \tag{13}$$

where $p\{d_R > x\}$ is the risk curve of the reference system, and $p\{d > x\}$ is the risk curve of the system with expansion, p denotes probability, d_R and d denote the DNS of the reference data and the DNS obtained with the simulations, respectively, and x denotes DNS value observed on the x-axis. The risk curves are obtained using the CFS model of Section 3. The goal of the optimization is to maximize r.

The multi-objective TEP problem is formulated as:

$$min\{c, -r\} \tag{14}$$

subject to:

$$\sum_{j=1}^{B} b_j \leq \overline{N} \tag{15}$$

$$b_i \leq \overline{N_b} \tag{16}$$

Where $\overline{N}$ is the maximum number of branches that can be built and $\overline{N_b}$ is the maximum number of times a branch can be selected for building.

4.1.2. Solution of the TEP

The presented optimization problem is tackled via simulation-based optimization problems. We use the genetic algorithm (GA) method [84], which has shown promising results compared to Nonlinear Mesh Adaptive Direct Search (NOMAD) [85, 86] and Particle Swarm Pattern Search [87] in cascading failure simulation studies [50]. Furthermore, GA is successfully applied on TEP problem as shown in [88]. In particular, we use the NSGAII [89] to deal with the multi-objective optimization TEP problem defined with Eq. (14). Figure 7 details the steps of the TEP procedure. The procedure starts by generating a population of random solutions, l, of the TEP problem:

$$l = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,j} & \cdots & b_{1,B} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,j} & \cdots & b_{2,B} \\ \vdots & & & \ddots & & \vdots \\ b_{i,1} & b_{i,2} & \cdots & b_{i,j} & \cdots & b_{i,B} \\ \vdots & & & \ddots & & \vdots \\ b_{L,1} & b_{\mathrm{L},2} & \cdots & b_{\mathrm{L},j} & \cdots & b_{L,B} \end{bmatrix} \tag{17}$$

where each row of the population represents a single TEP solution i $(i = 1,2, \dots, L)$ and L is the population size, i.e., the number of potential TEP solutions, and $b_{i,j}$ is a uniformly distributed integer number taking values between zero and $\overline{N_b}$ for each candidate branch j in the candidate list. The total number of new builds per TEP solution is constrained by Eq. (15). The cost of grid expansion, i.e., the value of the cost objective function is calculated for each potential solution from the population using Eq. (12). After upgrading/modifying the electrical grid according to each individual solutions i from the population, l, the cascading failure simulation model is run on each of the newly composed grids. These results are utilized to calculate the risk improvement objective function, r, using Eq. (13). Therefore, the fitness [89] of each solution from the population is calculated based on the cost of TEP and the risk improvement objectives, subject to the constraint satisfaction.

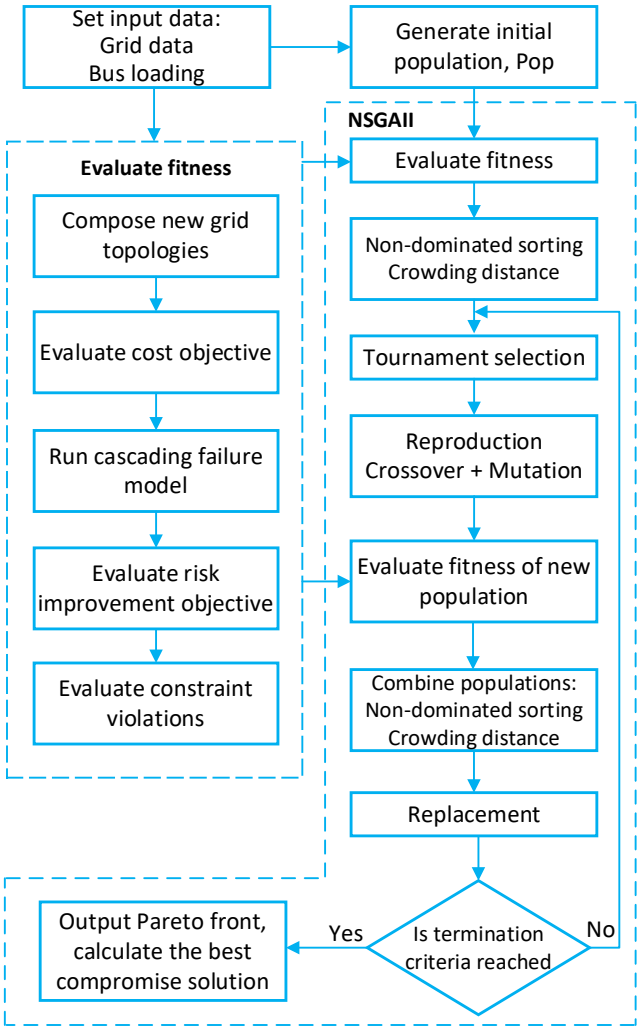


Figure 7**.** Flowchart of the TEP procedure.

The fitness of the solutions is used to sort the population by applying the concept of non-dominance [89]. If a solution outperforms all the other solutions with respect to all objective functions, this solution is non-dominated or non-inferior [90]. In this case, each solution is evaluated according to the cost objective (Eq. 12) and the risk improvement objective (Eq. 13). Then, the solutions are ranked based on their non-dominance and the population is sorted into Pareto frontiers. The solutions that satisfy the constraints always outrank those that do not satisfy the constraints. A solution is assigned to frontier $F + 1$, if it is dominated by at least one solution belonging to frontier F. The solutions belonging to the first frontier are called Pareto optimal solutions. Besides the non-dominant sorting procedure, the NSGAII uses the crowding distance operator to help maintain the diversity of the population. The crowding distance estimates the density of solutions surrounding a particular solution i by calculating the average distance between i and the two closest solutions [89]. Tournament selection is utilized to select the chromosomes (solutions) for reproduction. The selected chromosomes enter the reproduction step, integrating the blending crossover operator (BLX-α) and the non-uniform mutation operator [91]. In the next step, the non-dominated sorting and the crowding distance operators are applied to the new and the old populations. The replacement operator decides which solutions from the combined population will be passed to the next generation. Only half of them survive, a decision made according to the non-dominance rank and the crowding distance. Solutions with high rank and large crowding distance have higher survival chances [89]. The procedure is repeated until the stopping criteria, i.e., the predefined number of iterations is reached.

The multi-objective optimization of competing objectives results in a set of Pareto optimal solutions, none of which outperforms the others with respect to all objectives. Therefore, any of these solutions

can be selected based on the preferences of the decision maker. In this study, the best compromise solution (BCS), which satisfies both objectives to an acceptable degree, is selected via an approach that is based on the fuzzy set theory [92, 93]. The approach utilizes a membership function, which quantifies the degree of membership to a fuzzy set for an element from an input space by assigning it a value between 0 and 1 [94]. In particular, the objective functions of the Pareto optimal solutions are fuzzified using a trapezoidal membership function, and they are assigned values in the range [0, 1]:

$$\mu_k^m = \begin{cases} 1 & f_k(m) \leq f_k^{min} \\ \dfrac{f_k^{max} - f_k(m)}{f_k^{max} - f_k^{min}} & f_k^{min} < f_k(m) < f_k^{min} \\ 0 & f_k(m) \geq f_k^{min} \end{cases} \tag{18}$$

where $f_k(m)$ is the k^{th} objective function value for the m^{th} Pareto optimal solution, and $f_k^{\min}$and $f_k^{\max}$ are the minimum and maximum values of the k^{th} objective function across all optimal solutions, respectively [95]. The normalized membership function, μ^m, for the Pareto optimal solution m is calculated as

$$\mu^m = \frac{\sum_{k=1}^{K} \mu_k^m}{\sum_{m=1}^{M} \sum_{k=1}^{K} \mu_k^m} \tag{19}$$

where M is the number of non-dominated solutions and K is the number of objectives. The solution with the largest membership value, i.e., $max\ (\mu^m)$, is the best compromise solution.

4.1.3. Test system, case studies, and algorithm properties

The developed TEP methodology is tested with reference to the IEEE 118-bus benchmark system [96], which contains of three zones (shown in the one-line diagram in [97]) 186 branches, 9 transformers, 118 nodes and 54 generators with a combined maximum capacity of 7'200MW. The power demand across all load buses amounts to 3'733.07MW. This value is considered to be the mean total load over the course of a year and the yearly load curve is extracted from [98].

We perform two case studies to demonstrate the applicability of the proposed TEP methodology. In Case Study 1, all branches are selected as candidates. Although in practice not all branches are considered as candidates, we perform this assessment to show the robustness of the method when applied on a real size power system. In this case study, the costs of building/upgrading existing branches are estimated based on the method used in [99]. In Case Study 2, a list of pre-determined candidate branches is taken from [60]. This is done for the sake of comparison with established methods. The candidate branches in [60] only contain information on the corridor, cost, and rating of the proposed branches, but not on their resistance (R), reactance (X) or susceptance (B). Therefore, these properties are determined by scaling R, X and B of the already existing branch in a corridor as a function of the transmission capacity [70]. The candidate branch data is given in Table A.1, Appendix A [82].

In both case studies, a list of 1000 contingencies are used for each loading condition and 18 loading conditions are selected from the yearly load curve, resulting in 18000 cascading failure simulations for each evaluation of the objective function, i.e., execution of the cascading failure model. Given a population of 60 individuals and a maximum number of 300 iterations, a total of $3.24 \cdot 10^8$ cascading failure simulations are performed to identify the Pareto front in Case Study 1. In Case Study 2, the population is 40 individuals, and the maximum number of iterations is 200, leading to $1.44 \cdot 10^8$ cascading failure simulations. The maximum number of iterations is the stopping criterion for the NSGAII. The TEP algorithm is developed in MATLAB and executed on a personal computer with an Intel(R) Core(TM) i9-9980XE CPU and on the Euler computer cluster at ETH Zurich [100]. The calculation time for a single run of the cascading failure model, parallelized based on the loading conditions takes, respectively, 40s on the personal computer and 15s on the computer cluster. Furthermore, the TEP problem is parallelized based on the population, and the calculation time vary from 10 hours to 50

hours, depending on the population size (40 or 60), the number of iterations (200 or 300), and the used computer. This is suitable for a typical offline problem such as the TEP.

4.1.4. Results

Case Study 1: all branches

In this case study, the 186 branches (lines and transformers) belong to the candidate list, i.e., every corridor in which a branch exists is considered as a potential corridor for building another branch with the same properties as the existing one. In total 186 decision variables are considered in the optimization problem defined in Eq. (14). The maximum number of branches to be built ($\overline{N}$) is eight (Eq. (15)), while the maximum number of new branches per corridor ($\overline{N_b}$) is one (Eq. (16)). Figure 8 shows the Pareto optimal frontier obtained using the proposed multi-objective TEP method and demonstrates that the cost of the TEP and the risk improvement are indeed conflicting objective. In general, Pareto optimum solutions which upgrade many lines result in large costs and large risk improvement.

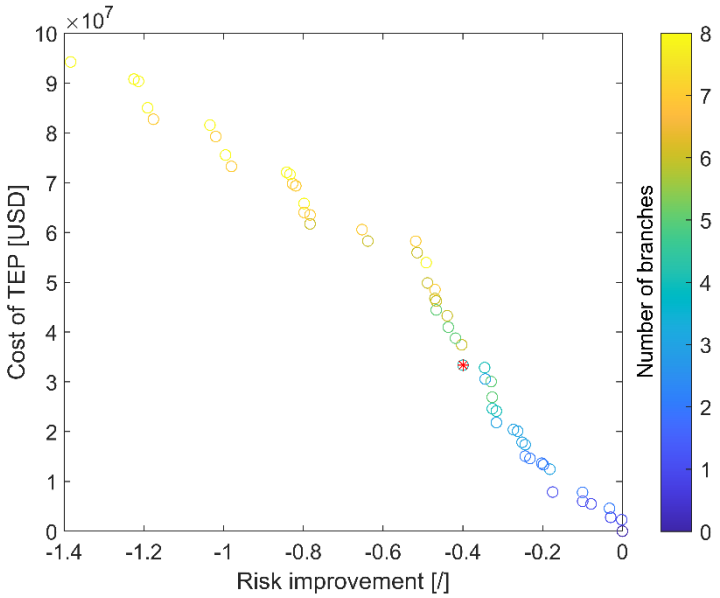


Figure 8. The Pareto front for Case Study 1. The color bar represents the number of branches in each TEP solution; the red star marks the BCS.

The risk improvement achieved by TEP ranges between 0 and 1.385 and the cost of TEP ranges from 0 to $94.25 \cdot 10^6$ USD. The red star marks the BCS, i.e., the expansion plan that is the best tradeoff between the objectives. Table 2 shows the list of the four branches to be built according to the BCS, and Table 3 shows the list of the eight branches to be built to achieve the best risk improvement solution.

Table 2. The list of the four branches to be built according to the BCS in Case Study 1. The first column denotes the buses between which a branch is connected, the second column denotes the type of branch, and the last column denotes the cost of the newly built/upgraded branches.

From – To bus	Type	Cost (USD)
5 – 8	Transformer	6’007’500
8 – 30	Line	9’072’000
26 – 30	Line	15’480’000
77 - 78	Line	2’790’000
		Total = 33’349’500

Table 3. The list of the eight branches to be built to achieve the best risk improvement solution in Case Study 1.

From – To bus	Type	Cost (USD)
2 – 12	Line	11’088’000
5 – 8	Transformer	6’007’500
8 – 30	Line	9’072’000
12 – 117	Line	21’000’000
23 – 32	Line	17’295’000
26 – 30	Line	15’480’000
77 – 78	Line	2’790’000
110 - 112	Line	11’520’000
		Total = 94’252’500

The risk improvement of the BCS is 0.399, while the risk improvement of the most expensive solution is 1.385. All the branches present in Table 2 are also present in Table 3, which emphasizes the importance of these branches in minimizing the risk of cascading failures. Branches 5 – 8 and 77 – 78 are the branches that most frequently show up in the obtained Pareto optimal solutions (Figure 8). In particular, the failure of transformer 5 – 8 during peak demand conditions causes the overload in line 8-30, which evolves into a cascade resulting in a DNS of 3190 MW. One of the reasons for the large DNS is the violation of steady state stability limits due to the lack of transfer capacity in zone one of the grid. The grid only stabilizes after five cascading stages and the formation of three islands. Two of the islands form in zone one and the third island spans across zone two, zone three and part of zone one.

