2019

Modeling Cascading Failures in Power Systems in the Presence of Uncertain Wind Generation

Mir Hadi Athari

Modeling Cascading Failures in Power Systems in the Presence of Uncertain Wind Generation

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

By

Mir Hadi Athari

M.Sc. in Electrical Engineering

Advisor: Zhifang Wang, Ph.D.

Associate Professor, Department of Electrical and Computer Engineering

Virginia Commonwealth University

Richmond, Virginia

May 2019

Dedication

I dedicate this work to my beloved wife *Shadi* whose unconditional encouragement and support made it possible for me to overcome the challenges of graduate school and life. I am truly thankful for having you in my life.

Acknowledgments

I express my sincere acknowledgment to my advisor Dr. Zhifang Wang for her guidance, endless support and encouragement of my scientific endeavors.

I would like to thank my committee members— Prof. Umit Ozgur, Dr. Weijun Xiao, Dr. Carol Fung, and Dr. Hong-Sheng Zhou for their constructive comments and suggestions.

The love and support of my parents have always been an encouragement for me to achieve what I aim for in my life. My appreciation for them is beyond something that I can write.

The support of my friends and facilities provided by the College of Engineering at VCU is much appreciated.

Table of Contents

List of figures

List of tables

Abstract

The power system is one of the most critical infrastructures in today's world. The reliability and security of such critical system are of great importance because of the huge economic and social cost of blackouts in modern societies and has induced a considerable research effort over the past two decades. One of the biggest threats to power systems is large-scale blackouts resulting from cascading failures (CF) in the grid. After major blackouts in U.S. history, many researchers have tried to study this phenomenon from different aspects. These efforts range from static modeling of the power system for topological studies to the dynamic simulation of the cascading process during the escalation phase of the failures. However, there still is a need for further studies on grid vulnerability to cascading failure especially with the ongoing changes in the portfolio of generation due to ever-increasing penetration of renewable energy (RE) resources.

The literature review on the existing CF simulation models suggests that these models are not suitable for the evaluation of blackout risk in power systems with high penetration of renewable energy resources. The uncertainty injected from RE units will affect the line flows that could potentially impact the tripping events and consequently the grid vulnerability. In this work, efforts have been made to develop tools and models to study CF in power systems in the presence of renewable generation. The uncertainty injected from renewable energy resources and electrical loads are modeled by analyzing actual high-resolution data from North American utilities. Next, two CF simulation models based on simplified DC power flow and full AC power flow (PF) to investigate grid vulnerability under different loading conditions are proposed. In order to estimate the line flow process of AC power flow model in a power system, the Unscented Transform (UT) technique is utilized. The simulations show that for high penetration of RE based on the proposed

UT method we can achieve up to 200, 77, and 65 times better results compared to the DC PF model in the estimation of flow uncertainty variance, bandwidth, and mean, respectively. Furthermore, the integration of RE resources is based on potential RE growth map and geographical information of the grid topology which is an enhancement over the random installation of RE units. The developed tools and models might be useful for the researchers in both academia and industry to investigate the impact of RE penetration on grid vulnerability under various loading levels and RE integration ratios.

Efforts have been made to perform a set of statistical analyses for some real power system data to extract salient characteristics of power grids such as electrical and non-electrical parameters of the transmission network and their interdependence on the nominal voltage level. Finally, in this dissertation a statistical toolkit to perform statistical analysis on power system cases, the *GridStat Analysis Toolkit*, is developed which is a MATLAB GUI-based application designed with an interactive interface for convenient use experience offering four sets of main statistical analysis: 1) Topological analysis, 2) Grid parameters statistics, 3) Voltage interdependence and 4) Grid scaling properties.

1 Introduction

Large-scale blackouts resulting from cascading failures induce considerable economic and social costs annually. Cascading failure (CF) is defined as a sequence of dependent failures of individual components that successively weakens the power system and could result in electrical instability and large-scale blackouts. CFs originate from strong interdependencies inside the grid. Transmission line overload due to contingency is the most common initial cause of CFs in power systems.

The ever-increasing penetration of renewable energy (RE) resources such as wind energy will have many impacts on grid vulnerability and operations. Due to the stochastic nature of wind energy, a large amount of uncertainty will be injected to the grid that will change its dynamic performance. One worrisome change is the increase in CFs involving wind farms [1]. The intermittent out power from wind generators will force other conventional generators to continuously alter their output resulting in big variation in line flows. This could, in turn, impact the loading level of transmission lines. During the escalation phase of a CF that overloaded lines are getting tripped, the change in line flow process due to high penetration of wind energy will affect the cascade of failures and propagation of trips thus consequently the overall blackout size. Therefore, it is becoming more and more crucial to develop tools and models that allow studying the impact of these changes on the risk of CFs leading to blackouts.

1.1 Motivation

Massive economic and social impacts of blackouts have motivated a great deal of research effort on studying the vulnerability of the power grids to CFs. With ever-increasing penetration of

renewable energy resources, the dynamic of the grid is changing. Due to their stochastic nature, the operators of the system face more uncertainty injected to the grid from these sources. The increased level of uncertainty will also change the behavior of the system during the escalation phase of cascading failures.

Therefore, we need to first find a suitable model for the uncertainty of renewable generation. Then, using the developed model the study of grid vulnerability to cascading overload failures under high penetration of renewables can identify their impacts on grid dynamics.

1.2 Literature review

Cascading failures in power systems happen due to several initial causes but they usually result in large-scale blackouts with a huge social and economic cost. After major blackout incidents in history, a great deal of research has been directed towards the modeling and analysis of large-scale blackouts resulting from cascading failures. Since the dynamics of the grid during a cascade process are still unknown, there exist, different models, to simulate and study this phenomenon. Many studies have tried to model the dynamics of the failure based on the power system normal operations regime like optimal power flow, while others proposed stochastic approaches to consider randomness in the evolution of the failures.

A) Existing models for cascading failure analysis

In the model proposed by Dobson, Carresras et al. in [2] the flow re-dispatch is determined by optimal power flow (OPF) calculation similar to the normal operation of the system. This model is the reference for works in [3] and [4]. Authors in [5] propose a CF model inspired by the model in [3] that has a different tripping mechanism in the simulation of CF.

Another approach to the problem in the literature is analyzing the interaction of power grids and communication networks with each other while studying the dynamics of the grid based on power flow analysis. The interdependencies of power and communication networks are modeled in [6] and [7] and it is found that power flow equations are a critical component in CF simulations. Rahnamay-Naeini and Hayat in [8] study the mutual dependencies of the two network using estate estimation techniques.

In the CF model proposed in [9], authors determine the criticality of lines by considering the changes of flow and structure of the network after each level of CF. [10] studies a blackout case in the Indian power grid due to voltage collapse in the inter-regional corridor. Transient stability analysis and cascading failures are considered together in [11] and [12]. Sensitivity analysis of power grids to cascading failures by means of considering their critical dependence on their operating characteristics is performed by Rahnamay-Naeini and Hayat in [13].

Another research line in dynamic failure analysis is the application of *Graph Theory* (GT) and *Complex Network Theory* (CNT).

For example, the models proposed by Motter and Lai in [14] and Holme and Kim in [15] analyze the blackouts resulting from intentional attacks assuming the shortest connection path for flow exchange. This model is modified by Crucitti, Latora, and Marchiori in [16] to more accurately simulate the behavior of the network and found that the difference in load distribution across the network might increase grid vulnerability to CF. Similarly, Wang and Ron [17] proposed a model

where the load of the attached node is redistributed to its neighbors proportionally. In this work, the mechanism of realistic power grid operation is not accurately considered in the model.

More recently, cascading failure models based on complex network theory (CNT) have been proposed in the literature. Authors in [18] proposed a model based on the CNT by combining the node overload failures and hidden failures of transmission lines in blackouts together. The robustness of the grid is studied by Wang et al. in [19] and found that sometimes the addition of a new line to the power system could result in a decreased robustness for the grid. Zhu et al. studied cascading failures of the power grids based on a new model that combines complex network theories with power flow models, called the extended model [20]. Authors in [21] established a cascading failure model based on the CNT where the development tendency of cascading failure is determined by the network topology, the power flow, and boundary conditions.

A few authors analyze the dynamics of cascading failure using stochastic approaches. Authors found a good fit for the cumulative number of trips based on the branching model in [22]. We should note that unlike the communication networks where packets can switch their route easily choosing the shortest path, the flows in a power grid are determined by the physical constraints imposed by Kirchhoff's Voltage/Current laws (KVL and KCL) and Ohm's law [23] which is unique for the electric networks. Another distinction is that in a power grid nodes are categorized into three classes: generators, loads and intermediate nodes and cannot be interchanged with each other. Wang *et al.* proposed a stochastic Markov model to capture the progression of CFs considering uncertainty coming from only electrical loads without accounting for thermal stability model for thermal relays tripping time [24]. More recently, a stochastic model is presented in [25] for describing cascading failure in a cyber-coupled smart grid where they combine stochastic

process with a state transition description. In this work, the CF is simulated considering the effects of overloading, interdependency between the power grid and cyber network, and malware contagion.

In the literature, there are several studies that focus on the static failure analysis by analyzing the topological robustness of the power system against cascading failures. These models analyze the topology change during the escalation phase of cascading failure due to intentional attacks. [26] and [27] examine the node removal effect on power system connectivity. Wang, Scaglione, and Thomas improved the nodal degree model and showed an excellent fit with a mixture distribution [28]. They find that the power system is even more vulnerable to disconnection under the realistic nodal degree distribution model. In [29] the cascading link failures are studied by using the random geometric graph model that does not accurately reflect the nature of the power system. Authors in [30] tried to identify the critical buses in terms of topology vulnerability by gauging the attack impact for several realistic grids. They only reported the direct simulation results and compared them with predictions of other models. More recently, Dey et al. in [31] suggest corrective actions to save the system from total collapse by considering topological characteristics of the grid. In this paper, the average propagation of failure is calculated as a branching process.

We can also categorize the CF models in the literature based on their approach in simulating power system response during cascades of tripping events [32]. In this regard, there exist two major approaches including dynamic transient models [33]–[38] and quasi-steady state (QSS) models [2], [24], [39]–[47] each has advantages and disadvantages.

1) Dynamic transient models

In the dynamic models, the dynamic components, such as rotating machines, exciters, and governors, as well as all protective components of the system along with their dynamic behavior, are modeled using differential equations. The computational burden of the simulation for large cases and the numerical failure in solving differential equations are disadvantages associated with dynamic models that prevent running multiple Monte Carlo (MC) simulations to assess the risk of blackouts for different planning scenarios. The models in [33], [34] and the COSMIC model in [35], [36] are examples of research-grade dynamic cascading failure models and [37], [38] are examples of existing commercial simulation tools that have introduced dynamic simulation to their cascading failure analysis.

2) Quasi-steady state models

On the other hand, the QSS models are widely used in the literature to study the cascading failure and evaluate the risk of large-scale blackouts. These models rely on the steady-state assumption for the system where the flow re-dispatch of the network is calculated based on power flow (PF) analysis. They differ from each other in terms of the assumptions they make to simulate the cascading failure and the power flow model used. The representation of the transmission system can be based on the full version of PF equations (i.e., ACPF), or on the linearized version (i.e., DCPF). The DCPF approximation is the common technique used in the QSS models due to its guaranteed convergence and low computational cost which allows the simulation of failures beyond any topology changes for the grid. In addition, its linear property makes the direct estimation of flow process statistics from those of the injected power possible as utilized in the proposed mixed OPF-stochastic CF model in [42]. However, the DC approximation comes at the expense of assuming flat voltage profiles for the entire network thus hindering the simulation of

voltage related failures during the CF. Large errors in flow estimation by DC model is also identified in the literature especially for larger networks [48]–[50]. The ORNL-PSerc-Alaska (OPA) models in [2], [39], the random chemistry model in [40], the Markov-transition model in [24], and more recently the mixed OPF-stochastic model in [41], [42] are examples of the cascading failure models that employ DCPF. Whereas the AC OPA model in [43], Manchester model in [44], TRELSS model in [45], importance sampling model in [51], and more recently the AC-OPF-f model in [47] are among the models employing full AC power flow in the simulation of cascading failures.

B) Studies on the impact of Renewable Generation on cascading failure

While studies for modeling the cascading failure process in power grids have been prevalent in the literature, the study of impacts of stochastic renewable generation on grid vulnerability to cascading failures is lacking. Especially, the highly variable nature of renewable generation will have a large impact on the dynamics of voltage profiles that can totally change the behavior of the grid during cascading failures. The effects of replacing conventional generation by wind and solar generation on the grid voltage performance are examined in [52]. Authors used a western electricity coordination council (WECC) equivalent system to identify the issues with voltage performance after such generation alteration. But, there is a need to analyze the impact of this change in voltage dynamics on grid vulnerability and voltage stability during cascading failure. Henneaux *et al.* studied the impact of thermal effects on the risk of blackout for increased wind farms [53]. They found that high penetration of wind energy to the grid will increase the variability of cross-border flows, therefore, leading to a higher risk of cascading failure. The short circuit capacity margin of renewable energy sources is analyzed in [54] to evaluate the state vulnerability

of the grid. The false tripping of protection relays is examined in [55] to see the impact of high penetration of distributed generation (DG) on the performance of the protection system. The authors found that an increase in the number of DG units could result in an increase in the number of undervoltage tripping which will be even more problematic with the increase in power electronic devices in the distribution network. Authors in [56] proposed an online assessment system to evaluate, analyze and predict the CFs of a group of wind farms timely and effectively. Cascading tripping out of numerous wind turbines in China is analyzed in [57] to identify important factors contributing to the failures. Khazaei et al. in [58], [59] proposed renewable energy aggregation to reduce the impacts of uncertainty on the network. Scala *et al.* in [60] found that the presence of fluctuations due to erratic renewable sources and customer demands increase the instability within an isolated segment of a power grid. However, none of the existing studies examine the impact of intermittent renewable generation on voltage dynamics within the cascading failure context.

1.3 Contributions

The models developed for cascading failure analysis over the past two decades such as Motter-Lai model [14], [61], CASCADE model [62], Branching process model [63], ORNL-PSerc-Alaska (OPA) model [3], [39], [64], Manchester model [65], and stochastic models [22], [24], [66]–[68], do not allow considering high uncertainty level injected from RE sources on the simulation of line outages during cascading failures. There are a few works in the context of CF analysis considering the RE resources. For example, Henneaux *et al.* in [53] studies the impact of thermal effects on the risk of blackout for increased wind farms and found that high penetration of wind energy to the grid will increase the variability of cross-border flows, therefore, leading to a higher risk of cascading failure. In another work in [55], the impact of high penetration of distributed generation

(DG) on the performance of the protection system is investigated by analyzing the false tripping of protection relays. And Scala *et al.* in [60] found that the stability of a power grid within an isolated segment could increase due to power fluctuations caused by renewable sources. However, there is still a need for models and tools that can simulate CF in power system similar to traditional CF models in the literature (e.g. AC OPA, Manchester model, …) while allowing for modeling the impact of RE on line flow process. As the new technologies are integrating into the modern grids, they become more and more complicated with an unprecedented uncertainty level that the grid has to deal with. Especially, simulation of voltage profile behavior with high penetration of variable renewables during the cascading failure is lacking from the literature.

In this dissertation, models and tools are developed to study the cascading failures in power system in the presence of highly variable and uncertain renewable energy resources. Following is the list of our main contributions to the state of the art in cascading failure analysis.

- Analyzing uncertainty injected from different renewable energy resources and electrical loads and developing an uncertainty model for cascading failure simulations
- Proposing a mixed OPF-stochastic cascading failure model based on the simplified DC power flow model suitable for preliminary vulnerability studies
- Estimating the line flow process for AC power flow model utilizing the Unscented Transform (UT) method
- Proposing an AC cascading failure model based on UT methodology to simulate voltage-related failures during CF phenomenon
- Benchmarking the proposed ACUT model with the historical blackout data as well as a number of CF models in the literature

- Some preliminary results on the RE penetration impacting the grid vulnerability

The developed tools and models might be useful for experts both in academia and industry to evaluate the impact of penetration of renewables on grid vulnerability to cascading failure. Especially, after further examination and validation, from the long-term planning standpoint, it might also be able to identify the suitable locations of integration for potential renewable energy resources to minimize their adverse impacts on grid vulnerability.

Moreover, both the DC and AC proposed CF models are quasi-steady state (QSS) models. Since the dynamic transient models simulate the dynamics of the grid based on the differential equation, they are very computationally expensive preventing a large set of Monte-Carlo (MC) simulations. Whereas the QSS models (including our proposed models) are less computationally expensive and allow simulation of large set MC simulations of different N-2 contingencies to evaluate the average impact of RE penetration on grid vulnerability.

Finally, efforts have been made to perform a set of statistical analyses for some real power system data to extract salient characteristics of power grids such as electrical and non-electrical parameters of the transmission network and their interdependence on the nominal voltage level. These efforts have led to the development of a MATLAB GUI-based application for convenient statistical analysis on power grids. These analyses are presented in the Appendix.

1.4 Structure of the dissertation

In chapter 2 we present the proposed uncertainty modeling approach and analyze actual data to characterize uncertainty injected from different RE sources as well as electrical loads. In chapter 3, CF in power systems will be briefly discussed. Chapter 4 will present the proposed DC CF

model for preliminary analysis. The enhanced AC CF model called ACUT will be presented in chapter 5. Chapter 6 is dedicated to validation and benchmarking the proposed model. Finally, chapter 7 will conclude the dissertation and make recommendations for future works. Our work on statistical analysis on power grid parameters and variables is presented in the Appendix.

2 Uncertainty modeling

In this chapter, we present the procedure to develop an uncertainty model to represent the forecasting error of smart grid loads and wind generation as an example of erratic renewable generation. Note that, a similar procedure can be followed to analyze and characterize the uncertainty from other renewable energy sources such as solar PV and solar thermal units. This section includes a brief introduction to the problem, the discussion on the proposed model, description of the historical data used in model construction, and the transition from generation and loads uncertainty to line flow uncertainty in a standard power system test case.

2.1 Introduction

The uncertainty handling has been one of the main concerns of the decision makers (including governors, engineers, managers, and scientists) for many years [69]. Most of the decisions to be made by energy sector decision makers are subject to a significant level of data uncertainty [70]. The uncertain parameters in power system studies can be generally classified into two different categories including:

- Technical parameters: These parameters are generally categorized in two main classes, namely topological parameters and operational parameters. The topological parameters are those related to network topologies like a failure or forced outage of lines, generators or metering devices, etc. The operational parameters are tied with operating decisions like demand or generation values in power systems. The variability of renewable generation is an example of operational uncertainty that can be modeled with forecasting error.
- Economical parameters: The parameters which affect the economic indices fall in this

category. Microeconomics investigates the decisions of smaller business sectors like aggregators, domestic or industrial consumers while macroeconomics focuses on the entire power system industry. For example, uncertainty in fuel supply, costs of production, business taxes, labor, and raw materials are analyzed in microeconomics. On the other hand, issues like regulation or deregulation, environmental policies, economic growth, unemployment rates, gross domestic product (GDP) and interest rates are analyzed in macroeconomics. All of these parameters are subject to uncertainties and should be correctly addressed in economic studies.

There is various uncertainty handling methods developed for dealing with the aforementioned uncertain parameters. The main difference between these methods is in line with the different techniques used for describing the uncertainty of input parameters. For example, the fuzzy method uses membership functions for describing an uncertain parameter while the stochastic methods use probability density function. The similarity of them is that all of them try to quantify the effect of input parameters on the model's outputs. These methods and the way the uncertainty is handled by them are described as follows:

- Probabilistic approach [71]: It assumes that the input parameters of the model are random variables with a known probability density function (PDF).
- Possibilistic approach: The input parameters of the model are described using the membership function (MF) of input parameters based on Fuzzy logic.
- Hybrid possibilistic–probabilistic approaches: Both random and possibilistic parameters are present in the model.
- Information gap decision theory [72]: In this method, no PDF or membership function is

available for input parameters. It is based on the difference between what is known and what is vital to be known by quantification of severe lack of information in the decision-making process.