Figure 9 shows the risk curves for the reference grid configuration, the grid configuration based on the BCS, and the grid configuration as found in the best risk improvement (RI) solution. Either of the expansion plans yield a significant risk improvement, i.e., less cascading events with lower intensity, with the best risk improvement solution showing much higher improvement compared to the BCS regarding the risk of cascading failures. However, as shown in Table 2 and Table 3 the capital cost associated with the best solution is three times higher than the capital cost of the BCS. Since the Pareto front provides multiple solutions, the decision maker – usually the TSO – can choose an expansion plan depending on risk improvement preferences and available budget.

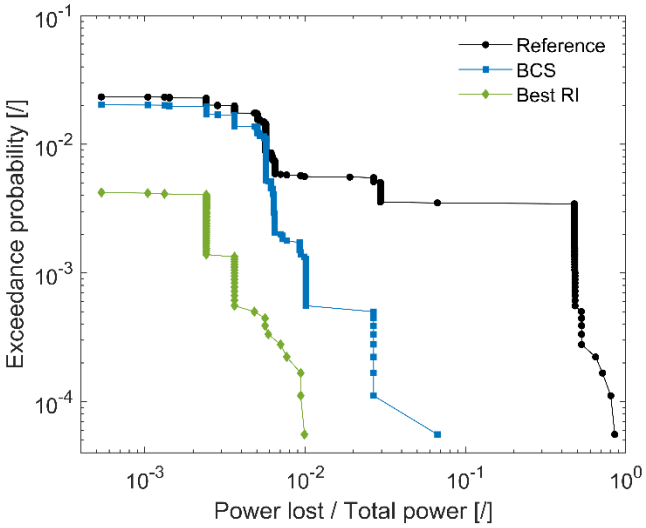


Figure 9. Risk curves obtained in case study 1. The black curve corresponds to the reference power grid, the blue curve shows the behavior of the power grid for the BCS, and the green curve depicts

power grid behavior for the best risk improvement (RI) solution. In total 18000 cascading failure simulations are performed to derive a single risk curve.

Figure 10 represents a risk surface consisting of all the risk curves calculated for each Pareto front solution given in Figure 8. The x-axis denotes the risk improvement for each Pareto solution. The y-z plane depicts the risk curves for each TEP solution from the Pareto front. Figure 10 quantifies the risk transition from the solution with the highest risk to the solution with the lowest risk and their associated costs. When the risk improvement exceeds 0.45, there is significant decrease of the cascading intensity (Power Lost / Total Power) of the most severe event form each risk curve. Figure 10 shows the importance of knowing the risk curve associated to the TEP solutions for making the most informative decision about transmission upgrades.

Table 4 and Table 5 show the cascading stages of the worst cascade observed, i.e., the cascade causing the highest DNS, for the reference grid configuration (Table 4) and for the BCS (Table 5), respectively, for the same list of contingencies. The worst DNS does not always correspond to the last event (point) in the risk curve, because the horizontal axis does not show the DNS in MW, but the ratio of power lost (DNS) and power demanded in the system. From Table 4 and Table 5 it is evident that for the reference grid the worst cascade causes much larger DNS. In addition, the number of cascading stages is higher compared to the grid expanded based on the BCS.

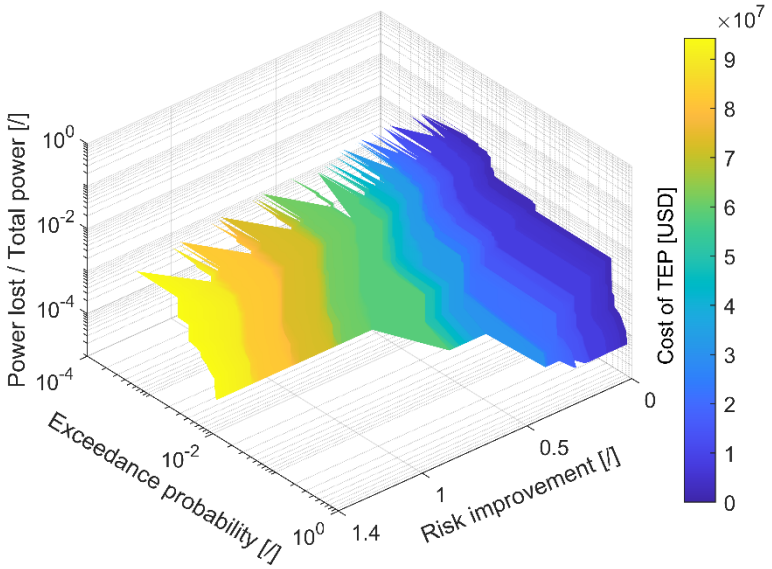


Figure 10. Risk surface obtained in case study 1. The color bar denotes the cost of TEP.

As for the best risk improvement solution, the worst DNS event (55 MW) is not a result of a cascade, but a load disconnection due to an initial set of contingencies, i.e., the simultaneous failure of two lines. In fact, most DNS events are the result of an initial contingency set disconnecting loads. Furthermore, there are no cascades with more than two cascading stages. In conclusion, expanding the grid based on the proposed TEP procedure effectively prevented the unfolding of cascades that result in large DNS events that can have negative effects on societal welfare.

Table 4. Cascading stages of the worst cascade observed in the reference grid; the total demand is 5681.75 MW.

Cascading stages	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Contingencies	26-30 37-38	23-32	30-38	22-23	25-27	23-24	42-49	42-49
DNS (MW)	0	0	53.4	101	2901.5	2916.6	3606.5	4848.2

Table 5. Cascading stages of the worst cascade observed in the BCS-improved grid; the total demand is 5681.75 MW.

Cascading stages	**1**	**2**	**3**
Contingencies	25-26 32-113	23-32	17-31
DNS (MW)	0	0	320

Case Study 2: pre-selected candidate branches

In this case study, the branches for the expansion can only be selected from a list of 17 candidate branches provided in [60]. Therefore, 17 decision variables are considered in the optimization problem defined in Eq. (14). The same list of contingencies and the same 18 load conditions as in Case Study 1 are used. The maximum number of branches to be built in total ($\overline{N}$) and maximum number of times a branch can be selected ($\overline{N_b}$) is set to eight (Eq. (15)) and one (Eq. (16)), respectively. Figure 11 shows the Pareto optimal front obtained using the proposed multi-objective TEP method.

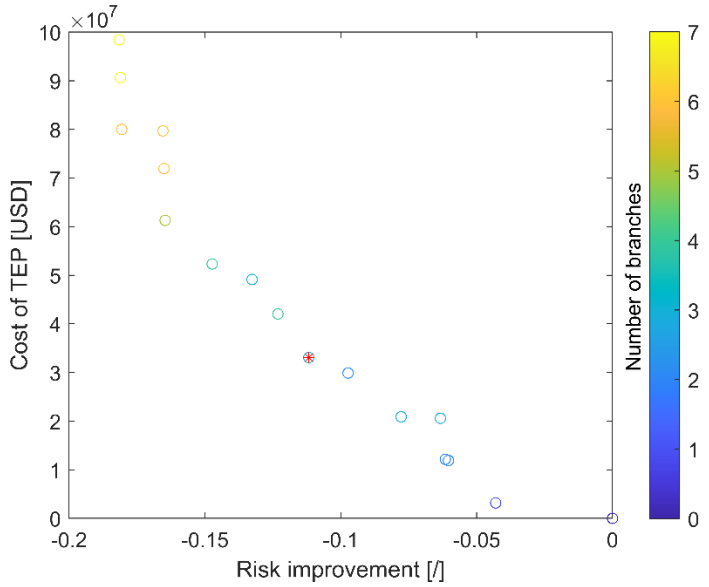


Figure 11. The Pareto front for Case Study 2. The color bar represents the number of branches in each TEP solution; the red star highlights the BCS.

Figure 11 shows that the obtained Pareto front does not reach the same risk improvements compared to Case Study 1, because the capacity of the branches in the list is smaller compared to the existing branches, and the pool of candidate branches to choose from is significantly smaller, i.e., 17 against 186 branches. Table 6 shows the list of three branches to be built according to the BCS, and Table 7 shows the list of eight branches to be built to achieve the best risk improvement solution in Case Study 2. The risk improvement for the BCS is 0.1118, while the risk improvement for the best risk improvement solution is 0.1816.

Table 6. The list of three branches to be built according to the BCS in Case Study 2.

From – To bus	**Type**	**Cost (USD)**
8 – 30	Line	8’670’000
12 - 117	Line	3’180’000
17 – 30	Transformer	21’220’000
		Total = 33’070’000

Table 7. The list of eight branches to be built to achieve the best risk improvement solution in Case Study 2.

From – To bus	Type	Cost (USD)
8 – 30	Line	8’670’000
12 – 117	Line	3’180’000
17 – 30	Transformer	21’220’000
37 – 38	Transformer	19’240’000
65 – 68	Line	18’410’000
77 – 78	Line	8’970’000
110 – 112	Line	18’710’000
		Total = 98’400’000

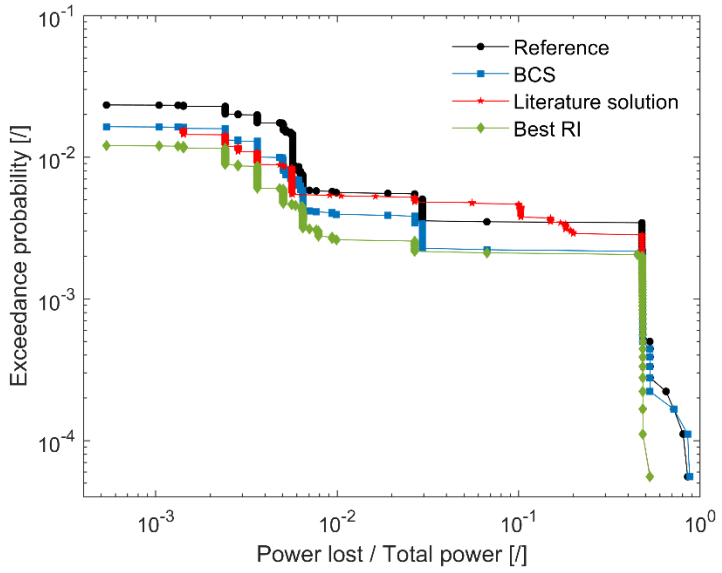


Figure 12. Risk curves obtained in Case Study 2 and the risk curve obtained for the reference system with the expansion solution given in [60].

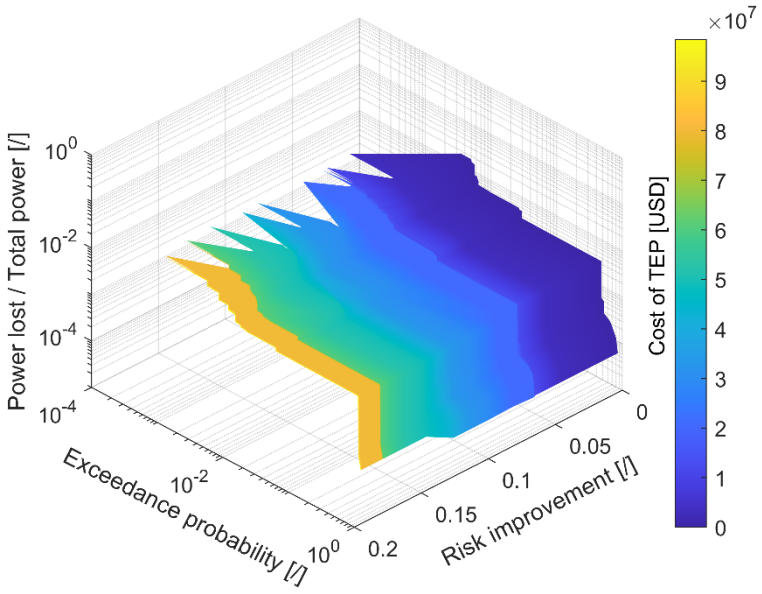


Figure 13. The risk surface obtained in Case Study 2. The color bar denotes the cost of TEP.

As in Case Study 1, the results of Case Study 2 show that the BCS expansion plan is a subset of the best risk improvement solution expansion plan (Table 7). Figure 12 shows the risk curves for the reference grid configuration, the grid configuration based on the BCS, and the grid configuration as found in the best risk improvement solution. Both solutions yield a significant risk improvement. Figure 13 shows the risk surface consisting of all the risk curves calculated for the Pareto front in Figure 11. Figure 13 quantifies the risk transition from the solution with the highest risk to the solution with the lowest risk and their associated costs.

Table 8 shows the cascading stages of the worst cascade observed with the best risk improvement solution. The worst cascading event unfold in a cascade with lower DNS compared to the reference grid (Table 4). However, despite BCS having a better overall risk improvement compared to the reference grid, the worst cascade has the very similar DNS as the reference grid (Table 4). Additional analyses and comparison with the literature are provided in [82].

Table 8. Cascading stages of the worst cascade observed with the grid after expanding based on the best risk improvement solution; the total demand is 5866.3 MW.

Cascading stages	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Contingencies	20-21 25-27	23-22	26-30	23-24	30-38	19-34	15-33
DNS (MW)	0	0	0	0	616	2955.4	3096.9

4.2. Importance list based transmission expansion planning method

The objective of the importance list based TEP method is to provide an expansion plan that achieves a predefined system security, i.e., risk of systemic failures. The selection of the target risk curve depends on the availability of historical data. If target security does not exist, the risk curve obtained for the base-case system configuration is set as reference security. In this case, either the expansion budget or maximum number of upgrades must be provided. The same three criteria as described in Section **Error! Reference source not found.** are valid here as well. Unlike the optimization-based expansion m ethod, the list based TEP method does not need optimization, which makes it simpler and faster. However, there is a risk that the obtained solution is not optimal. Note that this method only proposes upgrades of existing branches. Each upgrade is modeled as a parallel line or transformer.