In this thesis, the renewable generation uncertainty is modeled with operational parameters. This is reflected in the erratic renewable energy injected to the grid that can be represented by the forecasting error. Other sources of uncertainty such as economical parameters can be considered in the context of the day-ahead electricity market and generation dispatch. One way to model different uncertainty sources in cascading failure studies is Monte-Carlo simulations where the blackout risk assessment can be performed by a large number of scenarios accounting for both operational parameters (renewable generation variability) and economical parameters (unit dispatch).

Another interesting line of research is to study the correlation of operational parameters with weather data since especially for renewable generation weather information are critical. For example, the correlation can be seen in the output power of nearby wind farms and their variability to accurately represent the variability in their injected power.

In this chapter, we define wind power uncertainty by the high dynamic variability in its output power signal and model it by the forecasting error. Similarly, the electric loads' uncertainty is defined by the forecasting error for demanded power. The Autoregressive Moving Average (ARMA) method is simple and easy to implement technique widely used in time series forecasting. The several steps ahead values of a time series are calculated based on a linear relationship between previous values of the series. The historical data are used to determine the best parameters of the model by minimizing the root mean squared error (RMSE) of the output signal. The ARMA model

is used for both wind generation and electrical loads because of its good performance. It is able to capture the daily seasonality in load profile. Application of more complicated methods such as Artificial Neural Networks (ANN) to forecast and evaluate the uncertainties of wind generation may produce misleading results due to higher accuracy of the model compared to the actual forecasting techniques used in the industry.

We use very high-resolution data for both generation and loads and then analyze the error signal for each uncertainty source to evaluate their dynamic features. The occupied bandwidth of the error signal is an important metric that shows the frequency components of the uncertainty for the source and will be later used in the stochastic cascading failure model.

2.2 Proposed uncertainty model

A) Uncertainty representation

Figure 1 shows a hypothetical power grid with different kinds of generations. Wind turbines and PV plants are the main sources of renewable energy that are widely used in power grids. Along with renewable sources, there are many conventional generation units such as fossil fuel, hydro, co-generation, and steam units. Each of these generation units injects a different amount of uncertainty into the grid with different characteristics in terms of magnitude and frequency of occurrence. Despite recent developments in load forecasting techniques, there is still uncertainty coming from load forecasting error.

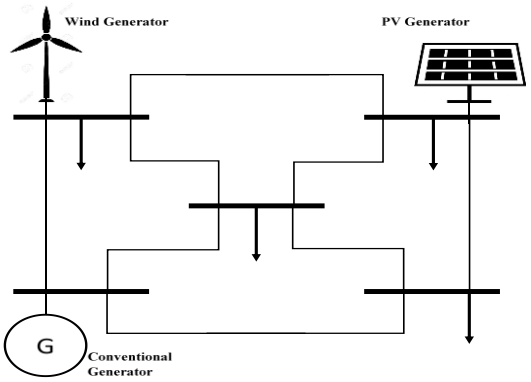


Figure 1. Schematic diagram of a hypothetical grid

In the proposed model, the power signal of each component (i.e. output power for generators and demand power for loads) is modeled with two terms as shown in Eq. (1).

$$P(t) = \mu(t) + \epsilon(t) \quad (1)$$

where $\mu(t)$ is the average of power signal (MW) at each time instant or in other words it is what we expect to have for each component ahead of time. For example, we can consider the forecasted value for load or renewable generation or scheduled output of a conventional generation unit for next day as $\mu(t)$. In this model $\epsilon(t)$ represents the uncertainty (MW) which can come from forecast error or mismatch in output power for conventional units. In this study, load and wind output power forecast errors are considered as uncertainty. Also, mismatch in output power from scheduled value for conventional generation units shows their uncertainty.

B) Model requirements

1) Data: since we are looking for a model which can accommodate the high dynamics of uncertainty associated with a different source of variations, high-resolution data with sampling rate in seconds range seem necessary. For this purpose, we obtained the data from the Electric Reliability Council of Texas (ERCOT) which offers high-resolution data that are measured every

4 seconds. These data are for numerous wind generators, load demand, and different types of conventional generations. In this study, fossil fuel generators and co-generation are used as a conventional generation. Figure 2 shows data for different components of the ERCOT network. Wind output power shows a highly variable signal which in turn will result in large uncertainty with high-frequency components.

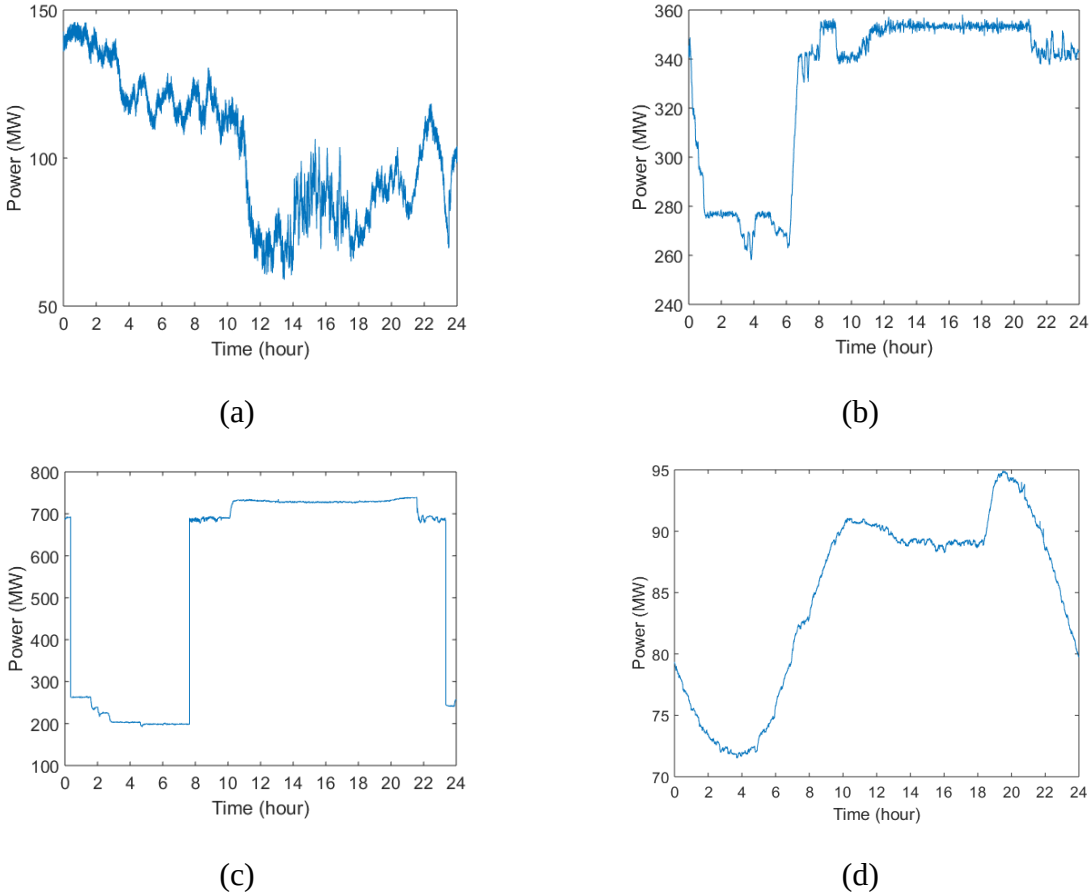


Figure 2. ERCOT 4 sec data for a) wind generator, b) fossil fuel generator, c) co-generation and d) load

2) Forecasting model: In this study in order to capture the high-frequency component of uncertainty associated with each source, a very short-term forecasting horizon is chosen for load and wind forecasting. Traditional Autoregressive Moving Average (ARMA) forecasting model offers a high accuracy when forecast time is short [73]. Therefore, the ARMA model is used to forecast load power wind output power. For a stationary time-series of load data, because the

properties of historical and future data are mutually similar, the historical data can be used as the reference to formulate an adequate ARMA model. The forecasting with ARMA model essentially can be divided into three steps, including model identification, parameter estimation, and adequacy validation.

System modeling can be expressed as the following ARMA form:

$$\phi(B)y_t = \theta(B)a_t \tag{2}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \tag{3}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \tag{4}$$

where y_t is the observed time series of load at time t, a_t is the white noise at time t, and B is the back-shift operator such that $By_t = y_t - 1$ and $B^m y_t = y_t - m$. ϕ_1, …, ϕ_p are AR part parameters and p is the AR order. θ_1, …, θ_q are MA part parameters and q is the MA order. The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are used as references to conjecture the appropriate model order [74].

Parameter estimation for ARMA model is done based on gradient-based method, in which they are estimated in order to have zero gradients of the mean squared sum of fitting errors to historical data. Finally, the adequacy of the model is validated after the appropriate estimation of model parameters. These parameters should be significantly different from zero and the residuals are the realization of the white noise process [74]. ARMA forecasting model parameters for load and wind forecasting are shown in Table I.

Table I. ARMA forecasting model parameters

	Model parameters			
	p	q	differencing order	seasonality
Load ARMA	10	10	2	Daily
Wind ARMA	50	50	1	None

3) DC power flow

To study the effects of uncertainty from different sources on transmission line flows, DC power flow approximation is used in this study. DC power flow is a standard approach widely used in optimizing flow dispatch and for assessing line overloads [75]. For a power grid with *n* nodes and *m* transmission lines, the network flow equation can be written as follows:

$$P(t) = B'(t)\theta(t) \tag{5}$$

$$F(t) = diag\big(y_l(t)\big)A\theta(t) \tag{6}$$

where $P(t)$ represents the vector of injected real power, $\theta(t)$ the phase angles, and $F(t)$ the flows on the lines. The matrix $B'(t)$ is defined as

$$B'(t) = A^T diag\big(y_l(t)\big)A \tag{7}$$

where $y_l(t) = 1/x_l$ is the line admittance; $diag\big(y_l(t)\big)$ represents a diagonal matrix with entries of $\{y_l(t), l = 1,2, \dots, m\}$. $A \coloneqq (A_{l,k})_{m\times n}$ is the line-node incidence matrix, arbitrarily oriented and defined as: $A_{l,i} = 1$; $A_{l,j} = -1$, if the lth line is from node i to node j and $A_{l,k} = 0, k \neq i, j$.

2.3 Uncertainty statistics for different sources

Based on the proposed model by [76], frequency components of injected power in buses which will translate into frequency components of line flows have a direct impact on grid vulnerability against random disturbances. In other words, the more the bandwidth of uncertainty, the more vulnerable the grid against failures.

The power spectrum $S_{xx}(f)$ of a time series $x(t)$ describes the distribution of frequency components composing that signal. The power spectral density (PSD) of a signal refers to the spectral energy distribution per unit of time. The spectrum of physical processes often contains essential information about the nature of them. One particular information that can be useful for our purpose is the occupied bandwidth of the signal. It can clearly represent the dynamics of signal that in our case shows the dynamics of uncertainty for different sources. In order to obtain the PSD of a signal (e.g. wind uncertainty) first, we define the autocorrelation function as [77]

$$R_{XX}(\tau) = \frac{1}{N\sigma^2}\sum_{t=1}^{N-\tau}(X_t - \mu)(X_{t+\tau} - \mu) \tag{8}$$

where N is the number of samples, σ^2 is the sample variance of the time series, and μ is the mean of the samples.

Next, the PSD can be calculated by applying the Fourier transform to the autocorrelation function:

$$S_{xx}(f) = \mathcal{F}[R_{XX}(\tau)] = \frac{\Delta t}{N}\left|\sum_{n=0}^{N-1} x_n e^{-j2\pi f n}\right|^2, \quad -1/2\Delta t < f < 1/2\Delta t \tag{9}$$

Where x_n is the uncertainty signal, Δt is the sampling interval, and N is the total number of samples. For a one-sided PSD, the values at all frequencies except 0 and the Nyquist, $1/2\Delta t$, are multiplied by 2 so that the total power is conserved [78], [79].

Finally, the 99% occupied bandwidth is the frequency range containing 99% of the spectral energy and can be calculated by solving:

$$\frac{\int_{-f^*}^{f^*} S_{xx}(f)df}{\int_{-\infty}^{+\infty} S_{xx}(f)df} = 0.99, \quad f^* = BW_{99\%} \tag{10}$$

In this study, the 99% occupied bandwidth is considered as a metric to show the dynamics of uncertainty.

The root mean square error (RMSE) which is also equal to the standard deviation of the distribution and mean absolute percentage error (MAPE) are metrics to show the magnitude of uncertainty from each source and can be calculated based on the following equations.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N} \epsilon_i^2} \tag{11}$$

$$MAPE = \left(\frac{1}{N}\sum_{i}^{N} \frac{\epsilon_i}{P_i}\right) \times 100 \tag{12}$$

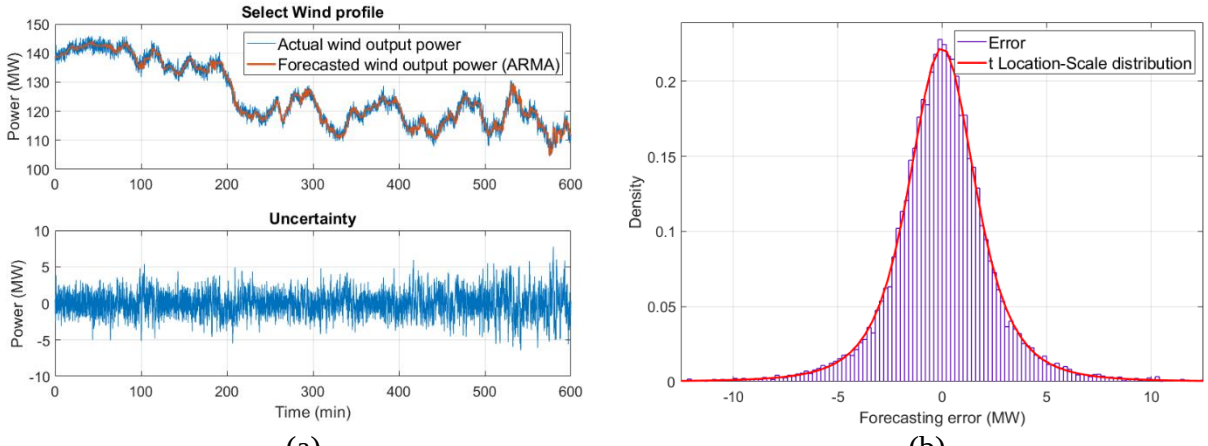
where ϵ is the uncertainty magnitude at each time step, N is the total number of samples (21600), and P is the magnitude of the actual power.

2.3.1 Wind generation uncertainty

Based on ARMA forecasting model, the uncertainties associated with wind generator output power and load are modeled. The uncertainty associated with conventional generation (i.e. fossil fuel generation and co-generation) is modeled based on the deviation of actual plant output from its scheduled value. Figure 3 shows the forecasting result and the distribution of forecasting error (i.e. uncertainty) for a select wind profile. Figure 3 (b) shows the probability distribution function of forecasting error for wind output power. The probability distribution function of forecasting error for wind output power can be approximated with *t location-scale* distribution with the following PDF:

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[\nu+\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)} \tag{13}$$

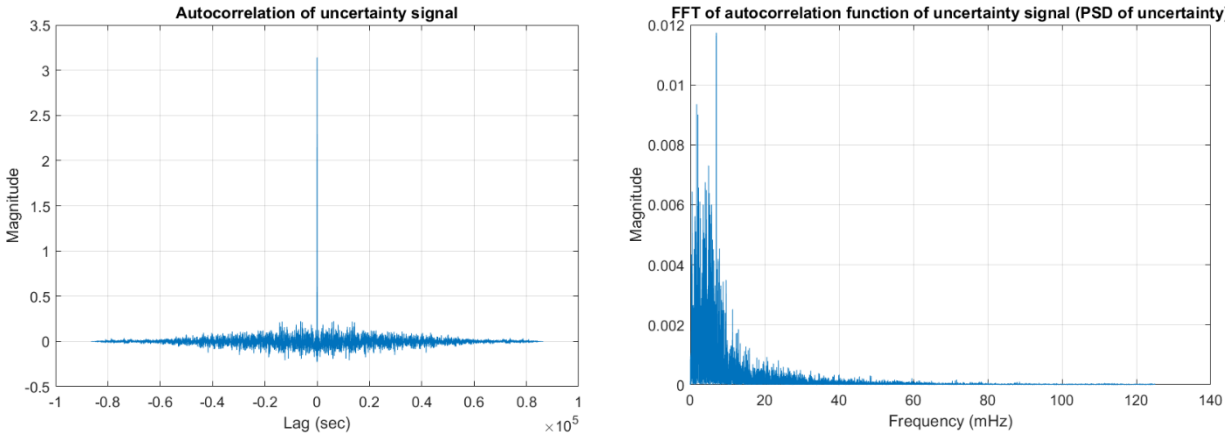
where Γ is the gamma function, μ is the location parameter, σ is the scale parameter, and ν is the shape parameter. The mean of the *t location-scale* distribution is μ and the variance is $var = \sigma^2\frac{\nu}{\nu-2}$. Note that if random variable x has a *t location-scale* distribution with parameters μ, σ, and ν, then $\frac{x-\mu}{\sigma}$ has a *student's t* distribution with ν degrees of freedom.



(a) (b)

Figure 3. Wind uncertainty results: a) forecasted and actual power and b) distribution of error

If we take the uncertainty signal showed in Figure 3 (b) and calculate its autocorrelation function, we then can obtain the PSD of the uncertainty and according to Eq. (12) we can calculate the 99% bandwidth of the error signal. Figure 4 shows the uncertainty signal analysis. Figure 4 (a) shows the autocorrelation of the uncertainty signal and Figure 4 (b) shows the PSD of the signal.



(a) (b)

Figure 4. Uncertainty signal analysis: a) autocorrelation of the signal and b) Fourier transform of the autocorrelation (PSD)

Figure 5 shows the PSD of the uncertainty signal for the selected wind profile. Based on Eq. (12) we have obtained the bandwidth that contains 99% of the spectral energy. This bandwidth is shown on the figure.

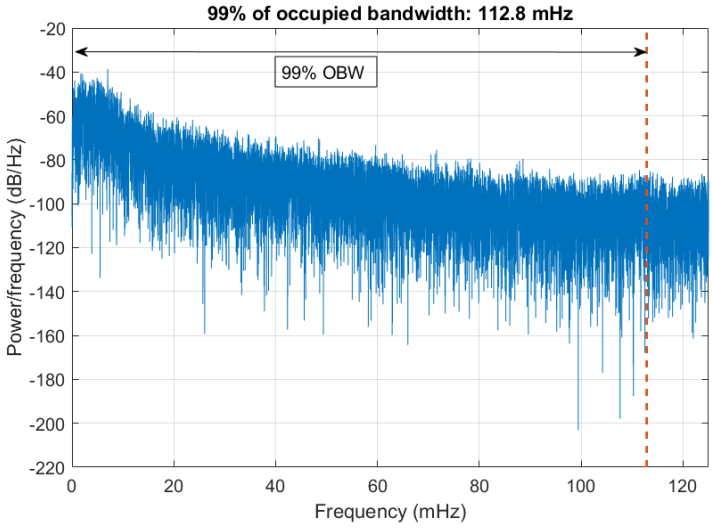


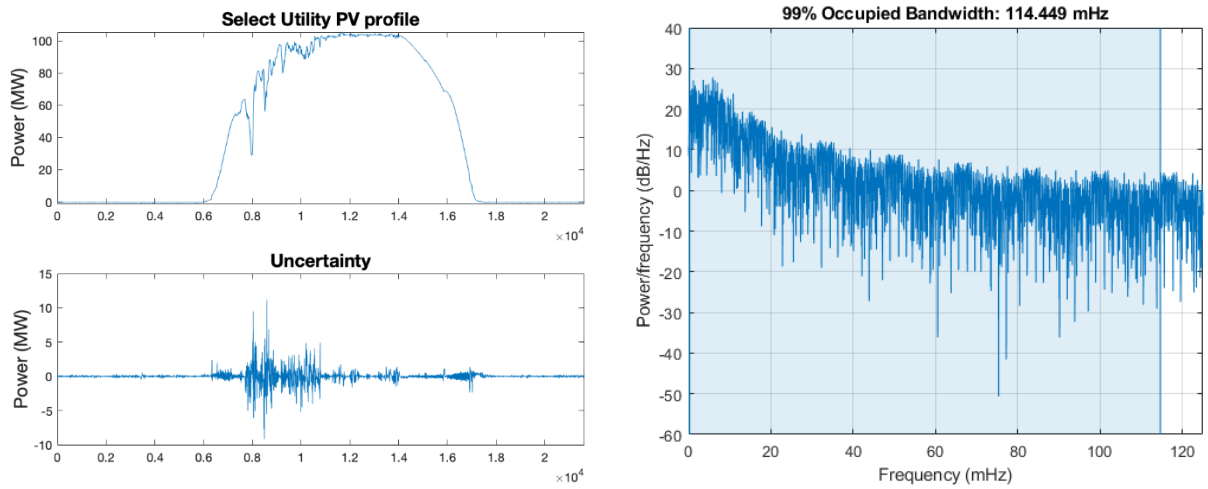
Figure 5. 99% occupied bandwidth of the uncertainty for a select wind profile

The signal processing toolbox of MATLAB offers a tool, called OBW, to directly estimate the occupied bandwidth of a signal. To determine the occupied bandwidth, this tool computes a periodogram power spectral density estimate using a rectangular window and integrates the estimate using the midpoint rule. The occupied bandwidth is the difference in frequency between the points where the integrated power crosses 0.5% and 99.5% of the total power in the spectrum. For the rest of this dissertation, we will use MATLAB's OBW command to calculate and plot PSD and occupied bandwidth of uncertainty signals.

2.3.2 PV generation (utility-scale and distributed PV) uncertainty

The high-resolution data received from ERCOT do not include solar PV generation. We obtained solar PV generation data from the local utility, Dominion Energy. The data have a resolution of 30 seconds and consist of PV generation data for both utility-scale and distributed PV. The utility-scale PV site has a capacity of 120 MW and the distributed PV sites capacity range from 5 MW to 20 MW. The challenge with the PV data set is to synchronize their resolution with ERCOT 4 second data. This was achieved by means of data padding based on linear interpolation and time table toolbox in MATLAB. After synchronizing the resolution of PV data, we can compare their statistics with those of ERCOT wind and load data.

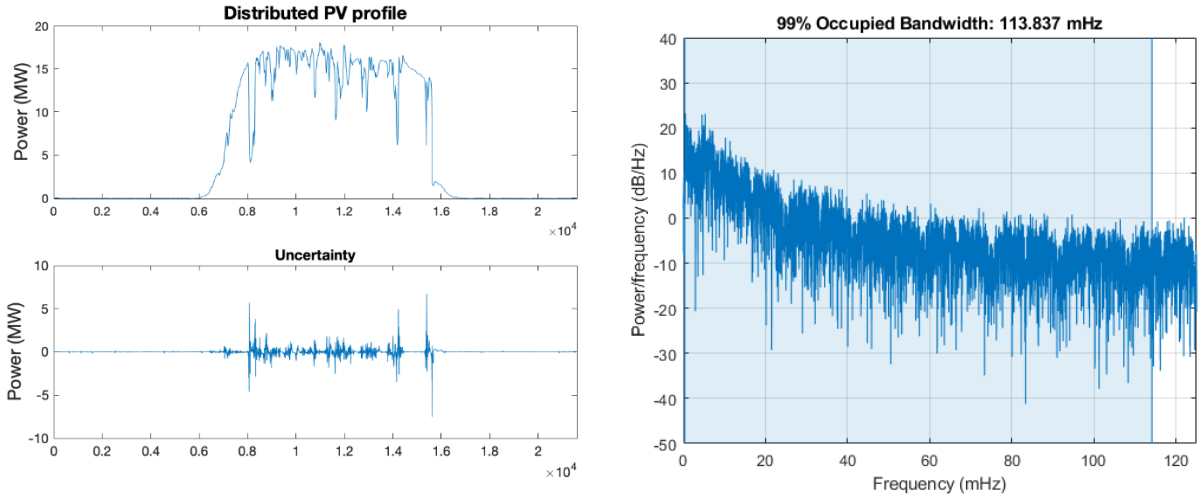
Figure 6 shows the daily profile of output power for the utility-scale PV generation site. The uncertainty signal (i.e. the forecasting error) is also depicted for this generation site.



(a) (b)

Figure 6. Utility-scale PV generation profile supplied by Dominion Energy: a) daily profile and the uncertainty signal and b) 99% OBW of the uncertainty signal

The 99% OBW is depicted in Figure 6 (b) for the utility-scale PV. The bandwidth for PV uncertainty is very close to the bandwidth obtained for wind generation which indicates that both wind and PV are highly variable generation units injecting the major portion of overall grid uncertainty.



(a) (b)

Figure 7. Distributed PV generation profile supplied by Dominion Energy: a) daily profile and the uncertainty signal and b) 99% OBW of the uncertainty signal

Figure 7 shows the distributed PV profile and its associated uncertainty signal for a select site with 20 MW capacity. As shown in Figure 7 (b), the OBW of the distributed PV is similar to the

bandwidth of the utility-scale PV and wind generation and can be considered a major source of uncertainty. However, one should note that the size of distributed PV generations is much smaller than utility-scale PV units, therefore, although the bandwidth of the injected uncertainty is comparable to utility-scale units, the size of uncertainty is significantly smaller. This means that the effect of uncertainty injected from distributed energy resources (DER) such as distributed PV units can be captured by the uncertainty of electrical loads. In other words, the local distributed generation will change the load profiles and their uncertainty add to the uncertainty of loads seen by the grid.

2.3.3 Conventional generation uncertainty

Conventional thermal units mainly include coal, gas, and nuclear generators. The hydropower generators also are categorized in the conventional generation units. These units can vary in size and depending on their cost function will operate as a baseload supplier or peak load supplier. For example, nuclear power plants always play the role of baseload supplier because they have to operate continuously without any interruption. This is due to the fact that their startup cost is high compared to other types of generation units while their operation cost is very small. In addition, due to safety and stability reasons, nuclear power plants are preferred to work with a steady output. On the other hand, the gas generators, usually play the role of peak takers because of their fast ramping capability. Figure 8 shows the scheduling of a fossil fuel generator that gives the uncertainty associated with each of this source. As compared to uncertainty injected from renewable energy sources like wind and PV, the uncertainty resulting from conventional generators are very small and can be neglected.

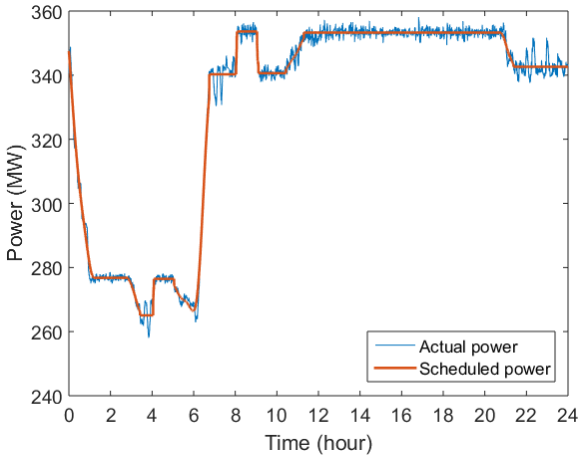


Figure 8. Fossil fuel generator actual and scheduled power

2.3.4 Electrical loads

Figure 9 shows the forecasting result for the load that gives the uncertainty associated with this source. The loads are the second biggest sources of uncertainty that are being injected to the grid. The uncertainty level of loads depends on the accuracy of the forecasting technique being employed. We utilize the ARMA method to achieve a reasonable amount of accuracy while not underestimating the amount of forecasting error with load forecasting.

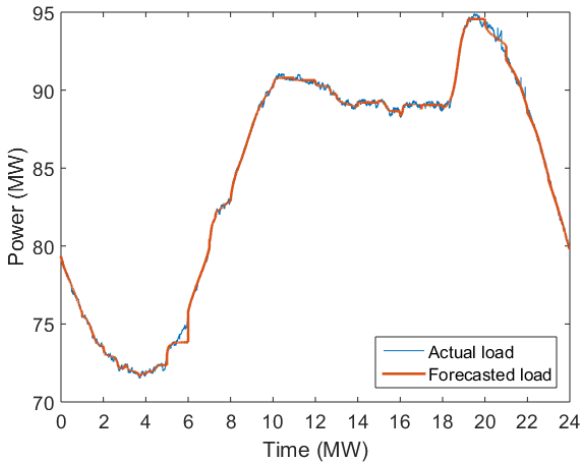


Figure 9. Actual and forecasted load profiles

Table II summarizes the uncertainty analysis results for different components of the power grid. Note that in this table, the PV data with 4-second resolution obtained by means data padding is utilized in uncertainty analysis. As can be seen from the table, the renewable energy sources (i.e. wind and PV) have the highest bandwidth, meaning that they inject faster dynamics to the grid. In terms of uncertainty size, the wind generation injects the largest, while the load injects the smallest uncertainty to the grid. Since, the bandwidth of PV and wind uncertainty are very close to each other, while wind uncertainty size is significantly larger than that of PV, for the purpose of simulation, we will use wind as the representative of renewable energy resources in the cascading failure simulations.

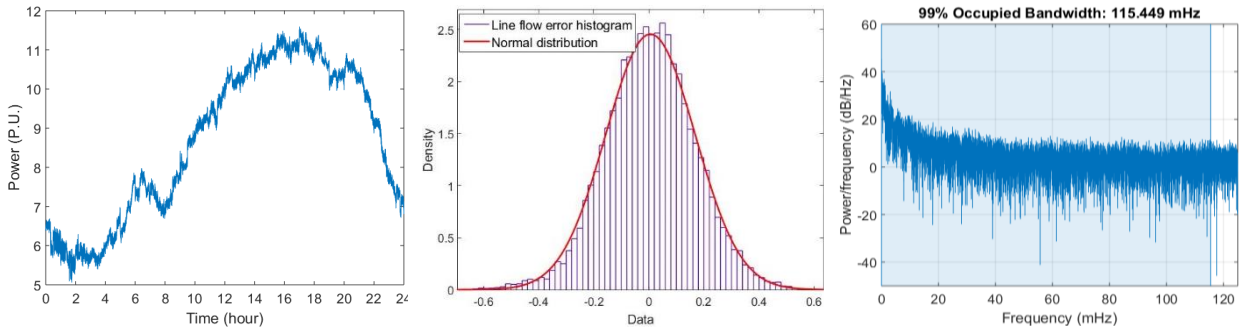
Table II. Summary of uncertainty analysis for different sources

Uncertainty	Avg. Bandwidth (Hz)	Avg. RMSE (MW)	Avg. MAPE (%)
Load	61.84	0.20	4.16
Co-generation	72.45	0.94	6.22
Fossil fuel generation	63.03	1.05	5.91
Wind generation	117.24	8.09	10.32
Utility PV generation	114.90	1.28	1.36
Distributed PV generation	112.20	0.45	5.26

2.4 From generation uncertainty to flow uncertainty

The proposed model is tested on IEEE 300 bus system to evaluate the impact of uncertainty on the flow of lines. To see the impacts of renewable energy penetration into the grid, 11 conventional generators of the original system are replaced with wind generators. Using DC power flow approximation, the power flowing from bus 4 to 16 is calculated as shown in Figure 10 (a). The distribution of line flow uncertainty and the associated bandwidth are shown in Figure 10 (b) and (c), respectively. It is found that line flows capture the highest bandwidth of uncertainties associated with different sources which in our case was wind generator with a bandwidth of 117

mHz. Also, the uncertainty probability distribution function for line flow found to approximately follow normal distribution which can be explained by Central Limit Theorem.



(a) (b) (c)

Figure 10. IEEE 300 bus system line 4-16 flow results: a) line flow, b) distribution of flow uncertainty and c) occupied bandwidth of flow uncertainty

3 Cascading failures in power systems

This chapter discusses the cascading failures phenomenon in power systems and their underlying causes and involved factors. Also, the simulation of cascading failures and the widely used simulations methods in the literature will be discussed briefly.

3.1 Introduction

Large-scale blackouts resulting from cascading failures induce considerable economic and social costs annually. Cascading failure (CF) is defined as a sequence of dependent failures of individual components that successively weakens the power system and could result in electrical instability and large-scale blackouts and they originate from strong interdependencies inside the grid. Massive economic and social impacts of such events have motivated a great deal of research effort on studying the vulnerability of the power grids to CFs. Transmission line overload due to contingency is the most common initial cause of CFs in power systems.

Renewable energy integration and power system deregulation may drive the electric grid closer to its operational limits and introduce a large amount of uncertainty coming from their stochastic nature that changes the grid's dynamic performance. One worrisome change is the increase in CFs involving wind farms [1]. Therefore, it is becoming more and more crucial to study the impact of these changes on the risk of CFs leading to blackouts. Henneaux *et al.* studied the impact of thermal effects on the risk of blackout for increased wind farms [53]. Authors in [56] proposed an online assessment system to evaluate, analyze and predict the CFs of a group of wind farms timely and effectively. Cascading tripping out of numerous wind turbines in China is analyzed in [57] to identify important factors contributing to the failures. Khazaei et al. in [58], [59] proposed

renewable energy aggregation to reduce the impacts of uncertainty on the network. Scala *et al.* in [60] found that the presence of fluctuations due to erratic renewable sources and customer demands increase the instability within an isolated segment of a power grid. However, none of these studies has evaluated the impacts of increased uncertainty injected from wind generation on the grid vulnerability to CFs in the complex interconnected power networks.

In this chapter, we analyze the dynamics of the grid during cascading failure and study the impact of highly variable renewable generation has on the transmission flow process.

3.2 Markovian model for cascading failure

In the Markovian state-based model the power system is analyzed based on the state of its components such as transmission lines and transformers. During the CF the state of the system transitions from one state to another state as the lines get tripped. This transition is usually shown by the state graph.

A) State graph

The state graph shows all possible states for a system. It also shows the probability of transitions from one state to the other [80]. For example, the two states for a transmission line could be "closed" and "open" states. The transition from "closed" to "open" means line trip or failure and the reverse transition means reclosing or repair.

In a Markovian process which is a continuous process, the future states depend only on the present state and not the past states [81]. The CF model proposed in [24] is based on Markovian process assumption for the cascade of failures in the power system.

B) Markovian model for cascading failure

Authors in [24] proposed a conditional Markov transition model for cascading failures in power system which is able to indicate which part in the network will be under stress and therefore most likely to break down given current network conditions and states. In this model, the grid state is defined as a vector of line states (0 for open and 1 for closed), and lines will transition from one state to another based on the transition rate (i.e. probability of failure for transitioning from "1" to "0"). Also, in this model, the tripped lines could transition to state "1" due to relay reclosing actions.

Next, by assuming a Gaussian assumption for line flow process, the authors connect the above Markovian transition model with line overloading and tripping mechanism to simulate the cascading failure in a power system. Note that, the power re-distribution after each line trip is calculated based on power flow analysis.

3.3 Cascading failures simulation

Cascading failures simulation has been the subject of research in the past two decades. Since cascading failures happen rarely in power systems, there is still little knowledge about their underlying causes and promoting factors. Therefore, researchers have come up with various simulation models to investigate different aspects of this phenomenon to reveal the weakness of the grid and potentially recommend remedial actions. There are many different models in the literature that each has its own advantage and disadvantage and is suitable for a special application.

The research on CF in the literature mainly focuses on modeling and analysis to assess the blackout risk for a given network [32]. There exist two major approaches in simulating CF including

dynamic transient models [33]–[38] and quasi-steady state (QSS) models [2], [24], [39]–[47] each have advantages and disadvantages.

As discussed in the literature review in the dynamic models, the dynamic components, such as rotating machines, exciters, and governors, as well as all protective components of the system along with their dynamic behavior, are modeled using differential equations. The computational burden and numerical failure in solving differential equations are disadvantages of these models.

QSS models rely on the steady-state assumption for the system where the flow re-dispatch of the network is calculated based on power flow (PF) analysis. The main difference among QSS models is the choice of the PF model they incorporate in their simulation. Most of the QSS models use DC approximation due to its fast and guaranteed convergence to calculate redistribution of flow after line trips. However, this comes at the expense of assuming flat profiles for voltage and thus being unable to capture voltage-related failures. Full ACPF is also incorporated in several QSS models however, the convergence of ACPF is a challenging issue especially when many lines go offline during the escalation of CF.

3.4 Uncertainty analysis for line flow

Uncertainties coming from different sources such as renewable generation and loads show different characteristics in terms of magnitude and frequency of occurrence. This work adopted an uncertainty model proposed in [82] that represents the injection power from each component (i.e. generator output power and load demand power) with two terms as shown in (14).

$$P(t) = \mu_P(t) + \epsilon_P(t) \tag{14}$$

where $\mu_P(t)$ is the time-varying mean of the power signal or in other words it is what we expect to have for each component ahead of time and $\epsilon_P(t)$ which is a zero mean signal, representing the uncertainty that may come from forecast error or mismatch in output power for conventional generators. Note that in this study, the output power of generators including conventional and wind generation and demand power from loads are modeled with the above representation. Figure 11 shows this representation for output power signal of a wind generator as an example.

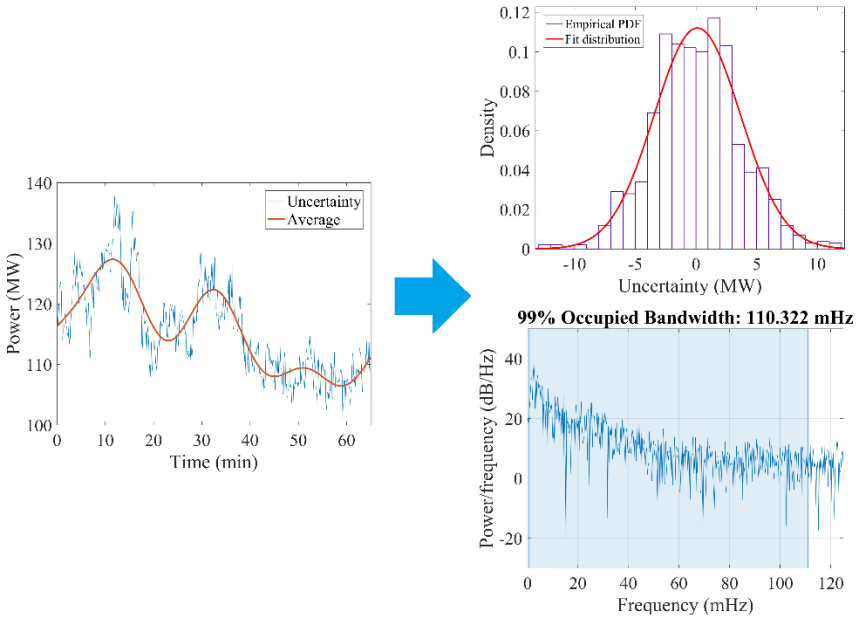


Figure 11. Uncertainty representation and characterization based on its frequency components and PDF for a selected wind generator

3.4.1 Normality verification for line flow

In the line outage model presented in the next section, the line overload distance is calculated based on the Gaussian assumption for line flow uncertainty as proposed by the authors in [24]. Authors in [24] assume Gaussian or normal distribution for the line flows without accounting for wind generation uncertainty in the network. Therefore, it is of interest to verify this assumption for the line flows of a grid with high penetration of wind energy with the uncertainty model proposed

earlier. This enables us to calculate the overload distance for the line flows which will be later used in the line tripping model. The IEEE 300 bus system with 411 transmission branches and several added wind generators is selected to implement the proposed uncertainty model and investigate the normality assumption for the line flows.