4.2.1. Modeling features

The importance list based TEP relies on the cascading failure analysis model to derive expansion strategies. Two concepts are used: (1) the first concept is based on the impact of line/transformer failures on the DNS, which relies on the cascading failure model recording all events where an initial contingency results in load shedding. (2) The second concept is based on the frequency of line overloads; this concept relies on the cascading failure model recording all line overloads occurring after an initial contingency. The model performs the following steps (Figure 14):

1. *Creation of lists*
   The impact of line/transformer failures on the DNS, as well as the frequency of line overloads are stored in two lists, after which the transmission system lines/transformers are ranked from highest to lowest frequency/impact. Additionally, a third list is created as a hybrid of both lists. The hybrid list must not have repeating lines/transformers, i.e., if a line/transformer exists in both list the one with the highest rank is kept in the hybrid list.

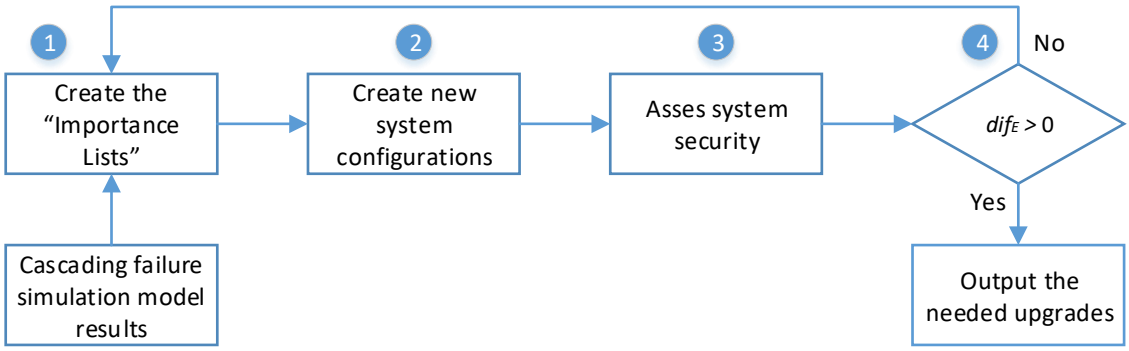


Figure 14. Expansion planning model flow chart.

2. *Expansion strategy*
   Each list is used to create an expansion strategy, and each strategy is used to create a new/updated network topology. Therefore, three new network configurations are formed. The number of lines/transformers to be upgraded from each list is predefined, with the recommended number of line/transformer upgrades being equal to one, i.e., only one line/transformer should be upgraded in each iteration. This procedure prevents overinvestments in the system; however, upgrading only one component per iteration can also lead to high computing times.

3. *Asses the expansion impact*
   After each network configuration has been evaluated separately using the cascading failure simulation algorithm, the obtained results are compared. This comparison is based on the difference between the simulation results and the (desired) reference system security. One of the main results from the cascading failure model is the distribution of DNS events. Consequently, we calculate the difference between the simulated and the reference DNS distributions. The algorithm has several options for calculating this difference, such as distance correlation [101], Kolmogorov-Smirnov test [102, 103], Kullback-Leibler divergence [104], Wasserstein distance [105], and Minkowski distance [83]. The latter is computed as follows:

$$dif_E = \left(\sum \left| CCDF_{DNS_{DATA}}(dns_{DATA}) - CCDF_{DNS}(dns) \right|^p \right)^{1/p} \quad (20)$$

   where dif_E is a measure of the differences between reference data and simulation results, $CCDF_{DNS_{DATA}}$ is the complementary cumulative distribution function (CCDF) of the historical DNS, $CCDF_{DNS}$ is the CCDF of the simulation results, dns_{DATA} and dns are the DNS value vectors, $p = 1$ results in the Manhattan distance, while $p = 2$ yields the Euclidean distance. Because the reference data and the simulation data may not contain the same number of DNS values, both vectors are interpolated using linear polynomials to construct new data points within a discrete range, thus making the subtraction possible. Equation (20) always results in a positive number for both $p = 1$ and $p = 2$. It is, therefore, crucial to know the sign of the function difference, which in this case is determined only as the sum of the differences between the CCDF of the reference data and the CCDF of the simulation results. A negative sign means that the security obtained with the simulations is lower than the security of the system for which the reference data is given.

   The function difference (distance) is calculated for all three new system configurations and only the system resulting in the highest function difference is kept, the others are discarded. In case the required system security is not yet reached, the results obtained from the cascading failure simulations of the other two system configurations are used to create the new importance lists, which will be further evaluated.

4. *Stopping criteria*
   The function difference is also used as a stopping criterion. In other words, the expansion planning continues until the function difference becomes a positive value or a predefined number of iterations is reached.

4.2.2. Algorithm features

The default setting of the transmission system expansion planning algorithm is to calculate and use the function difference based on the DNS results. However, the line outage probability function difference can be used as an alternative. This difference is calculated as the difference between the line outage data (historical or reference system configuration) and the simulation results. One must not confuse the failure probability of a line with the line outage probability. The latter is calculated based on the participation of each type of outage in the total number of outages. In other words, the number of all single component outages is divided by the total number of outages, the number of all double component outages is divided by the total number of outages etc.

Besides the system security, an expansion budget can also be used as the main criterion for the expansion plan, enabling the algorithm to find the best expansion plan considering a given budget. Violation of the budget is possible. If neither a security criterion nor budget exists, the user is motivated to use a maximum number of upgrades.

The expansion planning algorithm can be executed significantly faster if the user chooses to use only one expansion strategy and, thus, only one new system configuration is evaluated at a time. Once the expansion method has been applied on a given power system, it becomes evident which expansion strategy is giving the best result. Therefore, for further utilization of the algorithm on the same system, the user is encouraged to use the strategy that is clearly giving the best results without calculating the security impact of the other two strategies. However, if the system is significantly changed the user should reassess the expansion strategies and re-select the most adequate one. In case the calculation time is not a concern, it is recommended that the user selects the default option, i.e., assessing and comparing all strategies.

The application and obtained results of the list based TEP are presented in [106] and [107]. In both works we assess scenarios of the Swiss energy transitions and the need for transmission expansion. We show how particular upgrades of grid assets improve system security.

5 Power system vulnerability analyses

Power systems vulnerabilities are reviled either by accidental or deliberate events [108]. As accidental events are considered the random failures and natural hazards [16], and as deliberate events are considered the physical attacks, cyber-attacks, and electromagnetic pulses (EMP) [109]. An extensive body of research on power system vulnerability, performed in the past decades, highlights the relevance of the topic [16]. The vulnerability of a system can be defined as the "manifestation of the inherent states of the system that can be exploited by an adversary to harm or damage the system" [110].

Scientific literature reviles variety of models and methods used for power system vulnerability analysis. In general, they can be placed in two categories: (1) topological approaches; and (2) flow-based approaches. The topological approaches are based on network connectivity and the complex network concept [111, 112]. They rely only on information on the gird topology, and are computationally efficient [113]. The flow-based approaches are based on the power flow dynamics and physical characteristics of the power grid. They rely on the information of the physical characteristics of the grid, and can be computationally expensive [33]. A more detailed review of all types of methods used in power system vulnerability analysis is present in [16].

A topological model for the identification of groups of most critical elements in the Italian high-voltage grid is presented in [113]. The model identifies the important groups by applying the betweenness centrality of groups of nodes and groups of edges, and the variation in network connection efficiency. An integrated topological and reliability framework, comprising different centrality measures, for assessing the vulnerability of the high-voltage Iranian power grid is introduced in [114]. The paper shows that the reliability characteristics differ from the topological results, because they are mostly uncorrelated. Eusgeld et al. [115] use topological analysis to identify the most relevant parts for system vulnerability, supplemented by a physical analysis to understand better the mechanisms responsible for these vulnerabilities. Similarly, [116] presents an extended topological approach that, besides the regular topological metrics, considers the line flow limits and applies a real power-flow allocation over lines. Furthermore, topological models are used in combination with probabilistic safety assessment (PSA), where fault trees are applied to identify and rank critical components in the power grid [117]. A statistical model for estimating the probability of restoring power to the electric grid before a determined time after a blackout is presented in [118]. This developed knowledge can be used to inform event tree calculations in PSA models. In general, the topological approaches are criticized for their inability to capture the physical process occurring in the power grids, despite the attempts to perform additional physical analysis.

On the other hand, the main arguments against the flow-based approaches are their complexity and computational inefficiency. However, these approaches are widely used today, e.g. a DC power flow based method that exploits game theory is used to assess the vulnerability of a power system [119]. Similarly, [120] introduces an optimization method in a game theory framework, which utilizes linear approximation AC power flow and aims to identify critical components in a power system. The DC power flow based OPA model [18] and the AC power flow based Manchester model [22] are the most recognized cascading failure analyses models. In [121], the criticality of individual components of power systems is assessed via AC based cascading failure model considering multi-element failures. The proposed approach builds upon the Fussell-Vesely importance measure quantifying the performance decrease ratios due to the loss of grid elements. A flow-based approach that utilizes a novel mixed integer linear programming optimization framework is described in [122]. The study assesses the impact of several natural hazards on the grid infrastructure employing failure and recovery probabilities for system components. Similarly, [123] proposes a resilience enhancement framework for interdependent gas and power grids. The framework utilizes a multi-objective genetic algorithm to find the optimal resilience enhancement strategies. A study of the vulnerability and reliability of a power system with various levels of penetration of renewables is performed in [9]. The

results show that the vulnerability and reliability are affected by the renewable sources and the electrical interconnections. Abedi et al. [124] compare a DC and AC approaches for vulnerability, reliability, and contingency assessment of a power grid. The results show that the DC model significantly underestimates the reliability, and, therefore, conclude that the AC models should be prioritized.

A comparison of the criticality of power grid complements is performed using multiple cascading failure models in [34]. Different models show some inconsistencies when allocating grid assets criticality, which indicates that cascading failure analysis is still an active field of research. Cascading failure models are known for being capable of performing comprehensive cascading failure analysis and assessing the vulnerability of the power grids. However, their identification of the critical components mainly focuses on the impact on the demand not served (DNS) caused by a single branch failure or by sets of randomly generated contingencies. The methodology proposed in [125] goes further, combining a DC based cascading failure model and a stochastic "Random Chemistry" (RC) algorithm, to identify large collections of multiple contingencies that initiate large cascading failures. Yet, this approach does not utilize optimization to determine the sets of contingencies causing the worst blackouts (i.e., DNS). Furthermore, methods using optimization are also applied to uncover small subsets of events with high impact [126, 127]. However, these methods rely primarily on detecting limit violations and load shedding actions, and they do not simulate the full propagation of disturbance events, i.e., do not perform cascading failure analyses.

We address this research gap by introducing a vulnerability assessment methodology that couples an AC based cascading failure simulation model and a meta-heuristic optimization procedure. The use of cascading failure analyses is motivated by the fact that modern power systems, despite the multiple layers of protective schemes, are still experiencing blackouts that are mainly caused by cascading failures [2, 3]. The objectives of the proposed methodology are to: (1) provide ranking of the most important branches in the transmission grid, and (2) identify sets of branches that will cause the cascade with the highest intensity if simultaneously tripped. The first objective is achieved by ranking the criticality of the branches using two criteria (i) the impact that each branch failure has on the DNS and (ii) the frequency of line overload. The second objective is achieved by hard linking an AC based cascading failure simulation model and a meta-heurist based optimization procedure. This link allows for optimal identification of vulnerabilities that have the potential to spread within the power grid, while considering automatic and manual system responses. The algorithm developed for the purpose of this study is applied to the IEEE 118-bus test system and the Swiss power grid. The results provide a ranking of the branches according to the two different criteria and identify the sets of branches that are most critical to system security. The potential application of these results includes asset upgrade and guidance for operators on implementation of procedures for improving the system response during failures or malevolent acts. This work is presented in detail in [128].

5.1. Methodology

The vulnerability method comprises the CFS model from Section 3 and a meta-heuristic based optimization procedure.

5.1.1. Criticality measures

We utilize the CFS model to accomplish the first objective, i.e., to rank the branches in the transmission grid according to their criticality. For this purpose, the model relies on two criticality measures: (i) branch impact on DNS, and (ii) frequency of branch overload.

The CFS model initializes with a separate list of contingencies for each of the pre-selected loading conditions. Each list consists of pre-selected number of sets of contingencies. Thus, the total number of contingency sets, i.e., the total number of cascading failure simulations is equal to the number of loading conditions multiplied by the number of contingency sets per contingency list. For each cascading simulation, the model records the DNS caused by the initial contingency set. If the set consist of more than one contingency, same amount of DNS is assigned to all of them. The process is repeated

for all sets of contingencies and loading conditions and the recorded DNS is aggregated for each contingency (branch):

$$BD_k = \sum_{m=1}^{M} \sum_{n=1}^{N} DNS_{n,m}^{k} \tag{21}$$

where, BD_k is the criticality of branch k associated with the branch impact on DNS criticality measure, M is the number of loading conditions (power demands), N is the number of contingency sets per loading condition, and $DNS_{n,m}^{k}$ is the DNS assigned to branch k after performing the cascading simulations for contingency set n and loading condition m. $DNS_{n,m}^{k}$ is equal to zero in case that the cascading failure analysis do not end with DNS.

Furthermore, the model records the frequency of branch overloads, i.e., the overloaded branches after the first cascading stage is completed. These overloads are important because they directly contribute to the evolution of a cascading event. The counting is performed for all sets of contingencies and loading conditions and the recorded overloads are aggregated for each branch:

$$BF_k = \sum_{m=1}^{M} \sum_{n=1}^{N} F_{n,m}^{k} \tag{22}$$

where, BD_k is the criticality of branch k associated with the frequency of branch overload criticality measure, and $F_{n,m}^{k}$ is equal to 1 if branch k is overloaded after the contingency set n is introduced into the system, otherwise $F_{n,m}^{k}$ is equal to zero.

5.1.2. Vulnerability identification method

To accomplish the second objective, we introduce the vulnerability identification method, which aims to find the branches that if simultaneously tripped will cause the blackout with the highest intensity. We are looking for a set of two, three, or more lines/transformers that if simultaneously tripped will have the highest risk impact. The method uses the CFS model to assess the impact of an initial set of contingencies, and a meta-heuristic based optimization procedure to explore the contingency space.