The initial observations on the line flows uncertainty distributions reveals a close proximity to normal distribution. There is numerous approach for normality test in the literature each suitable for different needs. *Kurtosis* and *Skewness* coefficients are used widely for normality test for large samples as many other tests such as Kolmogorov-Smirnov test, Jarque-Bera test, and Shapiro-Wilk test almost always reject the null hypothesis of the normal distribution for large samples. This is because when the number of samples gets larger, even the smallest deviation from perfect normality will lead to a significant result. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to normal distribution. Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A perfect normal distribution would have both kurtosis and skewness coefficients equal to zero. However, in [83], [84] the ± 2 range for these measures are introduced as the acceptable range in order to prove normal univariate distribution. We will use this range in the verification step.

The absolute value for both measures are calculated for flows of all 411 lines in the IEEE 300 bus system and then sorted in ascending order. Figure 12 shows the results for the line flows uncertainty kurtosis and skewness. Kurtosis coefficients are sorted in ascending order and the corresponding skewness coefficients are superimposed in the figure to consider both measures simultaneously. It is found that 95% of the lines have kurtosis and skewness coefficients less than 5 with 85% (349 lines) within an acceptable range of 2. This means that the majority of lines satisfy

the normality criteria with a trivial deviation from a perfect normal distribution. Table III summarizes the results of the normality test for grid lines based on kurtosis measure. It is also found that 59% of the lines that are outside of the acceptable bound (with coefficient larger than 2) are connected directly to generation buses as shown in Table III. The statistical analysis on line capacities for the actual grid data from the Federal Energy Regulatory Commission (FERC) suggests that these lines tend to have larger loading margin compared to other lines. Hence, their overload probability is relatively low. It implies that we can use Gaussian assumption for the line flows in the line outage model considering the fact that the lines with larger coefficients will not frequently get tripped and don't have much impact on the proposed stochastic model.

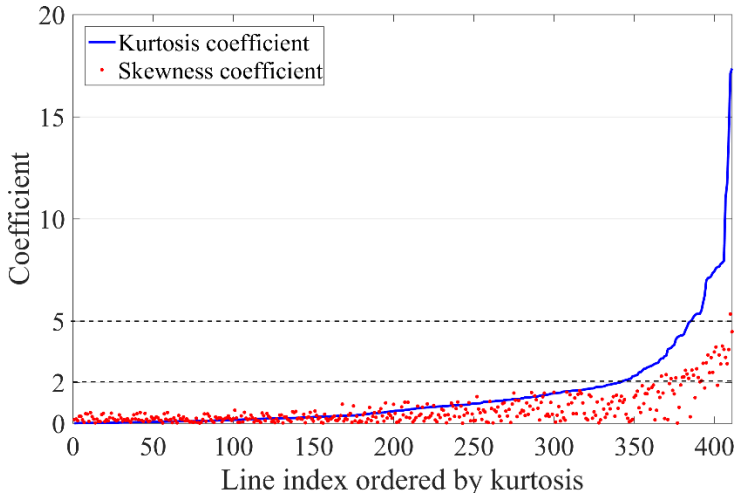


Figure 12. The IEEE 300 bus system line flow uncertainty normality test results.

Table III. Summary of normality verification for IEEE 300 system line flows

	Total Count	Generation connected	Load connected	Connection
Normal lines	349	49 (14%)	199 (57%)	101 (29%)
Out of normal bound lines	62	37 (59%)	20 (33%)	5 (8%)
Total	411	86 (21%)	238 (58%)	87 (21%)

4 Proposed DC cascading failures model

In this chapter, the cascading failures model proposed based on DC power flow is introduced. The assumptions made for model development and flow process estimation will be discussed. The simulation results based on the proposed method will be presented and the impact of high penetration of wind energy as a common form of renewable energy will be examined based on the proposed DC model and single N-2 contingency.

4.1 AC and DC power flow models

The power flow problem is the computation of voltage magnitude and phase angle at each bus in a power system under balanced three-phase steady-state conditions. As a by-product of this calculation, real and reactive power flows in equipment such as transmission lines and transformers, as well as equipment losses, can be determined. For a power grid with *N* nodes, the nodal equations for a power system network enforced by Kirchhoff's law are written as

$$\boldsymbol{I} = \boldsymbol{Y}_{bus}\boldsymbol{V} \tag{15}$$

where $\boldsymbol{I}$ is the *N* vector of source currents injected into each bus and $\boldsymbol{V}$ is the *N* vector of bus voltages, and $\boldsymbol{Y}_{bus}$ is the network admittance matrix. Then, the complex power delivered to bus *k* can be written as $S_k = P_k + jQ_k = V_k I_k^*$, I_k^* being the conjugate of the injected current at bus *k*. By taking the real and imaginary parts of the power balance equation and doing some simplifications, the nonlinear power flow equations are given by

$$P_k = V_k \sum_{n=1}^{N} Y_{kn} V_n \cos(\delta_k - \delta_n - \theta_{kn}) = P_{G_k} - P_{D_k} \tag{16}$$

$$Q_k = V_k \sum_{n=1}^{N} Y_{kn} V_n \sin(\delta_k - \delta_n - \theta_{kn}) = Q_{G_k} - Q_{D_k} \quad (17)$$

where $S_{G_k} = P_{G_k} + jQ_{G_k}$ is the generation and $S_{D_k} = P_{D_k} + jQ_{D_k}$ is the load demand at bus *k*. These nonlinear power balance equations are solved using iterative methods such as the Newton-Raphson algorithm [75].

The DCPF approximation is a standard approach widely used in the literature for assessing line overloads without the need for solving the full AC equations. The DCPF assumes flat voltage profiles for the entire network and neglects the reactive power. It also assumes that the angular separation across any transmission line is small enough so that $\sin(\theta_i - \theta_j) \approx \theta_i - \theta_j$. These assumptions lead to the linearization of Eqs. (16) and (17) that can be solved without the need for iterative approaches.

4.2 Tripping mechanism

Here we present the proposed tripping mechanism that allows for consideration of uncertainty injected from highly variable renewable energy sources. The proposed method consists of stochastic line overloading model and thermal overloading relay simulation.

4.2.1 Stochastic line overloading model

Based on the adopted uncertainty model presented in chapter 2, we present each power signal in the network (e.g. injected power and line flow) with two terms accounting for time-varying mean and uncertainty. Next, with a Gaussian assumption for the distribution of $F_l(t)$ in [24] we can estimate the overloading probability of lines and ultimately derive an average overloading time

that will be used in the proposed tripping mechanism combined with the thermal balance of overhead lines. The following equations show the process to obtain average overloading time based on the model presented in [24]

$$\rho_l(t) \cong Q(a_l) \tag{18}$$

$$a_l = \frac{F_l^{max} - \mu_{F_l}(t)}{\sigma_{F_l}(t)} \tag{19}$$

Where $\rho_l(t)$ is the overloading probability, a_l is the *normalized overload distance* of the *l*th line and Q-function as $Q(x) = \int_x^\infty e^{-t^2/2}/(\sqrt{2\pi})dt$. $\mu_{F_l}(t)$ is the average flow at time t and $\sigma_{F_l}(t)$ is the variance of the flow process at time t. Finally, using the normalized overload distance (a_l) and overloading probability (ρ_l) for each line we can calculate the mean overload time for flow process $F_l(t)$ as follows [24]:

$$\bar{\tau}_l^u = \frac{2\pi\rho_l e^{a_l{}^2/2}}{BW_l} \tag{20}$$

where BW_l is the equivalent bandwidth of the flow process for the *l*th line and can be calculated using the spectral power density (SPD) of the flow process discussed earlier [82].

4.2.2 Thermal overloading relays

The trip time of thermal overload relays is determined based on the maximum allowable current flowing in the conductor without causing thermal instability. Generally, the overload relays for HV transmission lines have time-dependent tripping characteristic, which is determined using the well-known dynamic thermal balance between heat gains and losses in the conductor [85]. The

maximum or hot spot temperature determines the time to trip for thermal relays and considering initial operation current and applying necessary changes, the time to trip can be calculated using a variation form of the tripping mechanism introduced in [86] in which we replace the current with the flow measured in per unit value assuming a flat voltage profile V=1.0 p.u. across the whole network, which is valid for the DC flow analysis.

$$t_{tr} = T_{th}.\ln\left(\frac{F^2 - {F_{op}}^2}{F^2 - {F_{max}}^2}\right) \tag{21}$$

where F is overloaded line flow (p.u.), F_{op} is initial operating flow (p.u.), F_{max} is the line flow threshold, and T_{th} is thermal time constant which is related to conductor type and environmental parameters such as wind speed and ambient temperature [87]. In this study, it is assumed that all transmission lines use typical HAWK (477 kcmil) ACSR conductor with T_{th}=450 sec.

In the proposed CF models, for the tripping mechanism, both relay time to trip and overloading probability are considered simultaneously to select the most probable line trip during the escalation phase of CF. At every time step, first, the time to trip for each overloaded line is calculated, then using normalized overload distance (a_l) and overloading probability (ρ_l) the mean overload time ($\bar{\tau}_l^u$) is determined. If relay time to trip is larger than the mean overload time, the trip timer is set to zero, otherwise, the trip timer is set to the relay time to trip. This tripping mechanism enables us to model the stochastic process of CF and identify the most probable path for its propagation.

After every line trip, the topology of the grid changes and so as the flow distribution across the grid network. Therefore, some new lines may become overloaded and some of the overloaded lines may not be overloaded anymore. Therefore, an update of line states after every line trip in the relay

tripping time is necessary. The time to trip of each overloaded transmission line is determined according to (14) and then after every line outage, the time to trip for other overloaded lines are updated. Note that updating stage for trip time considers the overload duration for each relay from the first overloading instant. In other words, the relay model is with memory, since the overloaded lines are already heated up due to excess power flowing through them and the new time to trip accounts for the gained heat. This concept is illustrated in Figure 13.

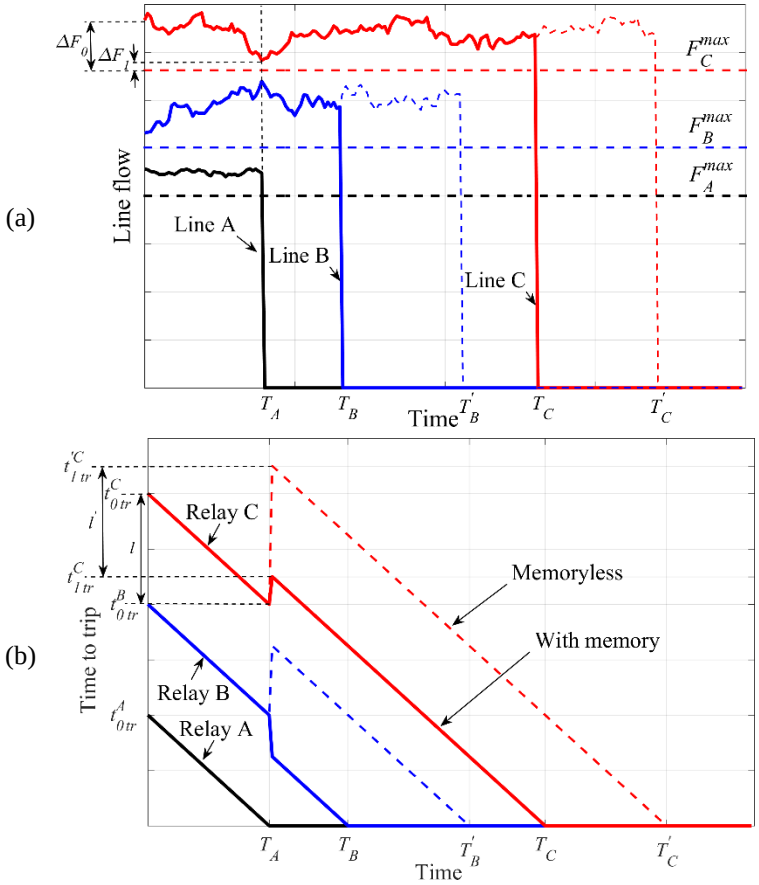


Figure 13. Memory effect in relay time to trip calculation: a) line flows for 3 different lines, b) relay time to trip for the corresponding lines

Figure 13 (a) shows the power flow of three different transmission lines each overloaded initially. The time to trip for each relay is shown in Figure 13 (b). According to the tripping mechanism explained earlier, after the first trip ($t = T_A$) all other relays need to update their timers. With memory effect assumption for relay operation, the new time to trip for relay C is $t_{1}^{C}{}_{tr}$. Note that

this updated time to trip is smaller than that of memoryless operation ($\acute{t}_{1\,tr}^{C}$). Also note that for relay B, the overload flow at $t = T_A$ is larger than initial overload flow at $t = 0$, hence the new time to trip is smaller while the overload flow for relay C at $t = T_A$ is smaller than initial overload flow at $t = 0$ which means larger time to trip. However, for both cases with memory effect for relay operation, the new time to trip equals the time to trip calculated by Eq. (21) minus the total time duration from the first overload instant till the updating time.

4.3 Island detection and power balance

Successive line tripping during the escalation phase of CF usually causes the formation of several islands in the power network. The electrical frequency of the system is driven based on the power balance according to the well-known electro-mechanical equation [88]

$$\sum P_{Gen} - \sum P_{Load} = H.2\pi \frac{df}{dt} \tag{22}$$

where $\sum P_{Gen}$ is the total produced power, $\sum P_{Load}$ is the total consumed power, H is the global inertia, and f is the electrical frequency of the system. The frequency in power systems is considered a global parameter and cannot be influenced by a small section. However, when a part of the network becomes islanded, the inertia and the load balance depend only on generators and loads inside the island where shedding actions may be necessary. Therefore, it detecting the formed islands during cascading failures process is a critical component of a CF simulation model that allows to model the corrective actions that is likely to take place in the real time operation of power systems.

4.3.1 Island detection method

An automatic island detection algorithm inspired by the approach proposed in [88] is used after each trip to identify newly formed islands. Clusters of the generator(s) and the load(s) that are not connected to the grid are called island(s). The algorithm consists of three steps; connectivity check, critical events identification, and island identification. Connectivity check determines how many islands are present in the power system, and their structure. Critical event detection identifies which breakers should create an island if opened. And the final step, island detection, identifies the buses belonging to each possible island and calculates their load balance. The actual dispatch of the network is not required for the algorithm since it only depends on gird topology (grid incidence matrix A) and generators location. Assuming the resistance for all lines in the network equal to 1 Ω and using the Kirchhoff's current and voltage laws, the equations describing system behavior are solved. To detect buses belonging to each island, generators are activated (assuming output current of 1 A) one at a time. After identifying all present separate islands in the grid, their power balance is maintained by shedding actions. For detailed island detection algorithm please refer to [88].

4.3.2 Power balance algorithm

In the power balance algorithm for any island, the total load and total generation capacity are compared to each other. If the total demand exceeds the maximum available generation, some load shedding is necessary to maintain the power balance. Similarly, if total demand is smaller than the current generation, one or several generation units should drop their generation. The flow chart of the automatic power balance algorithm is shown in Figure 14. Suppose that there are *k* separate islands in the grid at $t = t_0$. If a line trip at $t = t_0 + \Delta t$ results into the formation of a new island, it is necessary to run the power balance algorithm for both newly formed island and the mother

island that it separated from. Therefore, the power balance algorithm will balance the generation and load for the two clusters for the next power flow solution.

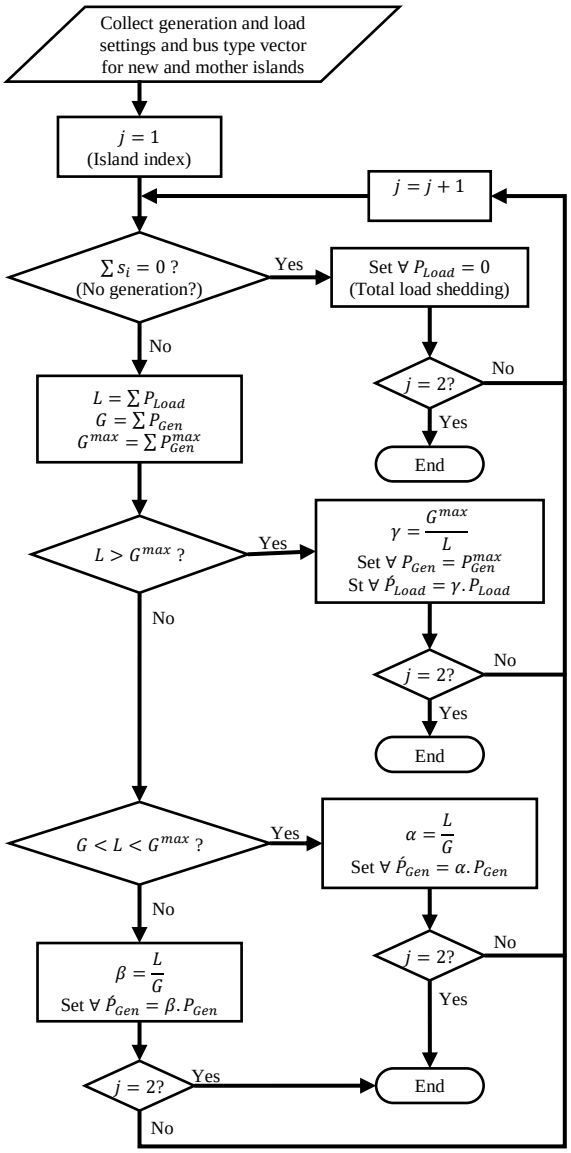


Figure 14. Flow chart of the power balance algorithm for newly formed islands in the power grid

The power balance algorithm starts with collecting generation and load settings of the two new islands. The bus type vector for each island is defined as $\boldsymbol{S} = \{s_1, s_2, \cdots, s_m\}$ where $s_i \in [0,1]$. For each bus s_i=1 represents a generation bus and s_i=0 is either load or connection bus. If there is no

generation bus in each island ($\sum s_i$=0), the total cluster is de-energized and the algorithm will cut the total load of the island. Otherwise, a comparison is made between the total load and the total generation and maximum available generation to balance the power accordingly to minimize the total load loss. Note that generation or load adjustment is distributed meaning that balancing adjustment is applied as a percentage to every generator or load.

4.4 Line flow process estimation

The most critical parameter in the simulation of cascading failure path and impact is the line flow process. The accurate estimation of the line flow process will have a big impact on overall grid vulnerability evaluation under different operating conditions. Moreover, the high level of uncertainty injected from renewable energy generation will directly affect the line flow estimation. Because of the nature of the random nature of renewable energy generation as well as the uncertainty coming from electrical load prediction, the line flow process can be considered as a random variable with variable mean and covariance. Earlier we assumed a Gaussian model for this random variable and showed that given the mean and covariance matrix, we can estimate the overloading status of the line during cascading failure. Therefore, the estimation of line flow means the estimation of its statistical parameters, mean and covariance. In this section, we will discuss how the flow process is estimated for DC and AC cascading failure model.

4.4.1 Line flow process based on DC power flow

The linear property of DCPF allows us to express the statistics of the flows as a linear function of the statistics of the operating conditions. If time-varying mean and covariance of the injected power are written as:

$$\mu_P(t) = \begin{bmatrix} \mu_g(t) \\ -\mu_l(t) \end{bmatrix}, \quad \Sigma_P(t,\tau) = \begin{bmatrix} \Sigma_{gg}(t,\tau) & \Sigma_{gl}(t,\tau) \\ \Sigma_{lg}(t,\tau) & \Sigma_{ll}(t,\tau) \end{bmatrix} \tag{23}$$

then, the line flows mean and covariance are [24]:

$$\mu_F(t) = \sqrt{y_t}\left(\tilde{A}_t^{\,T}\right)^{\dagger} \mu_P(t) \tag{24}$$

$$\Sigma_F(t,\tau) = \sqrt{y_t}\left(\tilde{A}_t^{\,T}\right)^{\dagger} \Sigma_P(t,\tau)\left(\tilde{A}_t\right)^{\dagger}\sqrt{y_t} \tag{25}$$

where $\sqrt{y_t} = diag\{\sqrt{y_l(t)}\}$ is the diagonal entries of the square matrix of line admittances and $\tilde{A} = \sqrt{y}A$ with A the line-node incidence matrix. Here $(.)^{\dagger}$ represents pseudo inverse [24].

The variance for the flow process of each line can then be calculated by taking square root of each diagonal element in the covariance matrix with $\sigma_{F_l}(t) = \sqrt{\Sigma_{F_{l,l}}(t,0)}$.

4.5 Mixed OPF-stochastic DC CF model

In this section, the simulation results based on the proposed mixed OPF-stochastic DC CF model are presented. Also, based on a single N-2 contingency simulation in a standard 300 bus system with different wind integration scenarios, the preliminary analysis of the impacts of penetration of renewable energy on grid vulnerability will be evaluated.

4.5.1 Cascading failure simulation with the DC model

Three different scenarios are considered to study the impact of wind uncertainty on grid vulnerability and for each of them, some of the conventional generators in the original IEEE 300 bus system are replaced with wind farms. In the first and second scenarios 11 conventional

generators are replaced with wind generators at buses 80, 88, 125, 128, 156, 199, 222, 256, 258, 262, 295. In the third scenario, 6 conventional generators are replaced with wind farms. In this scenario, as the wind generator capacity increases the capacity of the corresponding conventional generator on the bus decreases proportionally to maintain a fixed total generation capacity. The load and generation data are received from the Electric Reliability Council of Texas (ERCOT) with 4-second sampling rate. The high sampling rate for data allows us to capture high-frequency dynamics of different sources. These data are for numerous wind generators, load demand, and different types of conventional generations. In this study, fossil fuel generators and co-generation are used as conventional generation and load data are scaled according to the original settings of the IEEE 300 bus system. Uncertainty modeling of loads and generations are based on the model proposed earlier in chapter 2. The Autoregressive Moving Average (ARMA) forecasting technique is employed to model the initial uncertainty signal coming from wind generation and electrical loads based on the actual data. The increased uncertainty level models the use of different forecasting techniques and horizons with different accuracy and characteristics and illustratively shows how the accuracy of the forecasting method affects the results. Simulations of the CF scenarios are performed in the MATLAB environment and MATPOWER is used for OPF and PF calculations [89]. Wind installation settings are shown in Table IV for the three studied scenarios. Note that wind generation capacity is selected the same as the IEEE-300 bus system original setting for conventional generators in first and second scenarios.

Table IV. Wind generation locations for simulation scenarios

	Penetration level	Wind generation location (bus number)
First scenario	$\boldsymbol{\alpha = 0.036}$	A={80,88,125,128,156,199,222,256,258, 262,295}
Second scenario	$\boldsymbol{\alpha = 0.2}$	B={98,120,170,215,249,265}
Third scenario	$\boldsymbol{\alpha = 0.036}$	$\acute{A}$= A
	$\boldsymbol{\alpha = 0.05}$	$\acute{A}$={ A,69}
	$\boldsymbol{\alpha = 0.09}$	$\acute{A}$={ A,69,131,169}
	$\boldsymbol{\alpha = 0.105}$	$\acute{A}$={ A,69,131,169,254}
	$\boldsymbol{\alpha = 0.125}$	$\acute{A}$={ A,69,131,169,254,260}
	$\boldsymbol{\alpha = 0.150}$	$\acute{A}$={ A,69,131,169,254,260,215}
	$\boldsymbol{\alpha = 0.173}$	$\acute{A}$={ A,69,131,169,254,260,215,248}
	$\boldsymbol{\alpha = 0.223}$	$\acute{A}$={ A,69,131,169,254,260,215,248,122,255}
Forth scenario	$\boldsymbol{\alpha = 0.0}$	B={98,120,170,215,249,265}
	$\boldsymbol{\alpha = 0.1}$	B={98,120,170,215,249,265}
	$\boldsymbol{\alpha = 0.2}$	B={98,120,170,215,249,265}
	$\boldsymbol{\alpha = 0.3}$	B={98,120,170,215,249,265}

The initial operating equilibrium and conditions $(G(0), L(0), \theta(0), F(0))$ are taken or derived from the power flow solution. The equivalent bandwidth of the flow process for each line under the initial uncertainty level is then calculated and stored to use later on stochastic tripping mechanism. Since the original setting of the IEEE 300 bus system does not provide enough information on line capacities, they are set as $F^{max} = max\{\eta|F(0)|, 2.0(p.u.)\}$ with η=1.20. Here we take $F(0)$ as the rational flow distribution under normal operating conditions and assume that the line capacity allows a load increase up to 20% [24]. Note that we select a near congestion operating conditions for the grid to better see the impact of increased uncertainty from wind generation on multiple line overloads leading to CFs. The minimum of line capacity is set to be 2.0 p.u. so that the vibration in the lines which usually carry small flows will not cause frequent line trips.

A) Wind uncertainty level and grid vulnerability

The first and second scenarios are considered to study the impacts of forecasting relative error which comes into the picture in the form of uncertainty from wind generation. For the first scenario, the uncertainty signal magnitude for the wind generator is increased by factor $\gamma = \frac{{\epsilon_w}^{new}}{{\epsilon_w}^{int}}$ where ${\epsilon_w}^{new}$ is the new uncertainty of wind power and ${\epsilon_w}^{int}$ is the initial uncertainty.

In the first scenario to see the impact of larger forecasting errors on grid vulnerability to overload CFs, γ is increased from 1 to 5 with 0.25 steps to find the uncertainty level in which the first CF occurs. All other settings of the system remains the same during the first scenario. Table V shows the results for increased wind uncertainty level in the first scenario.

Table V. Cascading Failure results for the first scenario: wind uncertainty level

γ	Total trip count	# of formed islands	Total LS [a] (MW)	LS [a] (%)	First tripped line [b]	Second tripped line [b]
2.00	0	0	0	0.0	-	-
2.25	0	0	0	0.0	-	-
2.50	68	18	7941	32.0	365 @ 54.7	205 @ 55
2.75	70	19	8344	33.6	99 @ 44.8	205 @ 44.9
3.00	71	19	8496	34.2	117 @ 44.8	207 @ 44.9
3.25	72	20	8966	36.1	365 @ 44.6	205 @ 45
3.50	74	21	9135	36.8	117 @ 30.4	205 @ 30.5
3.75	75	21	9397	37.8	115 @ 29.2	205 @ 29.3
4.00	80	22	9749	39.2	365 @ 29.2	205 @ 29.4
4.25	83	23	10005	40.3	365 @ 28.2	205 @ 28.3
4.50	83	23	10092	40.6	205 @ 24.9	117 @ 25.1
4.75	83	24	10495	42.3	365 @ 15.8	205 @ 16.1
5.00	86	25	10665	43.0	99 @ 15.6	205 @ 15.7

[a]Load Shedding
[b]Line number @ time (min)

For γ between (1-2.25), there is no tripped line thus no CF happens for this uncertainty range. Moving beyond γ=2.25 multiple line overloads are observed that leads to a series of CFs that forms multiple islands and isolated buses. Automatic power balance on each island causes the load to be dropped to a certain level that can be supplied by generators inside the island. Successive line trips

continue until all line flows drop safely below line thresholds. Also, the first and second tripped lines and their respective times are given to identify the most vulnerable lines in the network for a given wind uncertainty.

γ is increased further to see the impacts of even larger uncertainty levels on the severity of CF. Figure 15 shows the total number of tripped lines and total load shedding percent for different wind uncertainty levels in the first scenario. As the γ increases, the more lines get tripped during CF which in turn leads into the formation of more islands and larger load shedding as shown in Figure 15.

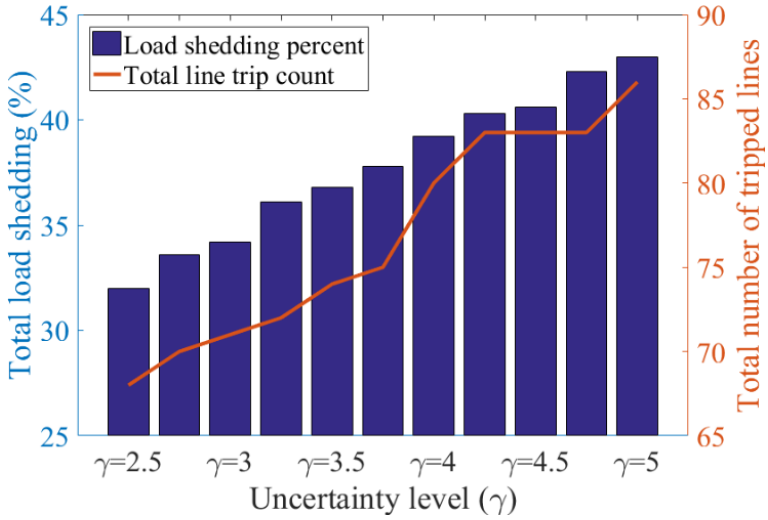


Figure 15. The total number of line trips and total load shedding versus uncertainty level (γ) for the first scenario

The evolution process of CFs for different wind uncertainty level is shown in Figure 16 for the first scenario. All the curves are comparable to actual failures recorded in history and reported in [22]. Each evolution curve consists of two phases, the escalation phase in which the line trip rate is as high as 12 lines per minute, and the damping phase with line trip rate of approximately one line per minute. Also, it is found that as wind uncertainty level increases, the first trip happens earlier than lower wind uncertainty level which indicates that the minimum safety time of the entire

network decreases under the same operating conditions. For example, the black bold line shows the cumulative number of line trips for uncertainty level increased by a factor of 5. As compared with uncertainty level increased by factor 2.5 (green line with square marker), the former results into a higher number of tripped lines due to the high level of wind uncertainty. Also, high uncertainty level causes contingency in multiple lines earlier compared to lower uncertainty levels. For example, the earliest cascading process is associated with the highest uncertainty level, γ=5, as indicated in Table IV and happens after 15 minutes of the beginning of the simulation, which implies that as more uncertainty is injected to the grid, its survival time gets shorter.

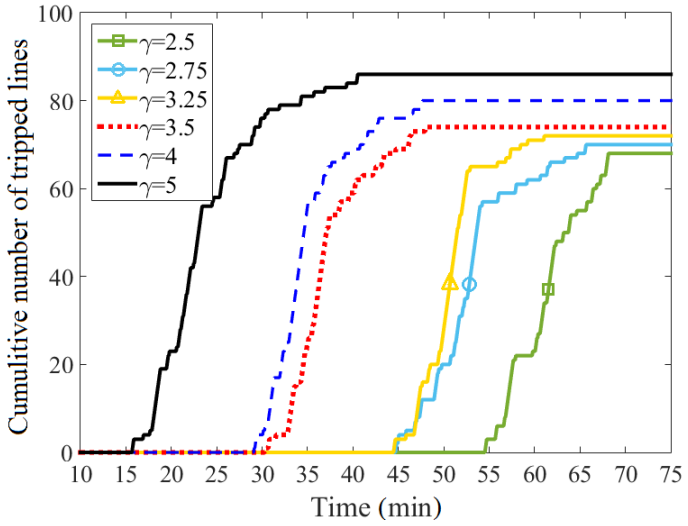


Figure 16. The evolution process of CFs for different wind uncertainty levels for the first scenario

For the second scenario, 6 wind generators are installed on the network accounting for 20% penetration ratio. Note that for this scenario, the total generation capacity of the grid before and after installation of the wind generators are the same. Next, a single N-2 contingency applied to initiate CF under various levels of wind uncertainty. Figure 17 shows the total number of tripped lines and total load shedding percent for different wind uncertainty levels in the second scenario. The total load shedding is almost the same for wind uncertainty level increased by up to 3.5 fold.

However, the load shedding dramatically increases as the wind uncertainty level increases by 4.5 fold. The same pattern is visible for the number of line trips with various wind uncertainty level.

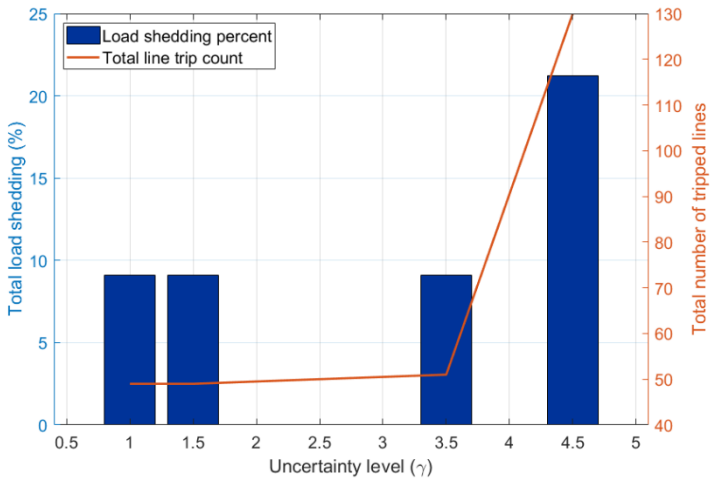


Figure 17. The total number of line trips and total load shedding versus uncertainty level (γ) for the second scenario

The evolution process of CFs for different wind uncertainty level is shown in Figure 18 for the second scenario. As we increase the wind uncertainty level, the line trips happen earlier before applying N-2 contingency. For the highest uncertainty level, $\gamma = 4.5$, the cascading trips escalades very quickly and propagates to a larger portion of the grid.

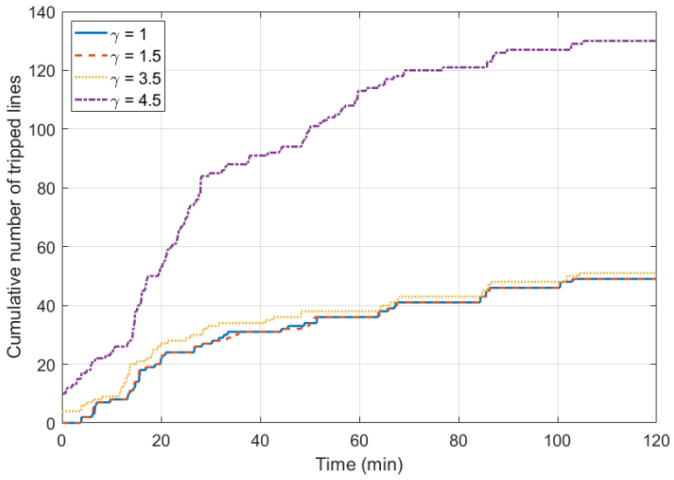


Figure 18. The evolution process of CFs for different wind uncertainty levels for the second scenario

B) Wind penetration level and grid vulnerability

The third and forth scenarios aim to investigate the impacts of increased penetration level of wind energy on grid vulnerability to cascading overload failures. For these scenarios, wind penetration ratio is defined as $\alpha = \frac{\sum P_{G,wind}^{max}}{\sum P_{G,total}^{max}}$, where $\sum P_{G,wind}^{max}$ is the total wind generator capacity and $\sum P_{G,total}^{max}$ is the total grid generation capacity. For the third scenario, by replacing more conventional generators from the original setting of the network with wind generators in addition to those already installed, α is increased to see the impacts of higher wind penetration on grid vulnerability. Note that small to medium generators are selected to be replaced with additional wind farms to have smaller steps for α. All other settings of the system remains the same.

The results for increased wind penetration for the third scenario are shown in Table VI. Starting from initial α=0.036, there are no line trips until α=0.09 where line 137 gets tripped at minute 2. To see further impacts of higher wind penetration, α is increased to 0.223 by replacing more mid-size conventional generators with wind generators. Figure 19 shows the total number of trip lines and load shedding for each penetration level α for the third scenario. It is found that the higher the wind penetration ratio, the more line trip and load shedding occurs in the network. In other words, given the same settings for all other generations and loads of the network, under congestion conditions in the network, installing more wind farms increases the risk of blackout due to cascading overload failures.

Table VI. Cascading failure results for the third scenario: wind penetration level

α	Total trip count	# of formed islands	Total LS [a] (MW)	LS [a] (%)	First tripped line [b]	Second tripped line [b]
0.036	0	0	0	0	-	-
0.05	0	0	0	0	-	-
0.09	68	20	8149	33	137 @ 2	101 @ 3.8
0.105	72	21	10156	41.1	137 @ 2	274 @ 3.8
0.125	75	23	10701	43.3	137 @ 2	83 @ 2.7
0.15	81	24	10890	44	137 @ 2	83 @ 2.5
0.173	94	27	11650	47.2	137 @ 2	83 @ 2.4
0.223	98	30	13958	56.5	137 @ 2	274 @ 2.3

[a]Load Shedding
[b]Line number @ time (min)

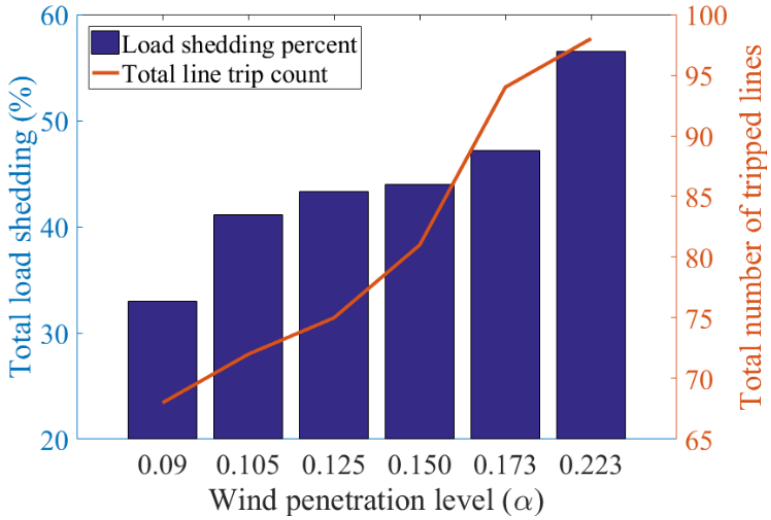


Figure 19. The total number of line trips and total load shedding versus wind penetration level (α) for the third scenario

It is also observed that for all CFs beyond α=0.09, line 137 is the first line getting tripped and it happens almost at the same time for all α above 0.09. This could be explained considering the location of the new wind farm installation. A certain wind farm added to the network at a particular location injects additional uncertainty to one of the backbone transmission lines in the network leading to further line trips and propagation of CFs. However, the second line trip is different for each penetration level which determines the cascading path and eventually the total number of trips and load shedding. This is particularly interesting for planning purposes, since this will detect the most vulnerable lines of the network.

The evolution process of CFs for the third scenario is shown in Figure 20 for each penetration level. The triggering event for every one of them is the same, however, due to a different level of uncertainty coming from wind generation, each failure evolves into a different path. This potentially identifies the weakest backbone line of the network according to the new configuration of wind generators. For example, the black bold line shows the cumulative number of line trips for 9% wind penetration level where starts nearly 5 minutes after the simulation start and stabilizes at minute 28. While increasing wind penetration to 22.3% (green line with square marker) results into higher tripping rate, as large as 20 lines per minute during escalation phase of cascading failure, and the more total number of tripped lines.