The objective function can either maximize the average DNS (DNS_{avr}) over a selected number of representable loading conditions from a yearly load curve or find the maximum DNS over these demands (DNS_{max}). Therefore, two objective functions are defined:

Objective function one:

$$\max\left(DNS_{avr}(C_{set})\right) \tag{23}$$

Objective function two:

$$\max\left(DNS_{max}(C_{set})\right) \tag{24}$$

where C_{set} is a set of contingencies with a size of N_{set}, i.e., the decision variables. We decide which objective function to use and the size of the contingency set, N_{set}, before starting the vulnerability identification.

Using the objective in Eq. (23), the algorithm identifies the contingency set that results in the largest average DNS over all of the simulated hours (i.e., system loadings). Therefore, the identified contingency set is expected to have high implication to system security at any hour and any loading of the system during the year. The measures undertaken against the effects of the identified contingency set will provide an adequate level of security during most operating conditions. Using the objective in Eq. (24), the algorithm identifies the contingency set that results in the largest maximum DNS over all of the simulated hours (i.e., system loadings), i.e. the largest disruption caused by the cascading event. The identified contingency set is expected to have strong implication only for some of the yearly loading conditions and hours. The measures undertaken against the implications of the contingency

set identified by Eq. (24) will provide an adequate level of security for the worst event identified by the proposed method.

The optimization procedure uses a Genetic Algorithm (GA), which is meta-heuristic optimizer capable to deal with simulation-based optimization problems as the one presented here. The flowchart of the vulnerability identification method is shown in Figure 15. The procedure starts with a population of random solutions, Pop:

$$Pop = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,N_{set}} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,N_{set}} \\ \vdots & & \ddots & \vdots \\ c_{N_p,1} & c_{N_p,2} & \cdots & c_{N_p,N_{set}} \end{bmatrix} \quad (25)$$

where $c_{i,j}$ is the contingency, i.e., branch j to be tripped ($j = 1,2,\dots,N_{set}$) at solution i ($i = 1,2,\dots,N_p$), and N_p is the population size, i.e., the number of potential solutions. Practically, each row in Pop represent a single solution, which consists of integer values between 1 and the number of branches in the system.

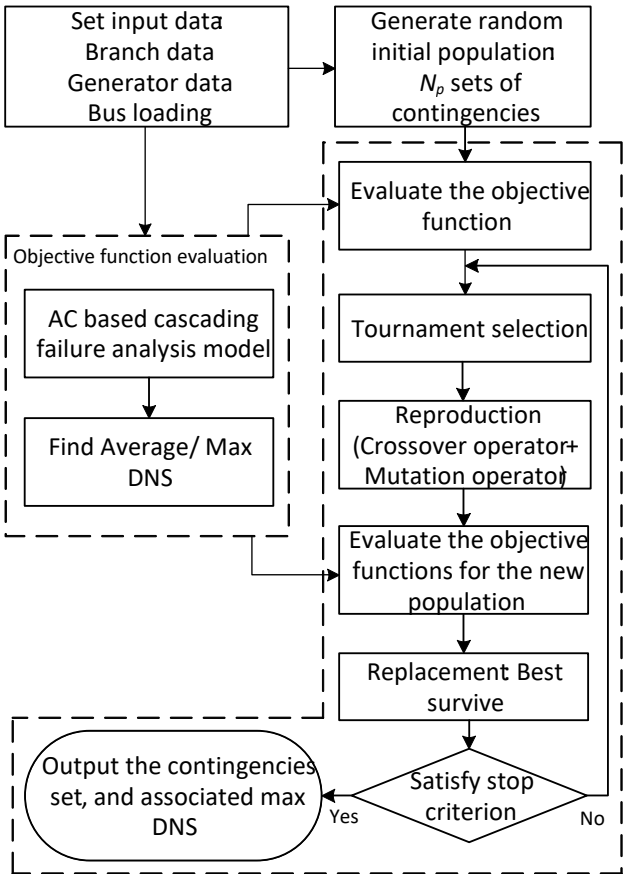


Figure 15. Flow chart of the vulnerability identification method.

The objective function given by Eq. (23) / Eq. (24) is evaluated for each of the randomly generated solutions. The tournament selection operator is utilized to select the solutions (chromosomes) for reproduction. The selected solutions enter the reproduction step, where the blending crossover operator (BLX-α) and the non-uniform mutation operator are applied to produce the population of new solutions. The fitness of the solutions from the new (child) population is calculated and compared with the fitness of the solutions from the old (parent) population. Only the best half of the solutions survive while the rest are discarded in a process known as elitist replacement. This procedure is repeated until the maximum number of iterations is reached. A detailed description of the applied GA is given in [129].

The main output of the algorithm is the set of contingencies that results in the highest DNS with respect of the selected objective function. Furthermore, the algorithm outputs the worst cascade including the cascading stages and the DNS at each stage.

5.2. Test case

Two test systems are used: the IEEE 118-bus test system and Swiss power system. The 118-bus test system contains 186 branches, 9 transformers, 118 nodes and 54 generators with a combined maximum capacity of 7'200MW [96]. By default, the power demand across all load buses amounts to 3'733.07MW. This value is considered to be the mean total load over the course of a year, with the yearly load curve extracted from [98] being scaled accordingly. The one-line diagram of the IEEE 118-bus test system marking the zones in the grid is given in [96, 97]. The Swissgrid AG provides the Swiss transmission system data under an NDA agreement, including current branches and planned grid expansions up to 2025. The power plant data and power demand data are obtained from various open sources. We represent each neighboring country with a single node, e.g., all lines connecting Switzerland and Germany are connected to one representative node in Germany. Furthermore, at each of these nodes we impose the historical hourly import/export for the respective country. Due to sensitivity of the obtained results, detailed description of the vulnerabilities of the Swiss power system and the grid data are not presented here.

The algorithm assessing grid vulnerabilities is developed in MATLAB and executed on an Intel(R) Core(TM) i9-9980XE CPU and the Euler computer cluster at ETH Zurich [100]. Calculation times vary from one to 30 hours, depending on the size of the contingency set, the population size, the number of iterations, the size of the analyzed system, and the computing machine.

5.3. Analysis and results

Both systems are assessed with respect to both objectives, i.e., the importance of each branch is determined, and the most critical sets of contingencies are identified.

5.3.1. The IEEE 118-bus test system case study

5.3.1.1. Criticality ranking results

The criticality ranking results for the IEEE 118-bus test system are obtained with a single run of the cascading failure model. With a list of 1000 contingency sets per loading condition and 18 loading conditions, the total number of executed cascading simulations in a single run of Cascades is 18000. Out of these contingencies, 91.1% are single branch, 8.3% are double branch, 0.5% are triple branch failures, and the remaining involve the failure of four branches or more. Figure 16 shows branch criticality in the IEEE 118-bus test system according to the a) branch impact on DNS (Eq. 21), and b) frequency of line overload (Eq. 22).

Lines 8-9 and 9-10, with similar values, are the most critical components according to the branch impact on the DNS criticality measure. The failure of either of these two lines disconnects a 300 MW generator connected to bus 10. The loss of the generator causes frequency instability and results in DNS under some of the selected loading conditions. Transformer 5-8 is the third most critical component in the system according to the branch impact on the DNS criticality measure. This transformer is the main link between the generator at node 10 and the part of Zone 1 with the highest concentration of loads. Cascading failures resulting in significant DNS begin by a simultaneous failure of Line 8-9, Line 9-10, or Transformer 5-8 with other branches in the grid, in particular branches in Zone 1. Furthermore, the disconnection of Line 8-9, Line 9-10, or Transformer 5-8 initiates an overload of Line 8-30, which is the most frequently overloaded branch in the IEEE 118-bus test power system. This Line is one of the main links that connect the northern and southern part of Zone 1.

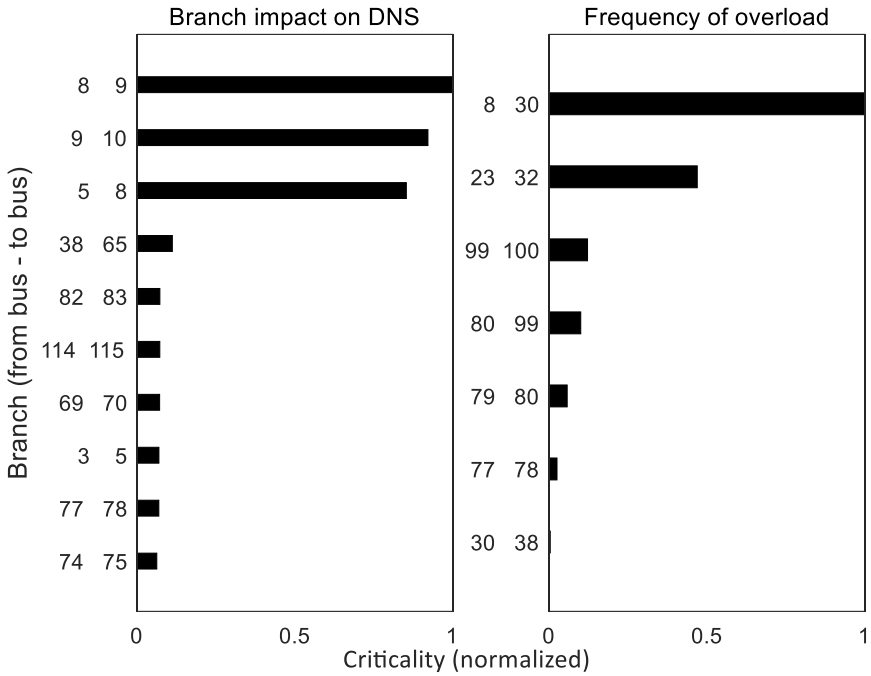


Figure 16. Branch criticality in the IEEE 118-bus test system according to: branch impact on DNS (left), and frequency of branch overload (right).

Table 9. A cascading event example with all cascading stage and the DNS at each stage.

Cascading stages	1	2	3	4	5	6	7
Contingencies	8-9 25-27	23-32	8-30	17-30	33-37	19-34	22-23
DNS (MW)	0	0	0	517.22	917.94	1478.96	1520.46

The cascading event in Table 9 is initiate by a simultaneous failure of Lines 8-9 and 25-27 causing a chain of failures with seven stages and a DNS of 1520 MW. The disconnection of Line 8-9 results in the separation of Buses 9 and 10 from the main grid, and therefore splits the grid in two parts (islands). Buses 9 and 10 have no connected loads, and therefore the generator at Bus 10 is shutdown. The first three stages of the cascade do not result in DNS, and each stage proceed to the next with a disconnection of a single branch. The first DNS occurs at stage four and is a result of the system operating outside the steady state stability limits. This problem is resolved by uniform load shedding at all buses in the grid. Similar operating conditions occur at stages five and six and are resolved with load shedding. Furthermore, at stage seven with the disconnection of Line 22-23 the main grid splits into two islands, where the first (smaller) island consists of most of Zone 1 and one bus from Zone 2 and the second (larger) island consist of part of Zone 1, most of Zone 2 and all of Zone 3. After the splitting, the smaller island has a frequency deviation, which is resolved with ~40 MW of load shedding. The larger island is stable and thus no additional actions are undertaken. At the end of stage seven, there are no overloads in any of the islands, and therefore the cascading simulation stops.

5.3.1.2. Most critical sets of contingences

To identify the sets of critical branches in IEEE 118-bus test system we run the vulnerability identification procedure described in Section 5.1.2. The set size for the IEEE 118-bus test system spans from two to seven contingencies, such that the algorithm is executed independently for each set size. To perform the analyses, we generate a random population of N_p contingencies for the designated set size, i.e., for a set size of two we randomly generated 90 samples of two simultaneous branch failures. Table 10 and Table 11 show the sets of contingencies resulting with the worst cascades in the IEEE 118-bus test system. The Table 10 contingency sets are obtained with the average DNS based objective function (Eq. 23), and the Table 11 sets are obtained with the maximum DNS based objective function (Eq. 24).

Table 10. List of the most critical sets of contingencies obtained using the average DNS based objective function (Eq. 23).

Set	From Bus	To Bus	Branch type
2-branch	5	8	Transformer
	25	27	Line
3-branch	17	30	Transformer
	25	27	Line
	37	38	Transformer
4-branch	5	8	Transformer
	22	23	Line
	25	27	Line
	37	38	Transformer
5-branch	5	8	Transformer
	21	22	Line
	25	27	Line
	37	38	Transformer
	45	46	Line
6-branch	5	8	Transformer
	22	23	Line
	24	70	Line
	25	27	Line
	37	38	Transformer
	45	46	Line
7-branch	5	8	Transformer
	17	30	Transformer
	22	23	Line
	25	27	Line
	37	38	Transformer
	42	49	Line
	45	46	Line

Table 11. List of the most critical sets of contingencies obtained using the maximum DNS based objective function (Eq. 24).

Set	From Bus	To Bus	Branch type
2-branch	17	30	Transformer
	37	38	Transformer
3-branch	26	30	Line
	37	38	Transformer
	45	46	Line
4-branch	5	8	Transformer
	25	27	Line
	37	38	Transformer
	45	46	Line
5-branch	5	8	Transformer
	22	23	Line
	25	27	Line
	38	65	Line
	45	46	Line
6-branch	5	8	Transformer
	22	23	Line

	25	27	Line
	38	65	Line
	45	46	Line
	60	62	Line
7-branch	9	10	Line
	23	25	Line
	25	27	Line
	37	38	Transformer
	45	46	Line
	69	70	Line
	69	75	Line

Table 10 and Table 11 show that some branches are participating in many of the identified contingency sets. Transformer 5-8, which is the third most critical branch according to the impact on DNS criticality measure, is part of almost all contingency sets. On the other hand, Line 8-9, which the most critical branch according to the impact on DNS criticality measure, does not show in any of the contingency sets. Whereas Line 9-10, which is the second most critical branch, only exists in the 7-branch contingency set from Table 11. This is so because in the identification of the contingency sets the algorithm is maximizing the total contribution of a set, instead of the contribution of a single component to the DNS. Furthermore, it is evident that the branches ranked according to the frequency of overload criticality measure does not show in any of the most critical contingency sets in Table 10 and Table 11. This is so because the objective of the vulnerability procedure is to determine the final impact of the initial set of failures, and therefore the contribution of the branches that aid the unfolding of a cascade is not directly measured. In other words, the objective functions (Eq. 23 and Eq. 24) are more in alignment with the impact on DNS criticality measure. A closer look of the branches given the Table 10 and Table 11 shows that all of these branches are part of Zones 1 and 2, and the rest are interconnectors between these two zones. Moreover, the analyses show that simulations failure of two or more of these branches often causes overloads that frequently result in voltage and frequency instabilities, such as the event show in Table 9. Consequently, the branches that comprise the most critical set of simultaneous failures belong to Zone 1 and Zone 2 of the IEEE 118-bus test system.

Figure 17 and Figure 18 show the average and maximum DNS per contingency set obtained using the average DNS based and maximum DNS based objective functions, respectively. Both figures show that the IEEE 118-bus test grid, with a peek demand of 6025 MW, can encounter large DNS with only two simultaneous failures. Figure 17 shows that with three and seven simultaneous failures, the average DNS is similar, i.e., the average DNS does not significantly grow with the number of simultaneous failures. Similarly, Figure 18 shows that the maximum DNS caused by the 2-branch contingency set does not differ significantly from the maximum DNS caused by the 7-branch contingency set. Table 12 shows the cascade initiated by the most critical set of three simultaneous failures according to the maximum DNS based objective function. The DNS observed at stages two, three, and four is a result of load shedding at busses 43, 44 and 45, at which the voltage drops below the predefined limit of 0.92 p.u. At stage six, after the disconnection of Line 25-27, the system splits into two islands, where the small island consists of Buses 23, 25 and 26, and the large island represents the rest of the power system. Furthermore, at stage five, seven, and eight, the large island operates outside the steady state stability limits, which results in significant load shedding.

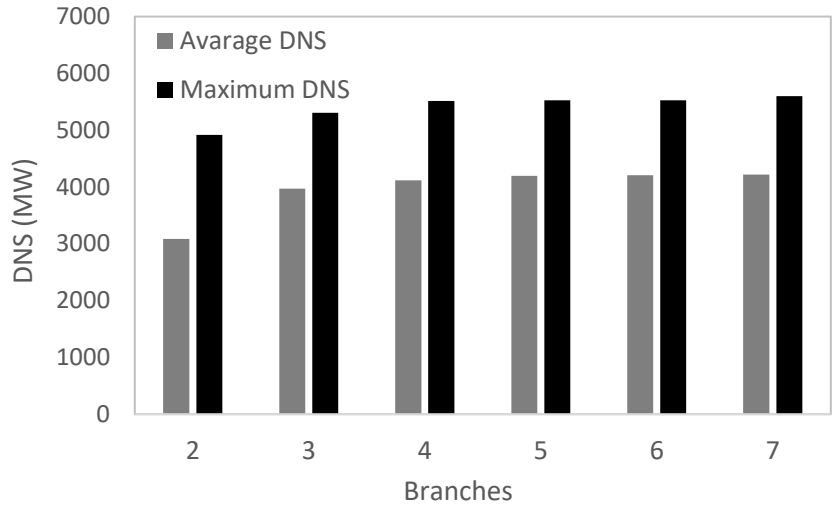


Figure 17. The DNS as function of the sets of critical contingencies obtained using the average DNS based objective function (Eq. 23).

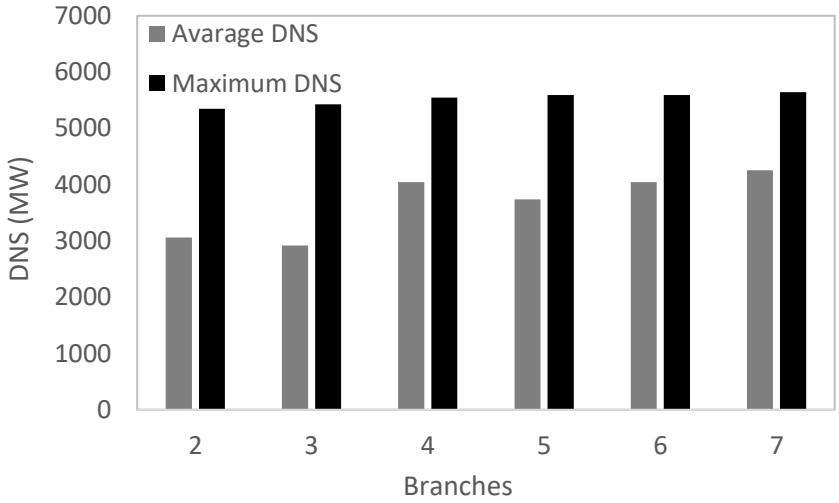


Figure 18. The DNS as function of the sets of critical contingencies obtained using the maximum DNS based objective function (Eq. 24).

Table 12. An IEEE 118-bus test system cascading event initiated by the 3-branch failure given in Table 11.

Cascading stages	1	2	3	4	5	6	7	8
Contingencies	26-30 37-38 45-46	30-38	23-32	22-23	25-27	23-24	42-49	42-49
DNS (MW)	0	27.84	48.72	121.12	3376.97	3376.97	4037.64	5425.05

5.3.2. The Swiss power grid test case

5.3.2.1. Criticality ranking results

The criticality ranking results for the Swiss power system are produced with a single execution of the cascading failure model. With a list of 1000 contingency sets per loading condition and 18 loading conditions, the total number of executed cascading simulations in a single run of Cascades is 18000. Out of these contingencies, 85.1% are single branch, 13.3% are double branch, 1.5% are triple branch failures, and the remaining involve the failure of four branches or more. Figure 19 shows branch

criticality in the Swiss power system according to a) branch impact on DNS, and b) frequency of line overload.

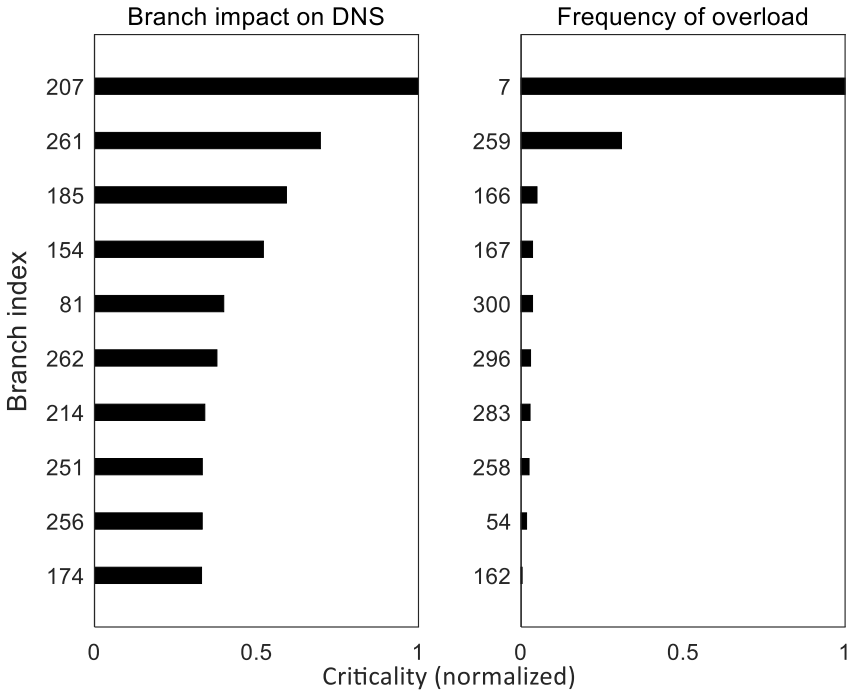


Figure 19. Branch criticality in the Swiss power system according to a) branch impact on DNS, and b) frequency of line overload.

Figure 19 shows that Line 207 is the most critical component according to the branch impact on DNS criticality measure. Line 207 is an interconnector to a neighboring country that after failure during hours with high exports is causing an overload in Line 7, which is another interconnector to the same neighboring country. In most case, there are no implications of such failure to the Swiss power grid. However, in some loading conditions/exports it causes voltage violation at single Swiss node, which is resolved by the UVLS. The failure of Lines 261 and 185 is causing similar effect to the power grid. In all these cases Line 7 overloads, which is the reason for becoming the most critical component according to the frequency of line overload criticality measure.

5.3.2.2. Most critical sets of contingences

To identify the sets of critical branches in the Swiss power system we run the vulnerability identification procedure described in Section 5.1.2. The set size spans from two to seven contingencies, such that the vulnerability identification is performed independently for each of them. To perform the analyses, we generate a random population of N_p contingencies for the designated set size, e.g., for a set size of five we randomly generated 1024 samples of two simultaneous branch failures. Figure 20 and Figure 21 show the average and maximum DNS per contingency set obtained using the average DNS based and maximum DNS based objective functions, respectively. Both figures show that up to three simultaneous failures the Swiss power grid does not encounter significant blackouts. The analysis show that the recorded DNS is a result of direct disconnection of loads in the system. However, for more than three simulations failures the algorithm identifies sets of contingencies that can cause cascades with high DNS. Furthermore, the results given in Figure 20 and Figure 21 show that there can be significant difference between the average and the maximum DNS, especially in the case when the maximum DNS based objective function is used. This shows that at some loading conditions a contingency set can cause severe blackouts while in others the same set will have little or no effect. Furthermore, Figure 20 shows a steady growth of the average DNS from four to seven contingencies. This is not the case when the critical contingency sets are obtained with the maximum DNS based

objective function, i.e., Figure 21 shows that the DNS is identical for the five, six, and seven most critical sets of contingencies.

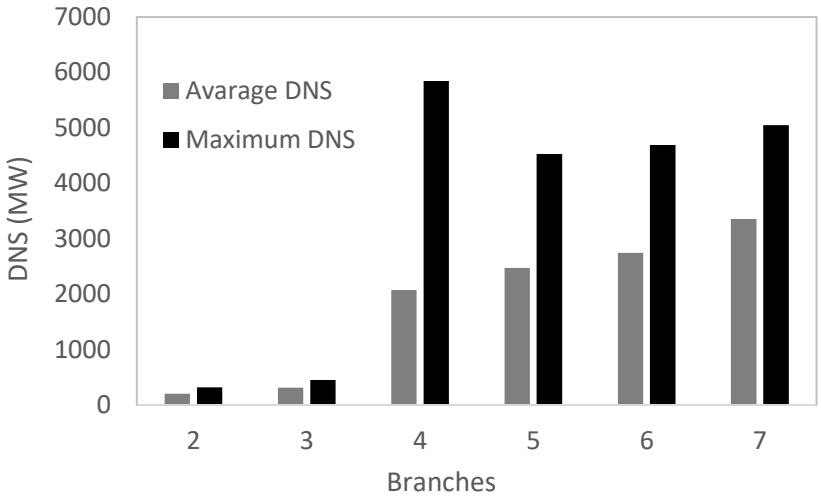


Figure 20. The DNS as function of the sets of critical contingencies obtained using the average DNS based objective function (Eq. 23).

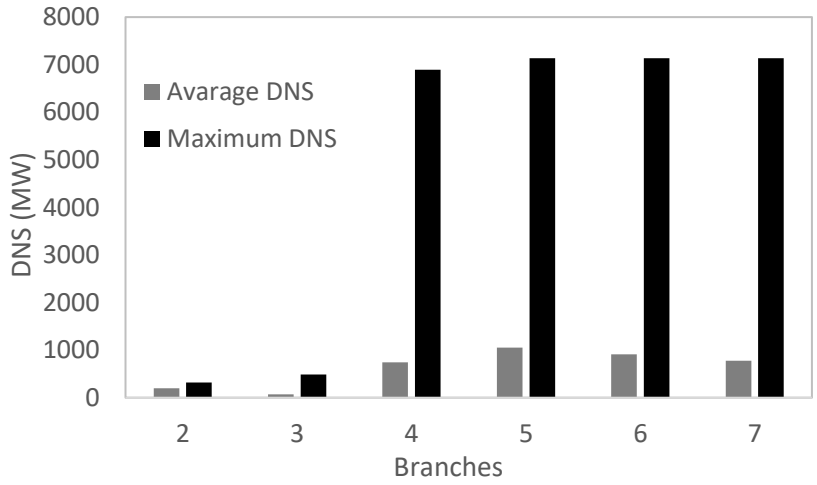


Figure 21. The DNS as function of the sets of critical contingencies obtained using the maximum DNS based objective function (Eq. 24).

6 CFS model calibration

We propose a generic framework for the calibration and validation of cascading failure analyses using meta-heuristic algorithm. The cascading failure analysis involves several parameters, i.e., line tripping threshold α, and random component failure probability p, whose values significantly affect the cascading outages propagation and its consequences. Therefore, it is critical that these parameters are calibrated to capture the statistics of cascading outages for the system to be assessed. Calibration can be based on expert elicitation of the model parameters, followed by checking the agreement between the simulation results and the historical data. In this work, we follow a systematic procedure and apply GA with the objective of minimizing the difference between the statistics of the real and simulated cascades for calibrating the appropriate values of the parameters, which produce cascading outages that are consistent with the cascading outage statistics. Metaheuristic is applied because the objective functions are highly irregular due to the discrete and threshold-driven dynamics of cascading outage model.

6.1. Modeling features

In particular, the calibration minimizes the function difference between the available historical blackout data and simulation results. The function difference methods elaborated in Section 4.2.1 are used here as well.

The cascading failure model calibration is defined as a multi-objective optimization problem where M objective functions are simultaneously minimized. M corresponds to the types of blackout data, i.e., number of customers affected, DNS, number of outage branches. If the statistics of DNS and number of tripped lines are used, the two objectives are:

$$\min\left\{dif_E^{DNS}(\alpha,p), dif_E^{outages}(\alpha,p)\right\} \tag{26}$$

where dif_E^{DNS} is the function difference between the CCDF's of the historical DNS and the DNS obtained with the cascading failure simulation model for model parameters α and p, and $dif_E^{outages}$ is the function difference between the PMF's of the historical branch failure data and the branch failure results obtained with the cascading failure simulation model for model parameters α and p. The NSGAII is selected due to its capability to obtain multiple Pareto optimal solutions in a single run [89]. The NSGAII based calibration procedure is illustrated in Figure 22.