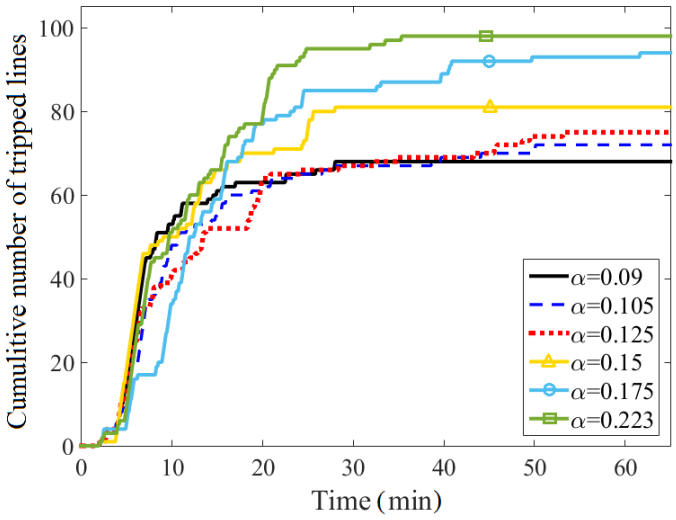


Figure 20. The evolution process of CFs for different wind penetration levels in the third scenario.

For the forth scenario, the wind penetration ratio, α, increases from 0 up to 0.3 by increasing the capacity of installed wind generators and reducing the capacities of conventional generators proportionally. This scenario is designed to examine the case where the installation locations are fixed while the penetration increases. A single N-2 contingency initiates the cascading failure in the simulations. The results for increased wind penetration for the forth scenario are shown in

Table VII. Starting from $\alpha = 0$, no cascading event happens in the grid and consequently no load shedding results at the end of simulation. As the capacity of wind increases, more lines get tripped and more load shedding occurs in the grid at the end of the simulation. Note that, for all wind penetration levels, a same N-2 contingency is applied to the grid at the same time to initiate the cascading failure.

Figure 21 shows the total number of trip lines and load shedding for each penetration level α for the forth scenario. Similar to the third scenario, it is found that the higher the wind penetration ratio, the more line trip and load shedding occurs in the network.

Table VII. Cascading failure results for the forth scenario: wind penetration level

α	Total trip count	# of formed islands	Total LS[a] (MW)	LS (%)	First tripped line[b]	Second tripped line
0	7	3	0	0	88 @ 4	194 @ 4
0.1	43	8	50.70	0.20	88 @ 4	194 @ 4
0.2	49	12	2231	9.07	88 @ 4	194 @ 4
0.3	116	26	4638	18.86	88 @ 4	194 @ 4

[a]Load Shedding
[b]Line number @ time (min)

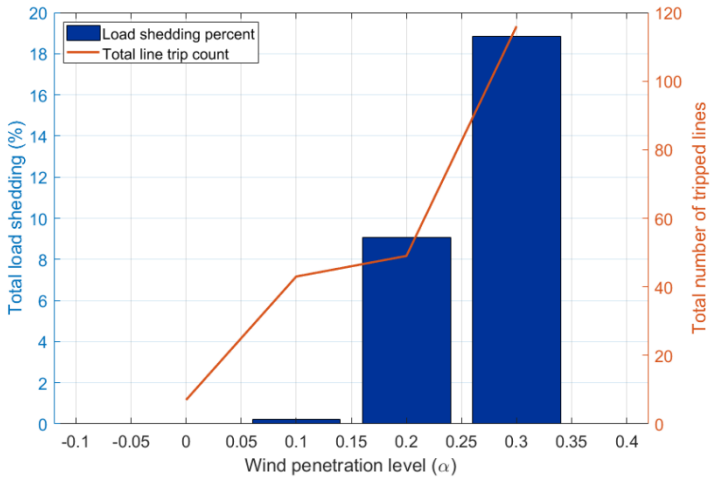


Figure 21. The total number of line trips and total load shedding versus wind penetration level (α) for the forth scenario

The evolution process of CFs for the forth scenario is shown in Figure 22 for each penetration level. The triggering event (N-2 contingency) for every one of them is the same, however, due to a different level of uncertainty coming from wind generation, each failure evolves into a different path. As can be seen from the figure, the increased wind penetration results into more tripping event and consequently more islands will be formed in the network as shown in Table VII.

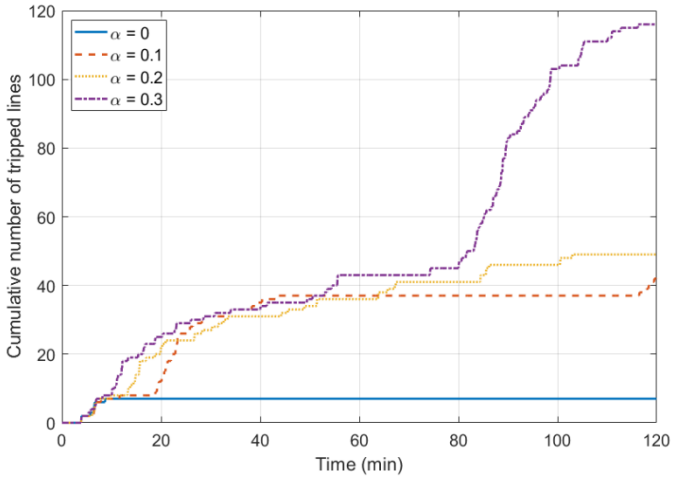


Figure 22. The evolution process of CFs for different wind penetration levels in the forth scenario

C) Overload distance analysis during cascading failure

As mentioned earlier, the overload distance of line flows can be a good indicator of the behavior of the power grid during normal operation and CF. The *Mahalanobis* overload distance (D_m) shows the overload distance of the whole network and considers the correlation between line flows using flow covariance matrix, while the *Euclidian* overload distance (D_e) assumes no correlation among the line flows and considers them as independent random variables. Figure 23 shows both D_m and D_e during normal operation of the system (without an increase in wind uncertainty) and CF resulting from increased wind uncertainty by the factor γ=4.5.

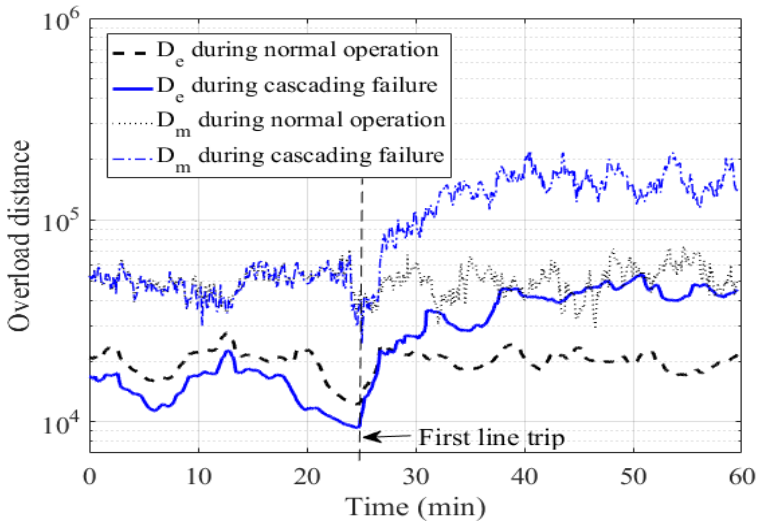


Figure 23. Mahalanobis and Euclidian overload distance during normal and CF in the IEEE 300 bus system.

From Figure 23 we can see that *Mahalanobis* overload distance (D_m) is always greater than *Euclidian* overload distance (D_e) either for normal operation or during the CF. This verifies the fact that in an actual power network, the line flows have a strong correlation with each other that causes a more robust and reliable operation compared to independent line flows. The first line trip happens at t=25 min. From t=0 to first tripping instant D_e is smaller for the case leading to CF. This is because of increased wind uncertainty compared to normal operation. In other words, increased uncertainty in line flows results into smaller overload distance and consequently higher chance for CF in the power system. Another interesting finding is that both D_m and D_e show an increasing trend as line tripping spreads throughout the grid. This is because of the load shedding actions to maintain power balance for newly formed islands in the network. As more generation units are separated from formed islands, larger portions of electrical loads get curtailed which means a reduction in line flows and according to $a_l = \frac{F_l^{max} - \mu_{F_l}(t)}{\sigma_{F_l}(t)}$ the overall distance of line flows from their threshold increases. This also explains why all CFs tend to stop after several line trips if appropriate load shedding mechanisms were employed in the operation and control of the grid.

4.5.2 Discussion on the results of the DC model

Two scenarios are considered to study the impact of wind generation uncertainty on CFs using the DC proposed model. First, it is found that increased uncertainty injected from wind generation could cause cascading failures in the grid and the higher the injected uncertainty the more severe the situation in terms of the total number of tripped lines and load shedding. Second, our analyses show that given the current operating condition of the grid, increasing wind penetration to a certain level may result in cascading overload failures and higher penetration makes the grid more vulnerable to failures. In addition, overload distance of the network as a measure of grid safety to CF is analyzed for normal and contingency operating conditions. Simulation results suggest that appropriate management of uncertainties via energy storage or advanced forecasting techniques is necessary in order to achieve sustained growth of renewable generation in current grid operation.

Despite all the advantages that DC approximation of power flow offers (e.g. fast and guaranteed convergence of power flow problem), we are not able to simulate and evaluate the voltage-related failures during cascading failure with this model. The voltage related failures could potentially play a significant role in determining the ultimate outcome of a cascading failure event and thus, the results obtained using DC model might not reflect the true impact of renewables on grid vulnerability. For example, in [90] the impacts of intermittent RE resources on voltage dynamics is investigated. In another work in [91], authors study the effects of higher and lower penetration of distributed wind generation on the voltage dynamics in a faulted system. The effects of replacing conventional generation by wind and solar generation on the grid voltage performance are examined in [52] and the issues resulting from such generation alteration are identified. These studies suggest that variability of RE generation can have a big impact on voltage dynamics

therefore, the DC model might underestimate the severity of CF due to its flat voltage profile assumption.

Therefore, our next step is to enhance the proposed CF model by incorporating the full AC power flow model to acquire voltage profiles and simulate under/over voltage relays during CF. Another enhancement that we will make on the proposed methodology involves determining the location of the integration of renewables based on actual GIS information of the test system and renewable growth potential maps.

5 Enhanced cascading failure model based on AC power flow

In this chapter, several enhancements on the proposed mixed OPF-stochastic CF model that has been presented in the previous chapter will be made. The enhancements come in the form of a) revising the way renewable energy resources are being integrated to the grid, b) incorporating full AC power flow model instead of DC approximation, and C) implementing under/over voltage relays to capture voltage related failures during a cascading failure event.

5.1 Modeling integration of renewable energy resources

The renewable generation accounts for 50% of the U.S. new energy installation capacity with the wind energy ranking second after hydropower in terms of percentage of total generation. It is critical to systematically evaluate the impacts this shifting in energy portfolio would have on grid vulnerability. In the DC model presented in the previous chapter, the spatial correlation of installed wind generators has been neglected and it was assumed that every new wind installation randomly replaces a conventional generator in the original setting of the network. In this paper, a methodology for more realistic modeling of wind penetration to the bulk energy systems (BES) by integrating geographical information of the network topology and wind installation potential for a given geographical area is proposed. The wind energy as a form of renewable energy resource will be used in our analyses that has the potential to be integrated into the grid in larger scales, thus having a more significant impact on grid operation. Other forms of renewable energy resources such as photovoltaic will have a similar impact considering its random nature and uncertainty in output power.

To accurately model wind penetration to BES, one should take into account the land potential of wind capacity to simulate the probable generation expansion planning scenarios in the near future. In other words, wind potential capacity maps such as the ones published by National Renewable Energy Laboratory (NREL) can give useful insights on the probable location and capacity of the future wind farms integrating to the existing grid (Figure 24) [92]. The potential wind capacity map shows land area with a gross capacity factor of 35% and higher, which may be suitable for wind energy development. AWS Truepower produced the wind resource data with a spatial resolution of 200-m, which was binned into 20-km grid cells. Map shading shows the amount of area with the potential to be developed within each 20-km cell: the darker the color, the larger the potentially developable area within each cell. These maps exclude areas that cannot be used for development (e.g. wilderness, urban areas, etc.) Potential wind capacity maps are provided for a 2014 industry standard wind turbine installed on a 110-m tower, which represents plausible current technology options.

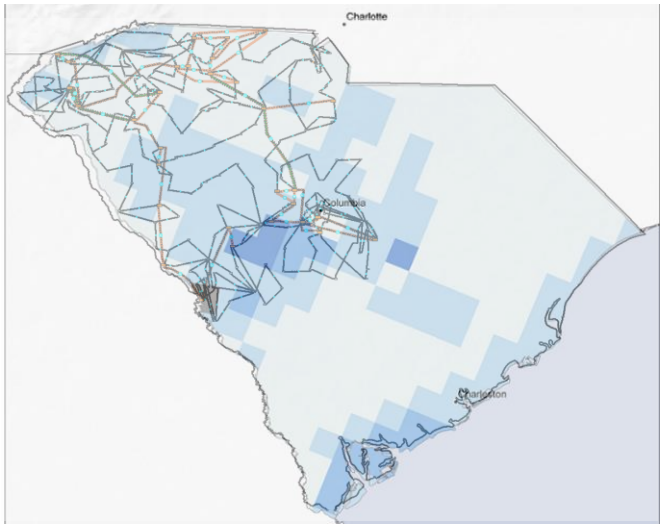


Figure 24. South Carolina 110-meter potential wind capacity map combined with ACTIVSg500 synthetic network [32], [33].

Using the potential wind capacity map and geographical information of the grid topology, we can determine the possible point of integration for potential wind farms. For this, we calculate the direct distance of each potential area from all substations of the grid. Then we choose the closest substation within a predefined radius of the area, say 20 km. Note that, this radius depends on various economic and technical parameters and is worthy of more investigation. Now, we have a potential area and substation pairs for the given network. Next, we sort the list of potential areas based on their maximum potential wind power (MW) installation capacity (Table VIII).

Table VIII. Potential MW of wind integration for substations

Substation ID	Potential MW of Wind	Candidate Buses
4	695	[**8**;7;9]
8	629	[20;**19**]
151	596	[373;**372**;371]
79	552	[190;**189**]
67	505	[159;**158**]

Inside every substation, there are multiple buses that can be selected as the point of coupling for wind farms. For this, we exclude load buses and select either generation or connection buses. Note that for the sake of consistency, we select buses with the same nominal voltage level throughout this process. The candidate buses for South Carolina 500-bus synthetic network are highlighted in Table VIII.

The one-line diagram of ACTIVSg500 network which is a synthetic power system model that does not represent the actual grid is superimposed on wind potential map for South Carolina in Figure 24. The ACTIVSg500 is developed as part of the ARPA-E Grid Data research project and contains no Critical Energy/Electric Infrastructure Information (CEII) [93]. Figure 25 shows the part of the network with a high concentration of wind farms. As an example to demonstrate the proposed

methodology we selected the top four locations to install wind farms. These locations are marked with green rectangles inside red circles.

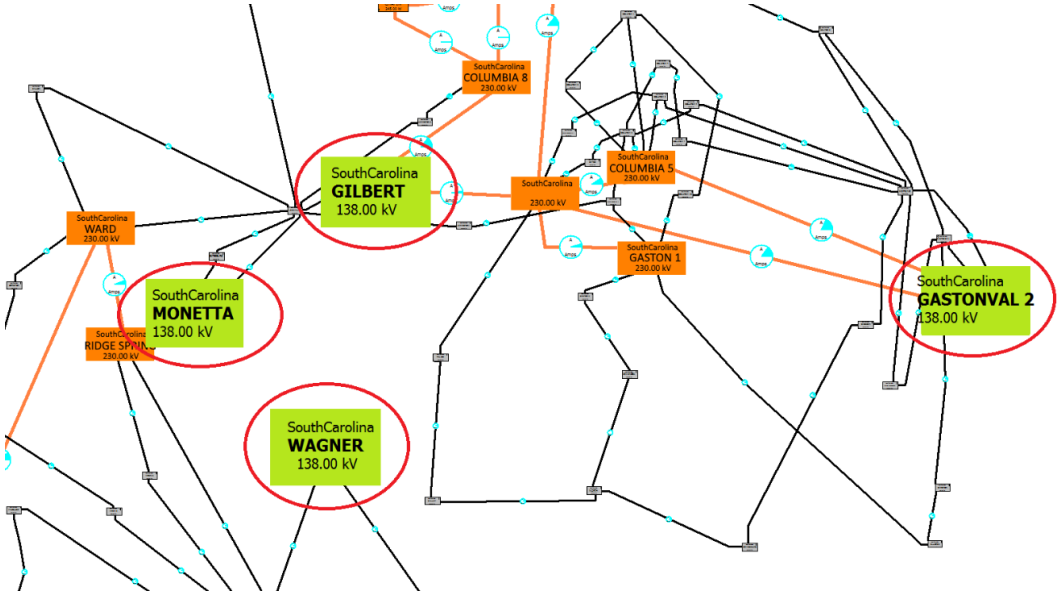


Figure 25. Location of four substations with integrated wind farms

5.2 Mixed OPF-stochastic AC CF model

Aside from high computational burden and non-convergent scenarios in ACPF, the direct estimation of the flow process for stochastic modeling of overloading relay is not possible due to the nonlinear relationship between injected power and line flow as random variables. Therefore, there is a need for a new methodology to estimate flow process statistics for AC model given the time-varying mean and covariance of the injected power to the grid.

The Unscented Transformation (UT) method calculates the statistics of an output random variable undergoing a set of nonlinear transformations (e.g. ACPF) and has been applied to probabilistic power flow problem and state estimation [94], [95]. This chapter presents a novel CF model where the UT method is recursively used to calculate time-varying mean and covariance of the flow

process in full ACPF for stochastic overloading relay simulation under uncertain generation. We call the proposed model AC Unscented Transform (ACUT). The incorporation of ACPF constraints makes it possible to simulate voltage related failures during the cascade of the failures by implementing under-voltage load shedding (UVLS) relays.

5.2.1 Line flow process based on AC power flow

In the stochastic CF model in [42], we define the line flow as a random variable in the form of $F_l(t) = \mu_{F_l}(t) + \epsilon_{F_l}(t)$. $\mu_{F_l}(t)$ is a the time-varying mean of the flow process which is deterministic and $\epsilon_{F_l}(t)$ is a zero mean component that has the same temporal (and spatial) covariance as the flow and can be assumed Gaussian within some small time window $[t_0, t_0 + T]$ [24]. According to (24) and (25), the DCPF enables us to directly model $\mu_{F_l}(t)$ from the time-varying mean of the injected power to the buses, $\mu_P(t)$, which can be estimated with forecasting. However, the linearity of the power flow equations does not hold true for full AC model, thus the direct estimation of flow mean from injected power is not valid. Next, we propose a novel approach to estimate the mean and covariance of the flow process based on the same statistics of the injected power with nonlinear ACPF equation based on the UT method. In the proposed approach, the active power for loads and wind generators are assumed to be random variables while the reactive power is assumed to be a function of the active power by means of randomly selected power factors within a pre-defined range (e.g. 0.9 to 0.95).