The calibration procedure starts by generating the initial population with $2xN_p$ size:

$$\text{Pop} = \begin{bmatrix} \alpha_1 & \dots & \alpha_i & \dots & \alpha_{N_p} \\ p_1 & \dots & p_i & \dots & p_{N_p} \end{bmatrix}, \quad i = 1, 2, 3, \dots, N \tag{27}$$

where $[\alpha_i, p_i]$ represent one chromosome, i.e., one combination of model parameters. The initial population is generated by random sampling from uniform distributions, i.e. $\alpha_i \in \left(\alpha^{min}, \alpha^{max}\right)$ and $p_i \in \left(p^{min}, p^{max}\right)$.

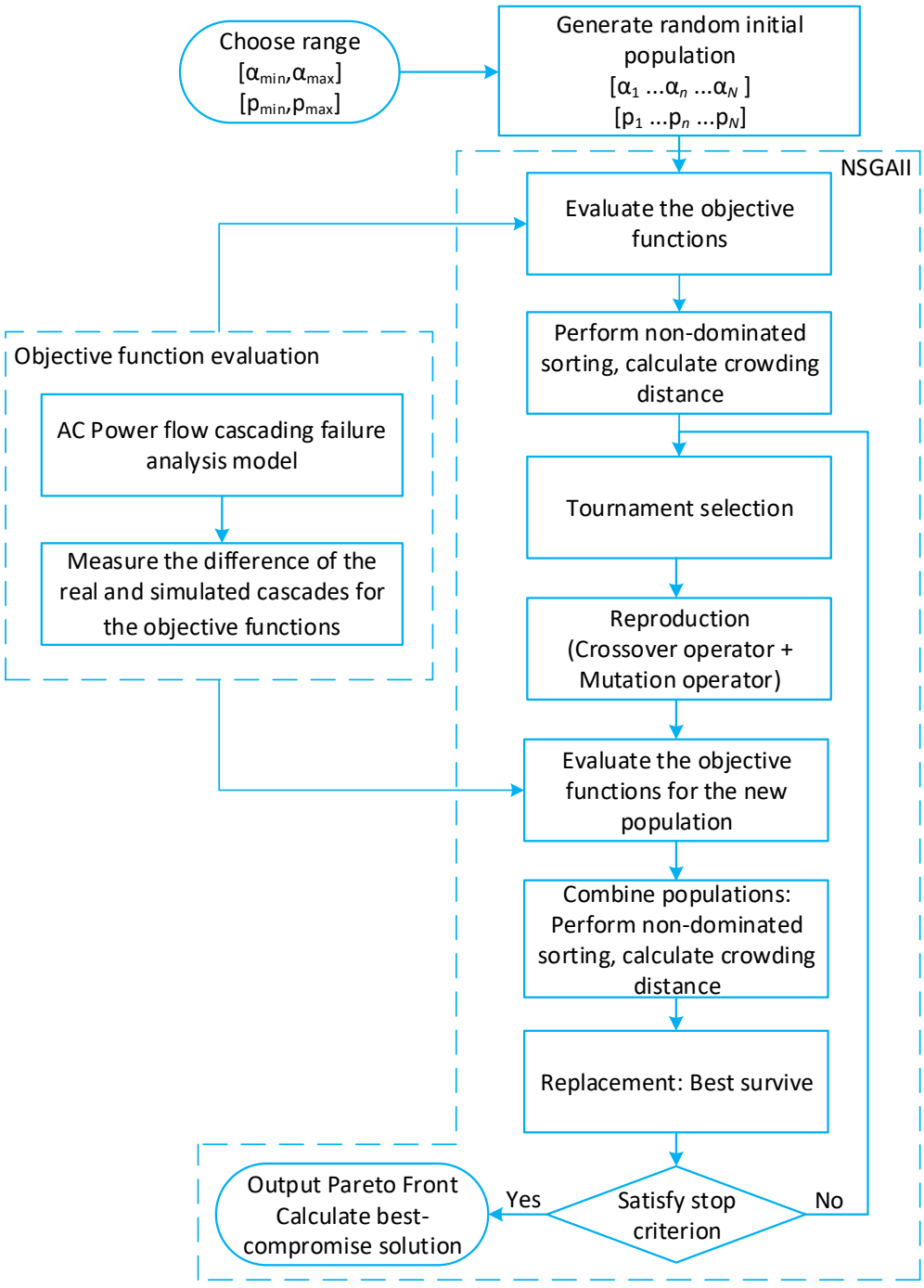


Figure 22. Flowchart of the NSGAII based calibration method.

The objective functions in Eq. (26) are evaluated for each of the randomly generated solutions, and their values are used to sort the population by applying the concept of non-dominance as described in Section 4.1.2. If a solution outperforms all the other solutions with respect to all objective functions, this solution is non-dominated or non-inferior [90]. Solutions are ranked based on their non-dominance and the population is sorted into fronts. The non-dominated solutions rank highest, and thus belong to the first front. A solution is assigned to front $F + 1$, if it is dominated by at least one solution belonging to front F. The solutions belonging to the first front are Pareto optimal solutions. Besides the non-dominant sorting procedure, the NSGAII uses the crowding distance operator to help in maintaining the diversity of the population. The crowding distance estimates the density of solutions surrounding a particular solution $[\alpha_i, p_i]$ by calculating the average distance between $[\alpha_i, p_i]$ and the two closest solutions [91]. The tournament selection operator is utilized to select the chromosomes

(solutions) for reproduction. The selected chromosomes enter the reproduction step, integrating the blending crossover operator (BLX-α) and the non-uniform mutation operator. In the next step, the non-dominated sorting and the crowding distance operators are applied to the new and the old populations. Only half of the combined population survives, according to the concept of survival of the fittest solutions, which is based on their non-dominance rank and the crowding distance. Solutions with high rank and large crowding distance have higher survival chances [89]. The procedure is repeated until the stopping criteria, i.e., the predefined number of iterations is reached. The best compromise solution (BCS) is calculated using the fuzzy set based selection procedure elaborated in Section 4.1.2.

6.2. Algorithm features

The user can select to perform the optimization using, only the line rating, α, only the failure probability, p, or both as described above. Furthermore, the user can choose between multi-objective optimization and single objective optimization. In the latter case, only the function difference, dif_E^{DNS}, between the CCDF's of the historical blackout data and the simulation results is minimized. For single objective optimization the user can selected between GA and Jaya.

7 Cascades platform simulations options

The cascading failure model can be utilized in two modes: (1) stand-alone mode, in which the module runs on its own and executes the analyses based on already existing data, and (2) interfaced mode, in which the cascading failure simulation model uses an interface with the market model, which supplies the generation dispatch and reserves scheduling. Both modes of the cascading failure model can be utilized with any of the methods (TEP, Vulnerability, Calibration) available in the Cascades platform.

Furthermore, the user can select the number of demand hours picked from the yearly demand curve and the number of contingencies per demand hour. Cascades has the option to select random hours from the load demand curve as well as different representative demand hours, e.g., number of peak loads from each season, and a combination of peak loads and random hours for each season. As explained in the N-K contingencies section, the number of contingencies is randomly selected based on the component failure probability. However, the algorithm also has the option to save contingency lists and reload them every time Cascades is executed. This option is valuable when one needs to compare the results from different cascading failure simulations of the same system and cannot afford to use a large number of simulations/contingencies. The number of contingencies and demand hours needed to thoroughly explore the system risk depends on the system size. For systems of the order of the 24-bus IEEE test system, the number of contingencies should be higher than 1000, and the number of demand hours at least 18.

The NSGAII and GA have multiple strategies and operators available. In particular, there are two reproduction strategies: (1) the new population is created in two shares, the first one is created with crossover and the second one with mutation, such the latter share is considerably small; (2) the new population is created solely with crossover, and then the mutation operator is applied on the new population (part or all), however the probability to select a variable for crossover is kept small. Two selection operators exist, tournament and rank selection. Two crossover operators exist, blending crossover operator (BLX-α) and two-point crossover. Three mutation operators exist, uniform mutation, non-uniform mutation, and new solution at random. Two replacement strategies exist, elitist (discard all the worst solutions), and semi-elitist (if some of the solution in the best population are repeating, keep one, and replace the rest with the top solutions from the worst population). This replacement strategy is also implemented in the Jaya algorithm.

Cascades can result in large computing times; to increase the processing speed the user can select a parallel execution option. Parallelization is performed on three levels, on the: 1) contingencies, 2) simulation hours, and 3) population (only when optimization is performed using one of the methods). Only one of these options can be selected since they are nested. When only CFS is performed, and large number of cores are available (>18) use option 1. In case the number of cores is smaller, use option 2). If any of the available parallelization options is selected, the algorithm distributes the work automatically to the available CPU cores or to the number of preselected cores (this setting is available in the MATLAB parallelization toolbox).

8 Cascades input data

The Cascades data comprise, power grid data, generator data, and load data. The platform is coded in Matlab and the data set is in Matpower format (see here). This format presents the system as an mpc Matpower case structure (mpc), consisting of matrices for the buses, generators, branches, and the generator cost in the system. The topology of the system is needed; therefore, every branch should have connecting buses and generators should be assigned to a bus. The necessary variables are presented in Table 13. Additional data that is desirable but not necessary is listed in Table 14. The data can be acquired in any format, and then organized into the Matpower's mpc structure. The model can pull the data either form saved .mat files or from a MySQL database.

Table 13. Necessary information for the Cascades platform

Variable	Resolution	Unit	Description
mpc base MVA	-	MVA	Base MVA assumed
mpc bus data	by bus	various	Bus data in format (Figure 23)
mpc branch data	by branch	various	Line data in format (Figure 24)
mpc gen data	by unit	various	Generator data in format (Figure 25)
mpc gen cost data	by unit	various	Generator cost data in format (Figure 26)
mpc gen info data	by unit	various	Additional generator information
Nominal frequency f_0	-	Hz	Nominal frequency of the system
Load realization	hourly, nodal	MW	Nodal power demand in each hour
Non-dispatchable units output	hourly, nodal	MW	Time series generation of non-dispatchable units at each node of the grid

Table 14. Desirable but not necessary

Variable		Resolution	Unit	Description
Generation dispatch		hourly, nodal	MW	Generator power injections in each hour (*)
Positive/Negative procurements	FRRa	hourly, nodal	MW	Generator power injections in each hour (*)
Positive/Negative procurements	FRRm	hourly, nodal	MW	Generator power injections in each hour (*)
Additional variables				
$\boldsymbol{\lambda_{area}}$		-	MW/Hz	Frequency characteristic of the sync. area (*)
$\boldsymbol{D_{reg}}$		-	%/Hz	Self-regulating characteristic of load (*)
UFLS		-	-	Under-frequency load shedding procedure

(*) – will be calculated by the model if not explicitly stated

Input data clarifications

- The Generator positions (dispatch) if not available is calculated in Cascades as an optimal power flow (OPF). However, this approach does not accurately reflect the market mechanisms. Therefore, either a dispatch provided by a market model, or a historical year of load and the realized dispatch is preferable.
- The hourly nodal load realization, if not available, is calculated in Cascades using the yearly load curve of the system and a snapshot of the load on each bus.
- The secondary (FRRa) and tertiary (FRRm) positive (negative) reserves can be explicitly stated, or the model will calculate them as the difference between dispatched power Pg and maximum (minimum) power of the generator Pmax (Pmin).

- The frequency characteristic of the synchronous area λ_{area} can be explicitly stated, or the model will calculate it as:

$$\lambda_{area} = \frac{0.05 * f_0}{\sum P_{maxDI}}$$

- where P_{maxDI} refers to the maximum power output of dispatchable units (e.g., thermal power plants).
- If the self-regulating characteristic of the load D_{reg} is not specified, the model will not consider this parameter.

Table B-1: Bus Data (`mpc.bus`)

name	column	description
BUS_I	1	bus number (positive integer)
BUS_TYPE	2	bus type (1 = PQ, 2 = PV, 3 = ref, 4 = isolated)
PD	3	real power demand (MW)
QD	4	reactive power demand (MVAr)
GS	5	shunt conductance (MW demanded at $V = 1.0$ p.u.)
BS	6	shunt susceptance (MVAr injected at $V = 1.0$ p.u.)
BUS_AREA	7	area number (positive integer)
VM	8	voltage magnitude (p.u.)
VA	9	voltage angle (degrees)
BASE_KV	10	base voltage (kV)
ZONE	11	loss zone (positive integer)
VMAX	12	maximum voltage magnitude (p.u.)
VMIN	13	minimum voltage magnitude (p.u.)
LAM_P†	14	Lagrange multiplier on real power mismatch (u/MW)
LAM_Q†	15	Lagrange multiplier on reactive power mismatch (u/MVAr)
MU_VMAX†	16	Kuhn-Tucker multiplier on upper voltage limit (u/p.u.)
MU_VMIN†	17	Kuhn-Tucker multiplier on lower voltage limit (u/p.u.)

† Included in OPF output, typically not included (or ignored) in input matrix. Here we assume the objective function has units u.

Figure 23. Data format of the bus structure. The parameters highlighted in red are necessary for the model to work properly.

Table B-3: Branch Data (`mpc.branch`)

name	column	description
F_BUS	1	"from" bus number
T_BUS	2	"to" bus number
BR_R	3	resistance (p.u.)
BR_X	4	reactance (p.u.)
BR_B	5	total line charging susceptance (p.u.)
RATE_A	6	MVA rating A (long term rating), set to 0 for unlimited
RATE_B	7	MVA rating B (short term rating), set to 0 for unlimited
RATE_C	8	MVA rating C (emergency rating), set to 0 for unlimited
TAP	9	transformer off nominal turns ratio, (taps at "from" bus, impedance at "to" bus, i.e. if $r = x = b = 0$, $tap = \frac{\lvert V_f \rvert}{\lvert V_t \rvert}$)
SHIFT	10	transformer phase shift angle (degrees), positive ⇒ delay
BR_STATUS	11	initial branch status, 1 = in-service, 0 = out-of-service
ANGMIN*	12	minimum angle difference, $\theta_f - \theta_t$ (degrees)
ANGMAX*	13	maximum angle difference, $\theta_f - \theta_t$ (degrees)
PF†	14	real power injected at "from" bus end (MW)
QF†	15	reactive power injected at "from" bus end (MVAr)
PT†	16	real power injected at "to" bus end (MW)
QT†	17	reactive power injected at "to" bus end (MVAr)
MU_SF‡	18	Kuhn-Tucker multiplier on MVA limit at "from" bus (u/MVA)
MU_ST‡	19	Kuhn-Tucker multiplier on MVA limit at "to" bus (u/MVA)
MU_ANGMIN‡	20	Kuhn-Tucker multiplier lower angle difference limit (u/degree)
MU_ANGMAX‡	21	Kuhn-Tucker multiplier upper angle difference limit (u/degree)

* Not included in version 1 case format. The voltage angle difference is taken to be unbounded below if ANGMIN < -360 and unbounded above if ANGMAX > 360. If both parameters are zero, the voltage angle difference is unconstrained.
† Included in power flow and OPF output, ignored on input.
‡ Included in OPF output, typically not included (or ignored) in input matrix. Here we assume the objective function has units u.