5.2.2 Unscented transform

The UT was developed to overcome the demerits associated with linearization process techniques and is applied to different uncertain problems with satisfactory performance [96]. The UT method

calculates the statistics of an output random variable undergoing a set of nonlinear transformations. Because it is easier to approximate a probability distribution than an arbitrary nonlinear function. The UT method produces appropriate samples of the input variables with sufficient information to accurately estimate the statistics of the output variable(s). Assume that $\boldsymbol{P}$ is the vector of n random variables of the injected power (loads and wind farms) with $\overline{\boldsymbol{P}}(t) = \mu_P(t)$ as the mean and $\boldsymbol{\Sigma_{PP}}(t)$ as the covariance. If $\boldsymbol{F} = \psi(\boldsymbol{P})$ is the line flows where ψ is the nonlinear ACPF functions, then the mean and covariance of $\boldsymbol{F}$, $\overline{\boldsymbol{F}}(t)$ and $\boldsymbol{\Sigma_{FF}}(t)$, respectively can be obtained through the following steps. For each time step t:

1) Obtain $2n+1$ samples of $\boldsymbol{P}$ called sigma points using:

$$p_t^0 = \mu_P(t) \tag{26}$$

$$p_t^k = \mu_P(t) + \left(\sqrt{\frac{n}{1-W^0}\boldsymbol{\Sigma_{PP}}(t)}\right)_k, \quad k = 1,2,\dots,n \tag{27}$$

$$p_t^{n+k} = \mu_P(t) - \left(\sqrt{\frac{n}{1-W^0}\boldsymbol{\Sigma_{PP}}(t)}\right)_k, \quad k = 1,2,\dots,n \tag{28}$$

2) Calculate the weights associated with each sigma point using the following:

$$W^0 = W^0 \tag{29}$$

$$W^k = \frac{1-W^0}{2n}, \quad k = 1,2,\dots,2n \tag{30}$$

Note that the associated weights must meet the following condition:

$$\sum_{k=0}^{2n} W^k = 1 \tag{31}$$

$\left(\sqrt{(n/1-W^0)\boldsymbol{\Sigma_{PP}}(t)}\right)_k$ in (7) and (8) is the *k*th row or column of matrix square root of $\left((n/1-W^0)\boldsymbol{\Sigma_{PP}}(t)\right)$. The matrix square root of positive definite matrix $\boldsymbol{X}$ means that there is a matrix $\boldsymbol{Y}=\sqrt{\boldsymbol{X}}$ such that $\boldsymbol{X}=\boldsymbol{YY^T}$ and it is calculated using numerically efficient and stable methods like Cholesky decomposition [96]. W^0 controls the location of the points around the mean of $\boldsymbol{P}$.

3) Obtain the line flow sample points by feeding the sigma points to the nonlinear ACPF function:

$$F_t^k=\psi(\boldsymbol{P}_t^k) \tag{32}$$

Note that, $\boldsymbol{P}_t^k$ is the *k*th sample of the injected power vector, $\boldsymbol{P}$, with $k=0,1,\dots,2n+1$ at time *t* calculated by eqs. (6), (7) and (8). It must be emphasized that in the UT method, the nonlinear function is considered as a black box; hence, no simplification or linearization is necessary.

4) Ultimately, the mean and covariance of flow variable $\boldsymbol{F}$ is calculated using:

$$\overline{\boldsymbol{F}}(t)=\sum_{k=0}^{2n}W^kF_t^k \tag{33}$$

$$\boldsymbol{\Sigma_{FF}}(t)=\sum_{k=0}^{2n}W^k\left(F_t^k-\overline{\boldsymbol{F}}(t)\right)\left(F_t^k-\overline{\boldsymbol{F}}(t)\right)^T \tag{34}$$

In the UT method, the sample points are not selected randomly, rather they are chosen so that they have a predefined mean and covariance. This leads to an accurate estimation of statistics of the output variable.

5.2.3 Flow uncertainty modeling based on ARMA technique

The stochastic overloading line tripping mechanism introduced in the previous chapter is based on modeling the uncertainty of line flow process. In order to accurately model the line flow uncertainty, it is necessary to model the uncertainty in generation and loads. The injected power to the grid is modeled as $P(t) = \mu_P(t) + \epsilon_P(t)$ where $\mu_P(t)$ is the mean of the injected power and can be modeled as the forecasted wind generation or load. The Autoregressive Moving Average (ARMA), a widely used technique in time-series analysis, is used to model the $\mu_P(t)$ and $\epsilon_P(t)$ as discussed earlier in chapter 2.

To capture the fast dynamics of the flow process and analyze the impacts of highly variable wind generation, actual load and wind generation data with a 4-second sampling rate is used in the simulation. Such high-resolution data along with the highly variable nature of wind generation cause the ARMA model to be unstable if it is constructed based on the original time-series. Therefore, it is necessary to smooth out the time-series before estimating ARMA model parameters for each forecasting step. The smoothing is performed by sliding window averaging with width T (e.g. T=15). The ARMA(p,d,q) with d being the differencing order is used to forecast and model time-varying mean and uncertainty of the injected power. Figure 26 shows the time-series of a select wind farm output power and ARMA forecasting with (p,d,q) = (3,1,0). At each step, using M=30 previous observations, the ARMA parameters are estimated and the next sample is forecasted. Performing estimation and forecasting recursively, the time-varying mean of the injected power is modeled as shown in Figure 26. The forecasting error and its associated bandwidth are also shown in Figure 26. The ARMA forecast results into zero mean error with a variance of 3.37 MW.

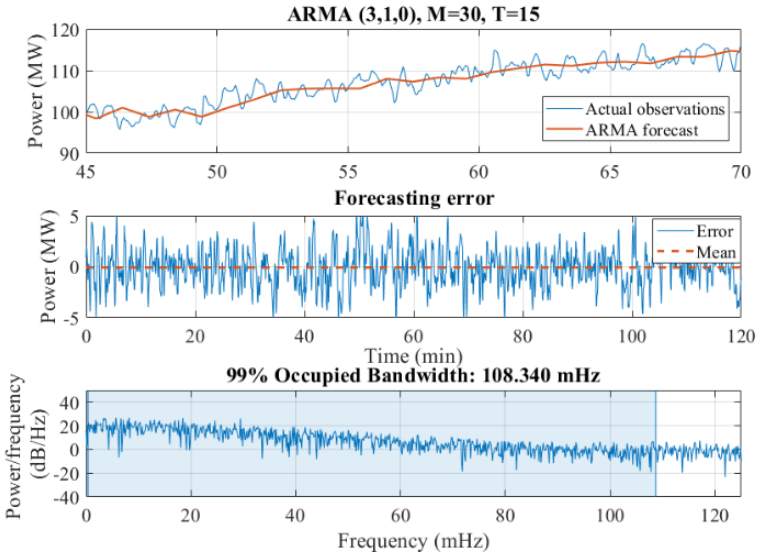


Figure 26. Time-varying mean and uncertainty modeling of a select wind farm using ARMA(3,1,0) with M=30 and T=15

5.2.4 Non-convergent AC power flow

Even after enforcing power balance for the islands formed in the network, the AC power flow may not converge. This is one of the challenges of incorporating AC power flow calculation in CF simulation and could be due to various reasons. Following discusses two of the most probable reasons for non-convergent AC power flow and the actions taken to address the situation.

- *AC power flow does not converge if the system load exceeds the steady-state loading limit:*

The steady-state loading limit is determined from a nose curve where the nose represents the maximum power transfer that the system can handle given a power transfer schedule. To determine the steady-state loading limit, the basic power flow equations

$$g(x) = \begin{bmatrix} P(x) - P^{inj} \\ Q(x) - Q^{inj} \end{bmatrix} = 0 \tag{35}$$

are restructured with a scaling factor λ as:

$$f(x,\lambda) = g(x) - \lambda b = 0 \tag{36}$$

where $x \equiv (\Theta, V_m)$, the vector of system state variables (i.e. voltage phase angles and magnitude), and b is a vector of power transfer given by

$$b = \begin{bmatrix} P_{target}^{inj} - P_{base}^{inj} \\ Q_{target}^{inj} - Q_{base}^{inj} \end{bmatrix} \tag{37}$$

where P_{base}^{inj} and Q_{base}^{inj} are injected real and reactive power for the base case, respectively (usually set to zero), and P_{target}^{inj} and Q_{target}^{inj} are target injected real and reactive power, respectively that for our case is the current dispatched power for the non-converged PF. The effects of the variation of loading or generation can be investigated using the continuation power flow (CPF) by composing the b vector appropriately [21]. To check if this is the case, we run a CPF that gradually increase the loading/generation. If the resulting scaling factor (λ_{max}^{cpf}) associated with the maximum loading the system can handle is less than 1, it indicates that the load for the case exceeds the steady-state loading limit, and loads must be scaled down at least by a factor of λ_{max}^{cpf} to get a convergent power flow solution.

- *The Newton-Raphson algorithm is sensitive to the initial guess:*

In the proposed model, the power flow is solved for each island using the current system state (i.e. voltage magnitudes and angles) as the initial guesses for the Newton-Raphson (NR) algorithm. If the PF does not converge, then an OPF is run for the island with voltage constraints relaxed. This can sometimes help find a new equilibrium for the system and calculate the new voltage profiles. Note that here the reactive power limits are imposed and if no solution within the generator limits can be found then the island will be flagged as total shut down. Also note that in the current model,

the switched shunts are not modeled and the simulations are for the network with all fixed shunts in service.

Figure 27 shows the flowchart of non-convergent power flow handling in the simulation of ACUT cascading failure model. As explained above, the first step when the power flow does not converge is to run a CPF to check if the load has exceeded the steady-state loading limit. If after this step, we still have a non-converged PF, we try to generate a new initial guess for the NR algorithm. At this stage, we create a loop that incrementally drops the load and runs an OPF to finally achieve a converged solution. The load shedding resulting from this step will be attributed to voltage-related failures since it is the voltage that constraints the problem.

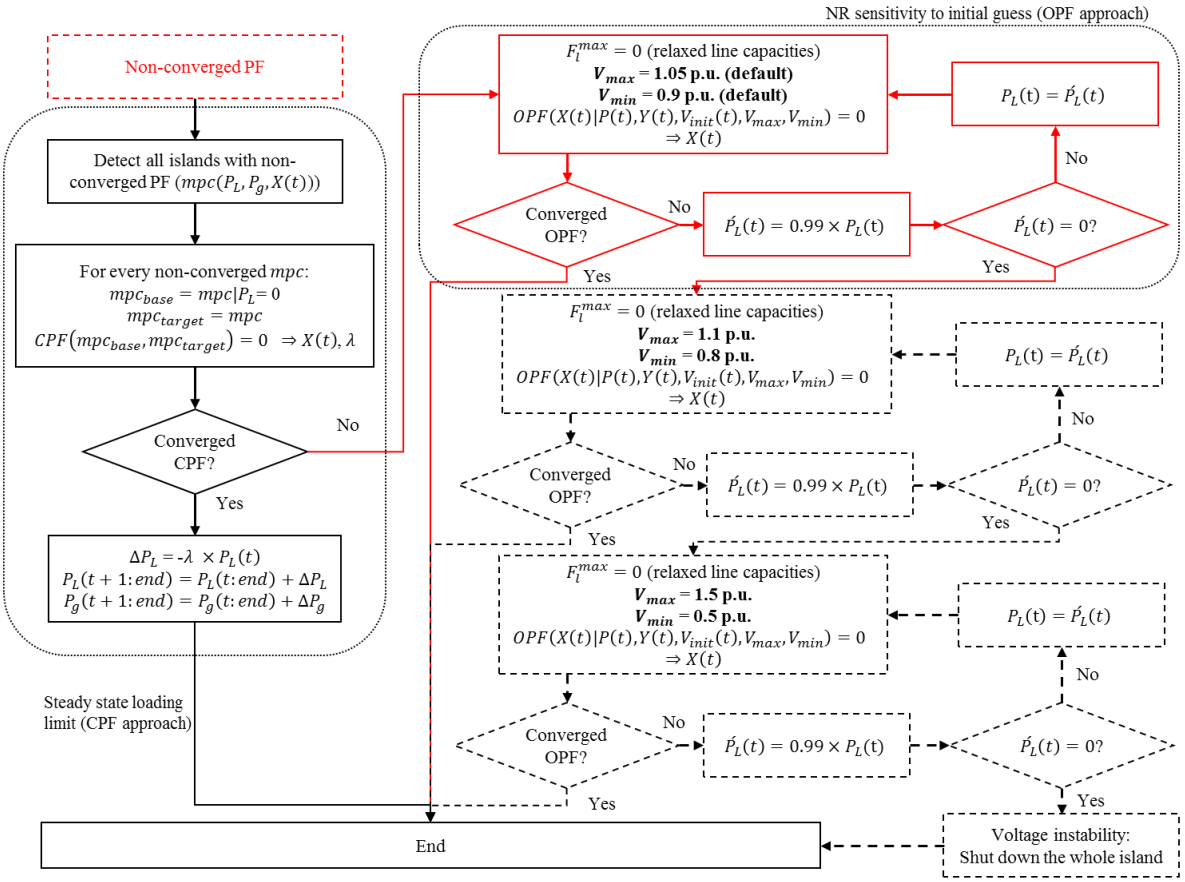


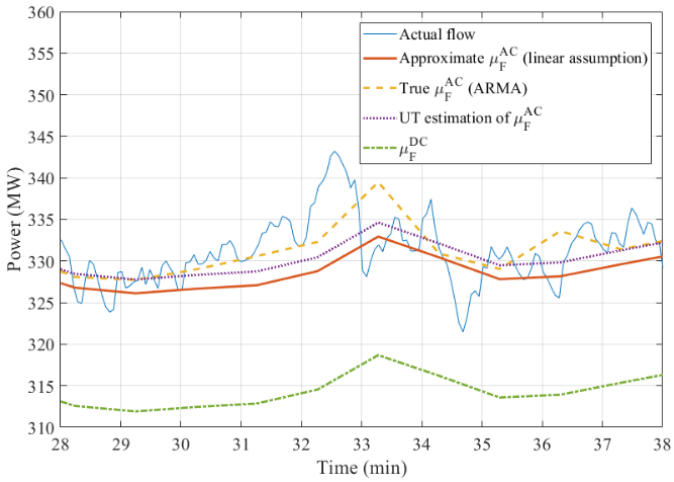
Figure 27. Flowchart of non-convergent power flow handling in CF simulation

5.3 Comparison of flow process with different models

Here we investigate how different assumptions and power flow models affect the estimation of line flow process in a power grid network. The ACTIVSg500 synthetic network is used for all power flow and CF simulation in this chapter.

The time-varying mean of the flow process, $\mu_F(t)$, is estimated for AC and DC power flow models. For the DC model, the linear property of the power flow equations results into the direct estimation of $\mu_F(t)$ and $\Sigma_F(t,\tau)$. While for the AC model the linear assumption is not valid anymore and the UT method is utilized to estimate the $\mu_F(t)$ and covariance of the flow process. For the sake of comparison, $\mu_F(t)$ is also calculated assuming linear relationship for ACPF. As a ground truth, $\mu_F(t)$ is also estimated using ARMA technique introduced earlier. The analysis is performed for two different wind penetration levels to investigate the impact of higher generation variability on the estimation of flow process. Figure 28 shows the flow process of a select line in ACTIVSg500 case.

(a)



(b)

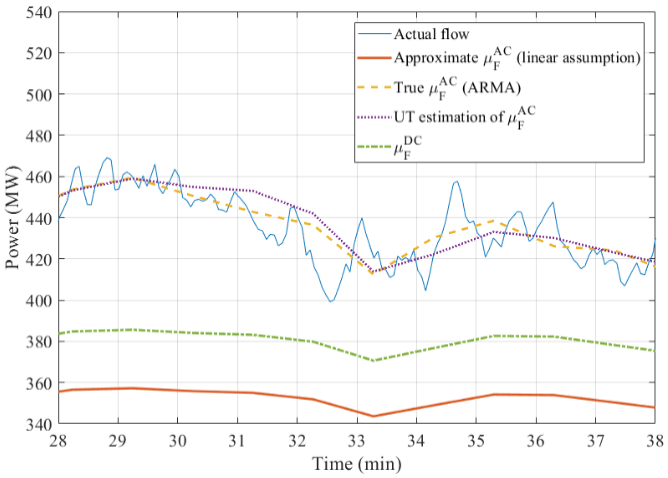


Figure 28. Flow process of line 426 for different models with a) 5% and b) 35% wind penetration level

Figure 28 (a) shows the flow process with 5% wind penetration level. The wind penetration level α is defined as $\alpha = \sum P_w / \sum P_{tot}$ where P_w is the installed wind power capacity and P_{tot} is the total generation capacity of the grid. In Figure 28 (a) the $\mu_F(t)$ estimated with both UT method (purple dotted line) and linear assumption for ACPF (solid red line) is close to the reference $\mu_F(t)$ obtained with ARMA modeling (yellow dashed line). However, as the wind penetration increases to 35% the $\mu_F(t)$ estimated with the linear assumption for ACPF shows a large difference from the true mean with an average of 100 MW difference while the $\mu_F(t)$ estimated with UT method still gives a close value to the true mean of the flow process (Figure 28 (b)). This confirms that as the penetration of wind generation to the grid increases the nonlinearity of ACPF emerges more evidently in the estimation of the flow process and the need for a statistical tool like UT can be justified. Note that the selected line is one of the inner branches of the grid carrying injected power from multiple wind farms. For both penetration levels, the $\mu_F(t)$ calculated with DCPF shows a big difference from the true mean which is due to neglecting line loss and other simplifications made in the model. This demonstrates the necessity of full ACPF model incorporation for accurate estimation of the flow process in the simulation of CF in power grids. Table IX shows the

realization of flow uncertainty in the same selected line for different models. Similar to $\mu_F(t)$, the estimation of flow uncertainty, $\epsilon_F(t)$, with the UT method generates accurate results compared to the ground truth. While, the DC approximation clearly results into a non-zero mean signal with erroneous bandwidth estimation. For the higher wind penetration levels, both DC and AC with linear assumption provide an erroneous estimation of the $\epsilon_F(t)$ both in variance and bandwidth. Figure 29 shows the uncertainty of the selected line and its associated bandwidth calculated using the UT method. As assumed in the proposed stochastic overloading model, the flow uncertainty is a zero mean signal.

Table IX. Estimated flow uncertainty based on different models

Model	5% wind penetration			35% wind penetration		
	Variance	**BW**	**Mean**	**Variance**	**BW**	**Mean**
ARMA	12.8	108	-0.08	131	109	-0.09
UT	11.4	107	0.24	127	108	-0.69
DC	11.6	25.9	16.26	916	32.3	39.23
AC (linear)	11.5	107	0.24	924	17.3	67.31

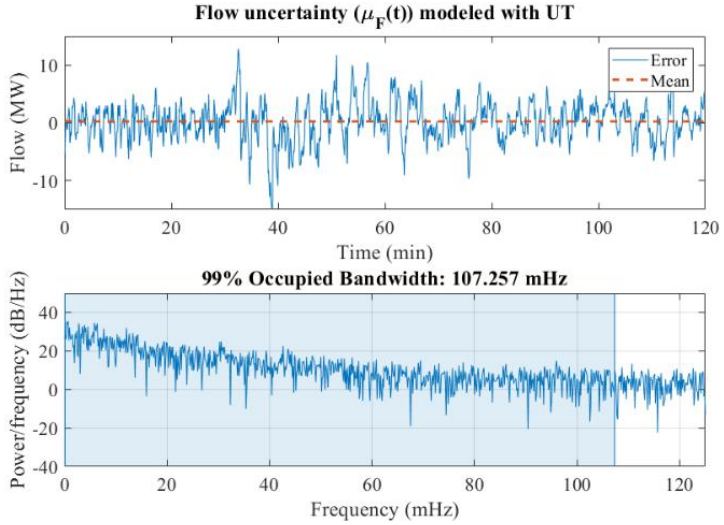


Figure 29. Line flow uncertainty and bandwidth for the selected line modeled with the UT method

5.4 Simulation of CF based on the ACUT model

In this section, cascading failure simulations under three N-2 contingency scenarios to evaluate the performance of the model and assess the blackout size are presented. We will show how the incorporation of AC power flow model can result in different vulnerability evaluation by enabling us to simulate and perform under/over voltage load shedding during the escalation phase of cascading failures. Figure 30 shows the flowchart of the simulation procedure based on the ACUT model. In this figure, the colored blocks represent developed algorithms to successfully implement AC power flow in cascading failure simulation.