Figure 24. Data format of the branch structure. The parameters highlighted in red are necessary for the model to work properly.

Table B-2: Generator Data (mpc.gen)

name	column	description
GEN_BUS	1	bus number
PG	2	real power output (MW)
QG	3	reactive power output (MVAr)
QMAX	4	maximum reactive power output (MVAr)
QMIN	5	minimum reactive power output (MVAr)
VG	6	voltage magnitude setpoint (p.u.)
MBASE	7	total MVA base of machine, defaults to baseMVA
GEN_STATUS	8	machine status, > 0 = machine in-service; ≤ 0 = machine out-of-service
PMAX	9	maximum real power output (MW)
PMIN	10	minimum real power output (MW)
PC1*	11	lower real power output of PQ capability curve (MW)
PC2*	12	upper real power output of PQ capability curve (MW)
QC1MIN*	13	minimum reactive power output at PC1 (MVAr)
QC1MAX*	14	maximum reactive power output at PC1 (MVAr)
QC2MIN*	15	minimum reactive power output at PC2 (MVAr)
QC2MAX*	16	maximum reactive power output at PC2 (MVAr)
RAMP_AGC*	17	ramp rate for load following/AGC (MW/min)
RAMP_10*	18	ramp rate for 10 minute reserves (MW)
RAMP_30*	19	ramp rate for 30 minute reserves (MW)
RAMP_Q*	20	ramp rate for reactive power (2 sec timescale) (MVAr/min)
APF*	21	area participation factor
MU_PMAX†	22	Kuhn-Tucker multiplier on upper P_g limit (u/MW)
MU_PMIN†	23	Kuhn-Tucker multiplier on lower P_g limit (u/MW)
MU_QMAX†	24	Kuhn-Tucker multiplier on upper Q_g limit (u/MVAr)
MU_QMIN†	25	Kuhn-Tucker multiplier on lower Q_g limit (u/MVAr)

* Not included in version 1 case format.
† Included in OPF output, typically not included (or ignored) in input matrix. Here we assume the objective function has units u.

Figure 25. Data format of the generator structure. The parameters highlighted in red are necessary for the model to work properly.

Table B-4: Generator Cost Data† (mpc.gencost)

name	column	description
MODEL	1	cost model, 1 = piecewise linear, 2 = polynomial
STARTUP	2	startup cost in US dollars*
SHUTDOWN	3	shutdown cost in US dollars*
NCOST	4	number of cost coefficients for polynomial cost function, or number of data points for piecewise linear
COST	5	parameters defining total cost function $f(p)$ begin in this column, units of f and p are \$/hr and MW (or MVAr), respectively (MODEL = 1) $\Rightarrow$ $p_0, f_0, p_1, f_1, \ldots, p_n, f_n$ where $p_0 < p_1 < \cdots < p_n$ and the cost $f(p)$ is defined by the coordinates (p_0, f_0), (p_1, f_1), $\ldots$, (p_n, f_n) of the end/break-points of the piecewise linear cost (MODEL = 2) $\Rightarrow$ $c_n, \ldots, c_1, c_0$ $n+1$ coefficients of n-th order polynomial cost, starting with highest order, where cost is $f(p) = c_n p^n + \cdots + c_1 p + c_0$

† If gen has n_g rows, then the first n_g rows of gencost contain the costs for active power produced by the corresponding generators. If gencost has $2n_g$ rows, then rows $n_g + 1$ through $2n_g$ contain the reactive power costs in the same format.
* Not currently used by any MATPOWER functions.

Figure 26. Data format of the generator cost structure. This information is not necessary, however, specifying the cost can lead to better results with the optimal power flow solver.

References

1. Čepin, M., *Assessment of Power System Reliability: Methods and Applications*. 2011, London: Springer-Verlag.
2. Haes Alhelou, H., et al., *A Survey on Power System Blackout and Cascading Events: Research Motivations and Challenges.* Energies, 2019. **12**(4): p. 682.
3. Andersson, G., et al., *Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance.* IEEE Transactions on Power Systems, 2005. **20**(4): p. 1922-1928.
4. ENTSO-E, *System Disturbance on 4 November 2006, Final Report*, in 2007, European Network of Transmission System Operators for Electricity: Brussels, Belgium.
5. ENTSO-E, *Continental Europe Synchronous Area Separation on 08 January 2021, Final Report*. 2021, European Network of Transmission System Operators for Electricity: Brussels, Belgium.
6. ENTSO-E, *Continental Europe Synchronous Area Separation on 24 July 2021, Technical Report*. 2021, European Network of Transmission System Operators for Electricity: Brussels, Belgium.
7. ENTSO-E, *Final Report of the Investigation Committee on the 28 September 2003 Blackout in Italy*. 2004, European Network of Transmission System Operators for Electricity: Brussels, Belgium.
8. ENTSO-E, *Report on Blackout in Turkey on 31st March 2015, Project Group Turkey*. 2015, European Network of Transmission System Operators for Electricity: Brussels, Belgium.
9. Beyza, J. and J.M. Yusta, *The effects of the high penetration of renewable energies on the reliability and vulnerability of interconnected electric power systems.* Reliability Engineering & System Safety, 2021. **215**: p. 107881.
10. Vrakopoulou, M., et al., *A Probabilistic Framework for Reserve Scheduling and N-1 Security Assessment of Systems With High Wind Power Penetration.* IEEE Transactions on Power Systems, 2013. **28**(4): p. 3885-3896.
11. Qiming, C. and J.D. McCalley, *Identifying high risk N-k contingencies for online security assessment.* IEEE Transactions on Power Systems, 2005. **20**(2): p. 823-834.
12. Guo, H., et al., *A critical review of cascading failure analysis and modeling of power system.* Renewable and Sustainable Energy Reviews, 2017. **80**: p. 9-22.
13. NERC, *Standard TPL-001-4 — Transmission System Planning Performance Requirements*. 2015, North American Electric Reliability Corporation (NERC): Atlanta, Georgia, USA.
14. Mitra, P., et al., *A Systematic Approach to n-1-1 Analysis for Power System Security Assessment.* IEEE Power and Energy Technology Systems Journal, 2016. **3**(2): p. 71-80.
15. Kröger, W. and E. Zio, *Vulnerable Systems*. 2011: Springer Publishing Company, Incorporated.
16. Abedi, A., L. Gaudard, and F. Romerio, *Review of major approaches to analyze vulnerability in power system.* Reliability Engineering & System Safety, 2019. **183**: p. 153-172.
17. Dobson, I., et al. *An initial model fo complex dynamics in electric power system blackouts*. in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. 2001.
18. Carreras, B.A., et al., *Critical points and transitions in an electric power transmission model for cascading failure blackouts.* Chaos: An Interdisciplinary Journal of Nonlinear Science, 2002. **12**(4): p. 985-994.
19. Chen, J., J.S. Thorp, and I. Dobson, *Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model.* International Journal of Electrical Power & Energy Systems, 2005. **27**(4): p. 318-326.
20. Zhang, H., et al., *Identifying critical risks of cascading failures in power systems.* IET Generation, Transmission & Distribution, 2019. **13**(12): p. 2438-2445.
21. Mehdizadeh, M., R. Ghazi, and M. Ghayeni, *New sequential approach based on graph traversal algorithm to investigate cascading outages considering correlated wind farms.* IET Generation, Transmission & Distribution, 2019. **13**(12): p. 2401-2410.
22. Nedic, D.P., et al., *Criticality in a cascading failure blackout model.* International Journal of Electrical Power & Energy Systems, 2006. **28**(9): p. 627-633.

23. Noebels, M., R. Preece, and M. Panteli, *AC Cascading Failure Model for Resilience Analysis in Power Networks.* IEEE Systems Journal, 2020: p. 1-12.
24. Salim, N.A., et al., *Determination of available transfer capability with implication of cascading collapse uncertainty.* IET Generation, Transmission & Distribution, 2014. **8**(4): p. 705-715.
25. Bing, L., et al. *Linear implicit AC PF cascading failure analysis with power system operations and automation*. in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. 2016.
26. Kamali, S., T. Amraee, and M. Khorsand, *Intentional power system islanding under cascading outages using energy function method.* IET Generation, Transmission & Distribution, 2020. **14**(20): p. 4553-4562.
27. Liu, Y., T. Wang, and X. Gu, *A risk-based multi-step corrective control method for mitigation of cascading failures.* IET Generation, Transmission & Distribution, 2022. **16**(4): p. 766-775.
28. Hoseinzadeh, B., et al., *Emergency wind power plant re-dispatching against transmission system cascading failures using reverse tracking of line power flow.* IET Generation, Transmission & Distribution, 2020. **14**(16): p. 3241-3249.
29. Fan, W.-L., et al., *Vulnerability analysis of power system by modified H-index method on cascading failure state transition graph.* Electric Power Systems Research, 2022. **209**: p. 107986.
30. Liu, H., et al., *A sequentially preventive model enhancing power system resilience against extreme-weather-triggered failures.* Renewable and Sustainable Energy Reviews, 2022. **156**: p. 111945.
31. Qorbani, M. and T. Amraee, *Long term transmission expansion planning to improve power system resilience against cascading outages.* Electric Power Systems Research, 2021. **192**: p. 106972.
32. Zhang, X., et al., *An integrated modeling framework for cascading failure study and robustness assessment of cyber-coupled power grids.* Reliability Engineering & System Safety, 2022. **226**: p. 108654.
33. David, A.E., B. Gjorgiev, and G. Sansavini, *Quantitative comparison of cascading failure models for risk-based decision making in power systems.* Reliability Engineering & System Safety, 2020: p. 106877.
34. Henneaux, P., et al. *Benchmarking Quasi-Steady State Cascading Outage Analysis Methodologies*. in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. 2018.
35. Bialek, J., et al., *Benchmarking and Validation of Cascading Failure Analysis Tools.* IEEE Transactions on Power Systems, 2016. **31**(6): p. 4887-4900.
36. ENTSO-E, *Policy 1: Load-Frequency Control and Performance*. 2009.
37. EuropeanCommission, *COMMISSION REGULATION (EU) 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation.* Official Journal of the European Union, 2017.
38. Alizadeh Mousavi, O., et al., *Inter-area frequency control reserve assessment regarding dynamics of cascading outages and blackouts.* Electric Power Systems Research, 2014. **107**: p. 144-152.
39. Swissgrid, *Transmission Code 2013*. 2014.
40. Kirschen, D.S., et al., *Computing the value of security.* IEE Proceedings - Generation, Transmission and Distribution, 2003. **150**(6): p. 673-678.
41. Zimmerman, R.D., C.E. Murillo-Sanchez, and R.J. Thomas, *MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education.* IEEE Transactions on Power Systems, 2011. **26**(1): p. 12-19.
42. Taylor, C.W., *Concepts of undervoltage load shedding for voltage stability.* IEEE Transactions on Power Delivery, 1992. **7**(2): p. 480-488.
43. Dola, H.M. and B.H. Chowdhury. *Intentional islanding and adaptive load shedding to avoid cascading outages*. in *2006 IEEE Power Engineering Society General Meeting*. 2006.
44. Stankovski, A., B. Gjorgiev, and G. Sansavini, *Multi-zonal method for cascading failure analyses in large interconnected power systems.* IET Generation, Transmission & Distribution, 2022. **16**(20): p. 4040-4053.

45. ENTSO-E, *Policy 5: Emergency Operations*. 2015.
46. ENTSO-E, *Continental Europe Synchronous Area Separation on 24 July 2021* in *ENTSO-E*. 2021.
47. AEMO, *Preliminary Report – Victoria and South Australia Separation Event, 31 January 2020*. 2020.
48. ENTSO-E, *Continental Europe Synchronous Area Separation on 08 January 2021*. 2021.
49. Li, B., B. Gjorgiev, and G. Sansavini. *Meta-heuristic approach for validation and calibration of cascading failure analysis*. in *2018 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2018 - Proceedings*. 2018.
50. Gjorgiev, B., B. Li, and G. Sansavini. *Calibration of Cascading Failure Simulation Models for Power System Risk Assessment*. in *Safety and Reliability - Safe Societies in a Changing World - Proceedings of the 28th International European Safety and Reliability Conference, ESREL 2019*. 2019. Germany.
51. Li, B., B. Gjorgiev, and G. Sansavini. *A genetic algorithm based calibration approach on validating cascading failure analysis*. in *2017 IEEE Power & Energy Society General Meeting*. 2017.
52. Price, J.E. and J. Goodin. *Reduced network modeling of WECC as a market design prototype*. in *2011 IEEE Power and Energy Society General Meeting*. 2011. IEEE.
53. North American Electric Reliability Council (NERC), P., New Jersey USA 08540-5731. *Information on Blackouts in North America, Disturbance Analysis Working Group (DAWG) Database*. Available from: http://www.nerc.com/~dawg/database.html.
54. Knapp, S.R., *Pumped storage the handmaiden of nuclear power.* IEEE Spectrum, 1969. **6**(4): p. 46-52.
55. Dobson, I., *Estimating the propagation and extent of cascading line outages from utility data with a branching process.* IEEE Transactions on Power Systems, 2012. **27**(4): p. 2146-2155.
56. Gomes, P.V. and J.T. Saraiva, *A novel efficient method for multiyear multiobjective dynamic transmission system planning.* Electrical Power and Energy Systems, 2018. **100**: p. 10-18.
57. Wu, Z., et al., *Automatic Selection Method for Candidate Lines in Transmission Expansion Planning.* IEEE Access, 2018. **6**: p. 11605-11613.
58. Verma, S. and V. Mukherjee, *Investigation of static transmission expansion planning using the symbiotic organisms search algorithm.* Engineering Optimization, 2018. **50**(9): p. 1544-1560.
59. Vatani, B., B. Chowdhury, and J. Lin, *The Role of Demand Response as an Alternative Transmission Expansion Solution in a Capacity Market.* IEEE Transactions on Industry Applications, 2018. **54**(2): p. 1039-1046.
60. Zhang, H., et al., *A Mixed-Integer Linear Programming Approach for Multi-Stage Security-Constrained Transmission Expansion Planning.* IEEE Transactions on Power Systems, 2012. **27**(2): p. 1125-1133.
61. Baharvandi, A., et al., *Bundled Generation and Transmission Planning Under Demand and Wind Generation Uncertainty Based on a Combination of Robust and Stochastic Optimization.* IEEE Transactions on Sustainable Energy, 2018. **9**(3): p. 1477-1486.
62. Orfanos, G.A., P.S. Georgilakis, and N.D. Hatziargyriou, *Transmission Expansion Planning of Systems With Increasing Wind Power Integration.* IEEE Transactions on Power Systems, 2013. **28**(2): p. 1355-1362.
63. Qiu, J., et al., *Power network planning considering trade-off between cost, risk, and reliability.* International Transactions on Electrical Energy Systems, 2017. **27**(12).
64. Kayal, P. and C.K. Chanda, *Optimal mix of solar and wind distributed generations considering performance improvement of electrical distribution network.* Renewable Energy, 2015. **75**: p. 173-186.
65. Lumbreras, S., A. Ramos, and F. Banez-Chicharro, *Optimal transmission network expansion planning in real-sized power systems with high renewable penetration.* Electric Power Systems Research, 2017. **149**: p. 76-88.
66. Taherkhani, M., et al., *Scenario-based probabilistic multi-stage optimization for transmission expansion planning incorporating wind generation integration.* Electric Power Systems Research, 2020. **189**: p. 106601.

67. Zeinaddini-Meymand, M., et al., *A Demand-Side Management-Based Model for G&TEP Problem Considering FSC Allocation.* IEEE Systems Journal, 2019. **13**(3): p. 3242-3253.
68. Baringo, L. and A. Baringo, *A Stochastic Adaptive Robust Optimization Approach for the Generation and Transmission Expansion Planning.* IEEE Transactions on Power Systems, 2018. **33**(1): p. 792-802.
69. Gomes, P.V. and J.T. Saraiva, *State-of-the-art of transmission expansion planning: A survey from restructuring to renewable and distributed electricity markets.* Electrical Power and Energy Systems, 2019. **111**: p. 411-424.
70. Glover, J.D., M.S. Sarma, and T.J. Overbye, *Power Systems Analysis and Design*. 1987, Boston, MA: PWS Publishing Co.
71. Overbye, T.J., X. Cheng, and Y. Sun. *A Comparison of the AC and DC Power Flow Models for LMP Calculations*. in *Proceedings of the 37th Hawaii International Conference on System Sciences*. 2004. Big Island, HI.
72. Kile, H., et al. *A comparison of AC and DC power flow models for contingency and reliability analysis*. in *2014 Power Systems Computation Conference*. 2014. Wroclaw.
73. Bent, R., et al. *Transmission Network Expansion Planning: Bridging the gap between AC heuristics and DC approximations*. in *Power Systems Computation Conference*. 2014. Wroclaw, PL.
74. Coffrin, C. and P. van Hentenryck, *A Linear-Programming Approximation of AC Power Flows.* Informs Journal on Computing, 2012. **26**(4).
75. Wu, Z., et al., *Automatic Selection Method for Candidate Lines in Transmission Expansion Planning.* IEEE Access, 2018. **6**: p. 11605-11613.
76. Zhang, Y., et al., *A Mixed-Integer Linear Programming Approach to Security-Constrained Co-Optimization Expansion Planning of Natural Gas and Electricity Transmission Systems.* IEEE Transactions on Power Systems, 2018. **33**(6): p. 6368-6378.
77. Rahmani, M., R. Romero, and M.J. Rider, *Strategies to Reduce the Number of Variables and the Combinatorial Search Space of the Multistage Transmission Expansion Planning Problem.* IEEE Transactions on Power Systems, 2013. **28**(3): p. 2164-2173.
78. Braga, A.S.D. and J.T. Saraiva, *A multiyear dynamic approach for transmission expansion planning and long-term marginal costs computation.* IEEE Transactions on Power Systems, 2005. **20**(3): p. 1631-1639.
79. Qiu, J., et al., *A Risk-Based Approach to Multi-Stage Probabilistic Transmission Network Planning.* IEEE Transactions on Power Systems, 2016. **31**(6): p. 4867-4876.
80. Karimi, E. and A. Ebrahimi, *Considering risk of cascading line outages in transmission expansion planning by benefit/cost analysis.* Electrical Power and Energy Systems, 2016. **78**: p. 480-488.
81. Zheng, J., et al., *Risk control in transmission system expansion planning with wind generators.* International Transactions on Electrical Energy Systems, 2014. **24**: p. 227-245.
82. Gjorgiev, B., A.E. David, and G. Sansavini, *Cascade-risk-informed transmission expansion planning of AC electric power systems.* Electric Power Systems Research, 2022. **204**: p. 107685.
83. Shahid, R., et al., *Comparison of distance measures in spatial analytical modeling for health service planning.* BMC health services research, 2009. **9**: p. 200-200.
84. Deb, K., *Multi-Objective Optimization using Evolutionary Algorithms*. 2001, Chichester: Jhon Wiley & Sons. 515.
85. Digabel, S.L., *Algorithm 909: NOMAD: Nonlinear Optimization with the MADS Algorithm.* ACM Trans. Math. Softw., 2011. **37**(4): p. 1-15.
86. Audet, C. and J.E. Dennis, *Mesh Adaptive Direct Search Algorithms for Constrained Optimization.* SIAM J. on Optimization, 2006. **17**(1): p. 188-217.
87. Vaz, A.I.F. and L.N. Vicente, *A particle swarm pattern search method for bound constrained global optimization.* J Glob Optim, 2007. **39**(197).
88. Arabali, A., et al., *A Multi-Objective Transmission Expansion Planning Framework in Deregulated Power Systems With Wind Generation.* IEEE Transactions on Power Systems, 2014. **29**(6): p. 3003-3011.

89. Deb, K., et al., *A fast and elitist multiobjective genetic algorithm: NSGA-II.* IEEE transactions on evolutionary computation, 2002. **6**(2): p. 182-197.
90. Srinivas, N. and K. Deb, *Muiltiobjective optimization using nondominated sorting in genetic algorithms.* Evolutionary computation, 1994. **2**(3): p. 221-248.
91. Deb, K., *Multi-objective optimization using evolutionary algorithms*. Vol. 16. 2001: John Wiley & Sons.
92. Bo, Z. and C. Yi-Jia, *Multiple objective particle swarm optimization technique for economic load dispatch.* Journal of Zhejiang University-Science A, 2005. **6**(5): p. 420-427.
93. Dhillon, J., S. Parti, and D. Kothari, *Stochastic economic emission load dispatch.* Electric Power Systems Research, 1993. **26**(3): p. 179-186.
94. Reznik, L., *Fuzzy controllers handbook: how to design them, how they work*. 1997: Newnes.
95. Benabid, R., M. Boudour, and M. Abido. *Optimal placement of FACTS devices for multi-objective voltage stability problem*. in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*. 2009. IEEE.
96. IIT. *"http://motor.ece.iit.edu/Data" Illinois Institute of Technology, Electrical and Computer Engineering Department, [Online]*. 06.08.2019].
97. Peña, I., C.B. Martinez-Anido, and B. Hodge, *An Extended IEEE 118-Bus Test System With High Renewable Penetration.* IEEE Transactions on Power Systems, 2018. **33**(1): p. 281-289.
98. USEIA. *U.S. Electric System Operating Data. https://www.eia.gov/realtime_grid/#/data/graphs?end=20170108T00&start=20170101T00&regions=04*. 06.08.2019].
99. Zhang, H., et al., *An Improved Network Model for Transmission Expansion Planning Considering Reactive Power and Network Losses.* IEEE Transactions on Power Systems, 2013. **28**(3): p. 3471-3479.
100. ETHZ. *Euler - Scientific computing*. 2020 [cited 2020 03.03]; Available from: https://scicomp.ethz.ch/wiki/Euler.
101. Szekely, G.J., M.L. Rizzo, and N.K. Bakirov, *Measuring and testing dependence by correlation of distances.* Ann. Statist., 2007. **35**(6): p. 2769-2794.
102. Smirnov, N., *Table for Estimating the Goodness of Fit of Empirical Distributions.* Ann. Math. Statist., 1948. **19**(2): p. 279-281.
103. Marsaglia, G., W.W. Tsang, and J. Wang, *Evaluating Kolmogorov's Distribution.* Journal of Statistical Software, 2003. **008**(i18).
104. Kullback, S. and R.A. Leibler, *On Information and Sufficiency.* Ann. Math. Statist., 1951. **22**(1): p. 79-86.
105. Kolouri, S., et al., *Optimal Mass Transport: Signal processing and machine-learning applications.* IEEE Signal Processing Magazine, 2017. **34**(4): p. 43-59.
106. Garrison, J., et al., *Nexus-e: Scenario Results Report*. 2020, Department of the Environment, Transport, Energy and Communication DETEC, Swiss Federal Office of Energy SFOE, Energy Research and Cleantech.
107. Gjorgiev, B., et al., *Nexus-e: A platform of interfaced high-resolution models for energy-economic assessments of future electricity systems.* Applied Energy, 2022. **307**: p. 118193.
108. Murray, A.T., Grubesic, Tony, *Critical infrastructure: reliability and vulnerability*. 2007, London: Springer.
109. Weiss, M. and M. Weiss, *An assessment of threats to the American power grid.* Energy, Sustainability and Society, 2019. **9**(1): p. 18.
110. Haimes, Y.Y. and B.M. Horowitz, *Modeling Interdependent Infrastructures for Sustainable Counterterrorism.* Journal of Infrastructure Systems, 2004. **10**(2): p. 33-42.
111. Cuadra, L., et al., *A Critical Review of Robustness in Power Grids Using Complex Networks Concepts.* Energies, 2015. **8**(9): p. 9211-9265.
112. Bompard, E., L. Luo, and E. Pons, *A perspective overview of topological approaches for vulnerability analysis of power transmission grids.* International Journal of Critical Infrastructures, 2015. **11**(1): p. 15-26.

113. Zio, E., L.R. Golea, and C.M. Rocco S, *Identifying groups of critical edges in a realistic electrical network by multi-objective genetic algorithms.* Reliability Engineering & System Safety, 2012. **99**: p. 172-177.
114. Alipour, Z., M.A.S. Monfared, and E. Zio, *Comparing topological and reliability-based vulnerability analysis of Iran power transmission network.* Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2014. **228**(2): p. 139-151.
115. Eusgeld, I., et al., *The role of network theory and object-oriented modeling within a framework for the vulnerability analysis of critical infrastructures.* Reliability Engineering & System Safety, 2009. **94**(5): p. 954-963.
116. Bompard, E., R. Napoli, and F. Xue, *Analysis of structural vulnerabilities in power transmission grids.* International Journal of Critical Infrastructure Protection, 2009. **2**(1): p. 5-12.
117. Volkanovski, A., M. Čepin, and B. Mavko, *Application of the fault tree analysis for assessment of power system reliability.* Reliability Engineering & System Safety, 2009. **94**(6): p. 1116-1127.
118. Čepin, M., *Probability of restoring power to the transmission power system and the time to restore power.* Reliability Engineering & System Safety, 2020. **193**: p. 106595.
119. Tas, S. and V.M. Bier, *Addressing vulnerability to cascading failure against intelligent adversaries in power networks.* Energy Systems, 2016. **7**(2): p. 193-213.
120. Cheng, M.X., M. Crow, and Q. Ye, *A game theory approach to vulnerability analysis: Integrating power flows with topological analysis.* International Journal of Electrical Power & Energy Systems, 2016. **82**: p. 29-36.
121. Li, J., et al., *AC power flow importance measures considering multi-element failures.* Reliability Engineering & System Safety, 2017. **160**: p. 89-97.
122. Fang, Y.-P., G. Sansavini, and E. Zio, *An Optimization-Based Framework for the Identification of Vulnerabilities in Electric Power Grids Exposed to Natural Hazards.* Risk Analysis, 2019. **39**(9): p. 1949-1969.
123. Liu, X., Y.-P. Fang, and E. Zio, *A Hierarchical Resilience Enhancement Framework for Interdependent Critical Infrastructures.* Reliability Engineering & System Safety, 2021. **215**: p. 107868.
124. Abedi, A., L. Gaudard, and F. Romerio, *Power flow-based approaches to assess vulnerability, reliability, and contingency of the power systems: The benefits and limitations.* Reliability Engineering & System Safety, 2020. **201**: p. 106961.
125. Eppstein, M. and P. Hines. *A "Random Chemistry" algorithm for identifying collections of multiple contingencies that initiate cascading failure*. in *2013 IEEE Power & Energy Society General Meeting*. 2013.
126. Rocco, C.M., et al., *Assessing the Vulnerability of a Power System Through a Multiple Objective Contingency Screening Approach.* IEEE Transactions on Reliability, 2011. **60**(2): p. 394-403.
127. Donde, V., et al., *Severe Multiple Contingency Screening in Electric Power Systems.* IEEE Transactions on Power Systems, 2008. **23**(2): p. 406-417.
128. Gjorgiev, B. and G. Sansavini, *Identifying and assessing power system vulnerabilities to transmission asset outages via cascading failure analysis.* Reliability Engineering & System Safety, 2022. **217**: p. 108085.
129. Gjorgiev, B. and M. Čepin, *A multi-objective optimization based solution for the combined economic-environmental power dispatch problem.* Engineering Applications of Artificial Intelligence, 2013. **26**(1): p. 417-429.