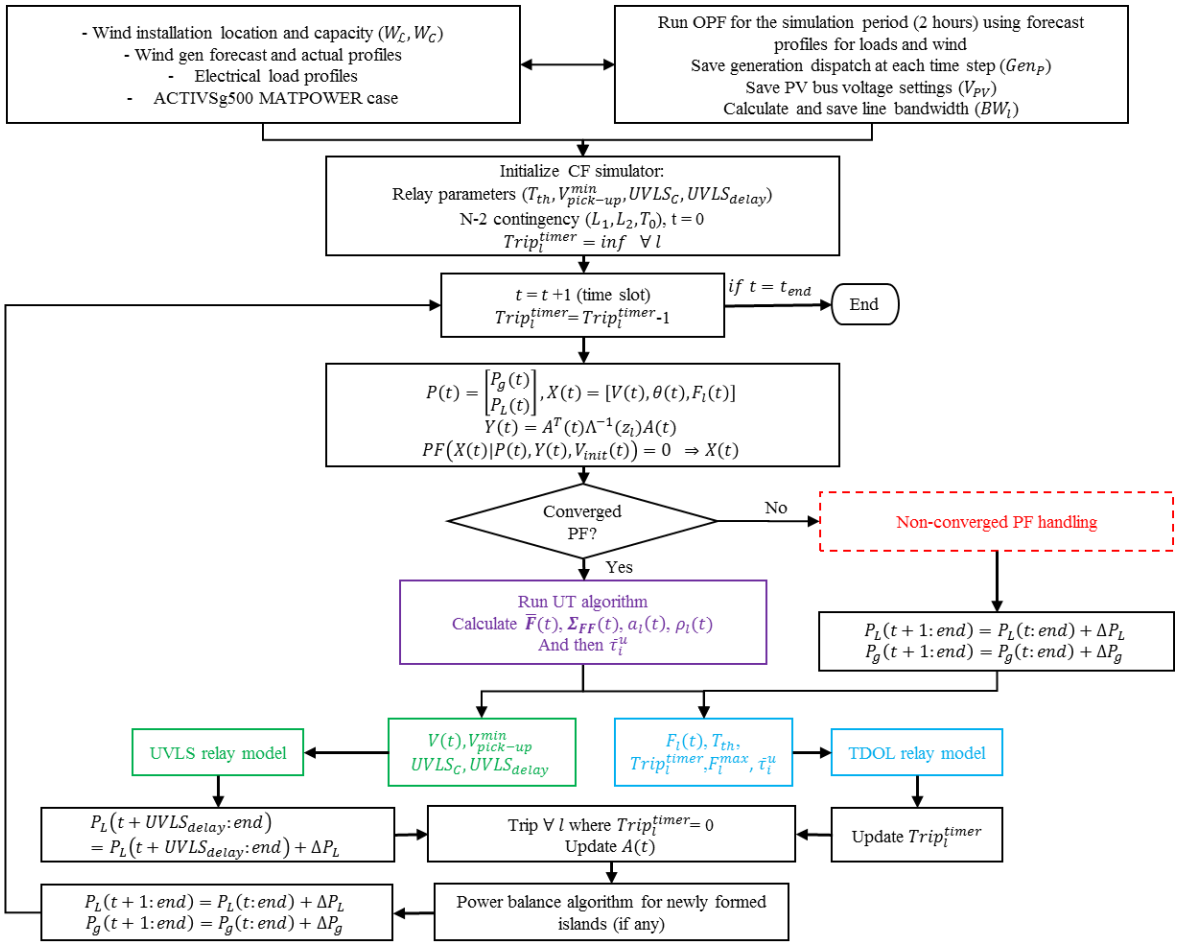


Figure 30. Flowchart of simulation procedure with ACUT model

5.4.1 N-2 contingency simulation results

In this section, the new proposed framework for grid vulnerability studies on a 500-bus synthetic power grid with 597 branches called ACTIVSg500 [93] is evaluated. The ACTIVSg500 is chosen because it offers detailed information on grid data including transmission line rate A capacities and geographical information of network substations which is based on the footprints of South Carolina. To illustrate the effectiveness of the proposed CF modeling with all discussed enhancement, three N-2 contingency scenarios are considered to study the overall grid vulnerability and performance of the relays. For all scenarios, four wind farms are installed in the buses identified in section 5.1, which corresponds to about 5% wind penetration to the grid. Table X shows the maximum wind capacity (MW), installation factor, and rated power of installed wind farms. Installation factor is defined as the fraction of potential wind capacity installed at each location.

Table X. Characteristics of installed wind farms at ACTIVSg500 synthetic network

Bus number	Max wind potential (MW)	Installation factor	Rated power of wind farm (MW)
8	695	0.15	104.25
19	629	0.15	94.35
372	596	0.18	107.28
189	552	0.20	110.40

Note that in the new time-delayed overload real implementation, tripping multiple lines at one instance is possible because of the memory operation of the relays. Therefore, the state of the system and probability of a line to get tripped depends both on the current state of the system as well as its past states. In other words, the proposed CF model is not a Markovian process anymore. Table XI shows the size and statistics of three N-2 contingency scenarios. As expected, the more line outage happens in the network the more islands are formed and consequently the more load shedding becomes necessary to maintain the power balance of each island. This is not necessarily

true for UVLS though. For the second scenario, we have a total of 348 MW UVLS which is larger than both other scenarios. This is due to the independent operation of under-voltage relays which are triggered by a certain threshold for every voltage profile.

Figure 31 shows the evolution process of the three scenarios and the total load shedding for them. All three curves are comparable with typical cascade evolution curves recorded in history in terms of rate of the outage. They usually consist of a slow start, the escalation phase, and settlement. However, sometimes after one or two line outage, the escalation phase starts (scenario 1). The impact of time-delay model for overload relay is visible in the evolution curve of the second scenario where at t = 20 we see a pause in trips but when the timer for multiple overloaded lines reaches zero, the second escalation phase starts (t = 36).

Table XI. Results of three N-2 contingency scenarios for different models

N-2 Contingency	Model	Total trip count	# of formed islands	Total LS (%)	Total UVLS (MW)
Scr1:{95,231}	ACUT	62	16	48.7	326
	Linear AC	28	6	21.9	74.6
	DC	10	3	2.03	0
Scr2:{63,231}	ACUT	51	12	44.6	348
	Linear AC	26	6	21.9	74.6
	DC	8	3	2.03	0
Scr3:{193,234}	ACUT	41	7	38.2	38
	Linear AC	53	13	62.9	49
	DC	11	4	8.20	0

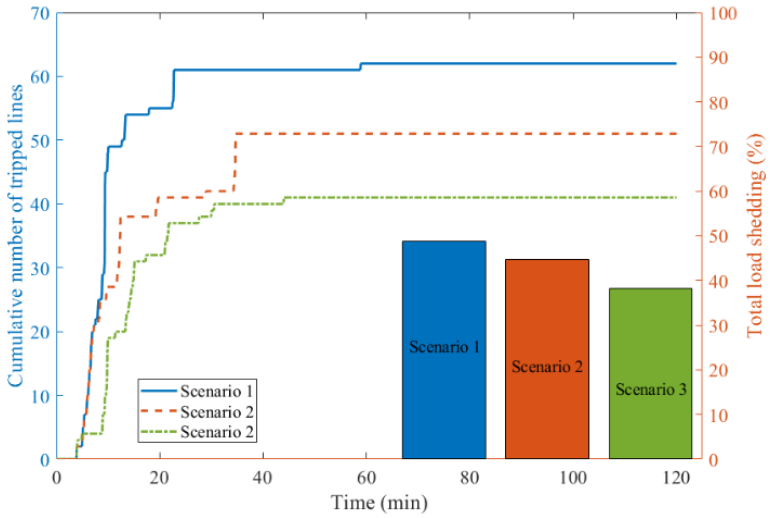


Figure 31. The evolution process and total load shedding percent of CFs for different scenarios based on the proposed model

Figure 32 demonstrates the performance of the UVLS relay during CF. At around minute 14, due to a line trip in the network, the voltages on bus 418 and bus 341 start to drop. When the voltage drops below the UVLS relay activation threshold (0.87 p.u.) the relay starts to shed the load on the two buses by 25% for each time step to recover the voltage. After about 30 seconds, the load on bus 418 drops to 4.8 MW which helps boost the voltage to 0.88 p.u. Then, the UVLS stops load shedding. However, for bus 341, the load shedding continues until all the load on the bus is shut down before boosting the voltage above the threshold.

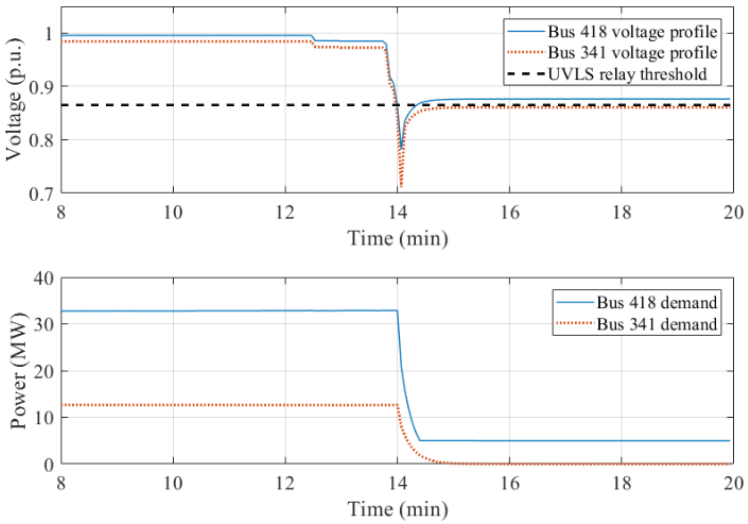


Figure 32. Voltage and load profiles for two select buses during failures

This is an example of a condition during CF where voltage related failures lead to further load shedding in order to maintain voltage stability of the network and prevent voltage collapse. Obviously, this condition could not be simulated based on the DC model and this confirms that the DC CF models may underestimate the severity of blackouts for real scenarios.

6 Validation of the proposed ACUT model

In this chapter, first it is emphasized why it is important to validate and benchmark the proposed cascading failure models and then choosing one of the available options for benchmarking, the performance of ACUT model will be compared with a number of existing and accepted CF models in the literature under fair comparison settings.

6.1 Introduction

Cascading failures are typically triggered by one or more disturbance events, such as a set of the transmission line or generator outages. Triggering events can result from a variety of exogenous threats, such as earthquakes, weather-related disasters, hidden failures, operator errors, and even deliberate acts of sabotage. The dependent outages in a cascade can result from a wide variety of different mechanisms including thermal overloads, voltage instability, and angular instability [32].

Because the resulting blackouts can be large and costly, utilities are increasingly required by reliability regulators to systematically study and manage the cascading outage risk in their system. In response to increasing regulations and several large cascading blackouts [98]–[100], a growing number of tools are being developed in industry and academia to address this analysis need. Given that these tools are increasingly being used to make large investment decisions, and the critical importance of managing the risk of massive cascading blackouts, it is important that cascading failure analysis tools be tested to ensure that they provide accurate and useful information. Doing so requires verification (ensuring that tools perform correctly), validation (checking the accuracy of the results), and benchmarking (a systematic, reproducible validation procedure).

A) Definitions

Benchmarking is a process for measuring the performance of a tool, such as a software program or a business process, using a trusted procedure and/or dataset, in a way that allows one to compare the performance of one tool to another [101].

Cascading failure analysis is a relatively immature power systems application area due to many uncertainties and challenges of CF simulations. This is not the case with well-known analysis line power flow problems. Therefore, there exist a few benchmarks for CF analysis. Benchmarking essentially mean validation and verification of a method [102]. Verification means we check that tool that if it solves the problem that it is intended to solve. Validation, on the other hand, means checking the system and its answers to make sure they are accurate according to some set of criteria [103]. Benchmarking is the combination of these two processes to create reproducible results and comparing different approaches with each other.

The following are a few examples of benchmarking approaches:

1. *Checking for internal validity*. In the internal validation, we check the degree of consistency of the assumptions of a model with reality. Internal validation determines which set of assumptions are in line with reality and which are likely to produce misleading results [101].

2. *Comparing simulation results with real data*. In the case of CF due to various thresholds for different actions like tripping lines, it is very hard to get similar results as in real system or even similar results on an actual system for different conditions. One model may decide to trip a line under certain circumstance while others may decide not to, which will affect the overall grid vulnerability evaluation. Therefore, it is very difficult to obtain the exact same results from models

compared to the actual system however, we can alleviate this by engineering judgment and comparing the statistics of CF instead of their exact sequence of events [101].

3. *Comparing the performance of one tool with another tool (cross-validation).* This is one of the more established benchmarking technique where we compare different models with each other and finding their similarities and differences [101].

4. *Checking for reproducibility*. It is important that a tool produces the same results with the same assumptions and data for multiple runs. Due to the randomness factor incorporated in many CF models, their results will be somewhat different for each run. Therefore, we need to find an optimal number of simulations to make sure that the results are dependable [101].

5. *Sensitivity analysis*. It is important to check how the results change with change in various input parameters. This way, we can identify those parameters that have big impacts on the outcomes.

6.2 Approaches to validating cascading failure simulations

There is a measure of consensus in the power system engineering community to effectively validate well-understood problems, such as power flow and standard contingency analysis. We measure the extent to which models align with actual measurements. However, this type of consensus does not yet exist for cascading failure simulation and analysis [101].

The diversity of mechanisms involved in CF and the difficulty in their accurate modeling is one of the main reasons for this lack of consensus. However, given all the difficulties and challenges with the simulation of CF, following is the list of approaches that can be usefully employed for CF analysis validation.

1. *Validation: Comparing Models to Real Data:* One approach to compare models to real data is to compare simulated event sequences to historical cascade sequences. When we can reproduce a similar sequence of tripping events, it means that the proposed model is a good representative of the actual grid. However, it is not an indicator of the general validity of the model across all operating conditions, rather it shows that there is no big difference in the evolution of the cascade between the actual events and the model.

Moreover, comparison of statistics of simulation which those of historical data is both feasible and easy. Indeed there are distinctive patterns in the observed statistics of historical cascading blackouts, which can be reproduced in simulators [104]–[107]. Therefore, we can run simulations for an appropriate sample size of initiating events and then calculate different statistics and compare them to those of actual events [108]. In this case, if there is a big gap between the simulation and actual results, we can say that the simulation model is not valid. Another useful statistical measure is the observed frequency of cascades of various sizes [101].

2. *Cross-Validation: Comparing Models to Each Other:* Another practical benchmarking method is the comparison of CF results statistics for two or more models. This way, we can see the impact of certain modeling assumptions and check if a parameter variation results in a significant disagreement between the two models. In this case, there is a need for a more detailed analysis of that parameter.

Based on the above discussions, we can list the following highlights that are important to notice when performing validation and benchmarking on the proposed cascading failure model:

- The useful measures to consider when comparing statistics from a simulation with those of actual events and results of other models include line outages, load shed, and energy

unserved. Publishing the statistics of other quantities (for example, propagation or cascade spreading) is helpful too.

- The simulation should clearly specify how the method samples from the potential operating conditions, initial faults, and the progress of the cascades.
- We need to distinguish models that try to reproduce in detail features of certain historical events with those that aim to asses overall risk of CF in long-term planning studies.

6.2.1 Data and test cases for the simulations

In order to have common ground with the existing cascading failure models, it is necessary to employ power grid test cases that first of all are accurately modeled and a true representative of the system behavior and second of all include enough details to accurately model them for dynamic analysis of cascading failure. One of the critical parameters for cascading failure simulation is the transmission line threshold or thermal stability capacity. Another important set of data especially if one aims to study the voltage/var dynamics, is the dynamic model of generators and controllers employed in the system.

Data:

Following is the set of data that can be very useful for the validation and benchmarking process:

1. *Historical Blackout Size Data*: The blackouts in North America since 1984 has been published by the North American Electric Reliability Corporation (NERC). These data indicate that there are approximately 13 very large blackouts (above ~ 300 MW) per year. The affected people and size of blackout in terms of MW load shed are reported in the NERC publications. These data are available on the Internet [109].

2. *Transmission Line Outage Data*: Transmission owners in the USA are required to report higher voltage transmission line and transformer outage data to NERC for the Transmission Availability Data System (TADS) [101]. These data describe the details of component outages (e.g. time and cause) within NERC region. BPA offers more than a decade worth of data for transmission element outage [110]. We can find a way of validating cascading failure simulations by quantifying the line outages propagation in real data.

3. *Reports on Historical Outages*: there are reports on actual outages in [98], [100], [111], [112] with many useful details. It helps with understanding the different and complex nature of mechanisms involved in actual cascading events.

Power system test cases:

As mentioned earlier, the power system test cases play an important role in the simulation of cascading failure. Since blackouts resulting from cascading failure are widespread, the size of the test case used for simulation needs to be large enough to accurately simulate the propagation of the cascades in the network. Another problem for many public test cases is the lack of coordinated line rating limits. There are some available sources for the test cases as listed below:

1. *Small, Publicly Available Test Cases*: there exist a number of public test cases that were mainly developed to serve as the standard test cases for methodology benchmarking and performance evaluation. Some of these test cases are suitable for cascading failure analysis.

2. *Public Test Cases Based on Industry Data*: there exist several test cases that either come with industrial power system analysis software or offered by some regulatory organization that sometimes needs a particular process in order to get access to them. The examples of these cases

are the *New Brunswick (NB) Test System* [113], the *NETS-NYPS 68-Bus Test System* [114], the *MATPOWER Polish Test Cases* [89], and *WECC Reduced 200-Bus System* [115].

3. *Synthetic Power Networks*: these power systems are designed to address the lack of access to accurate and detailed power system data for the research community due to security reasons. Examples of these test cases are the *RT-nestedSmallWorld* and ACTIVSg cases.

6.3 Validation and benchmarking of the ACUT model

In this section, the proposed model is compared with widely used CF models in the literature as well as historical data from real power systems. The benchmarking is based on the statistics of blackout size and total outage numbers. The benchmark includes DC OPA [2], AC OPA [43], Manchester model [44], and historical data from [116]. The AC OPA, DC OPA, and Manchester model data are from [117]. The CF simulations are performed on the standard RTS-96 3-area system model for AC OPA, DC OPA, and Manchester model and ACTIVSg500 synthetic grid is used for CF simulation of ACUT model. Table XII shows the details of the three systems used in validation procedure.

Table XII. Comparison of the three systems used in validation [117]

	ACTIVSg500	RTS-96 3-area	Historical data (WECC)
Number of buses	500	73	20131
Number of branches	597	120	25156

The load shedding is reported as per unit of peak demand. For the proposed method, we simulated 1000 randomly selected N-2 contingency scenarios with the same system state (i.e. same loading level and same line trip time) without any penetration of RE.

Figure 33 shows the distribution of load shed in the form of the survival function of data from the proposed method versus other methods in the literature. The survival function is equivalent to 1-(cumulative distribution function) and shows the probability that the demand loss is larger than a given value, given that the demand loss occurred in the system. The mathematical definition of survival function is given below:

$$S(t) = P(\{T > t\}) = \int_t^{\infty} f(u)du = 1 - F(t) \tag{38}$$

where $f(u)$ is the probability distribution function of the demand loss and $F(t)$ is the cumulative distribution function (CDF) of the load shedding.

Therefore, for all methodologies, the limit of this probability is 1 when demand loss approaches 0. The way in which the survival function decreases as blackout size increases shows the decreasing frequencies of large-scale blackouts. The figure is plotted on a log-log scale so that the smaller probabilities of the larger blackouts can be seen.

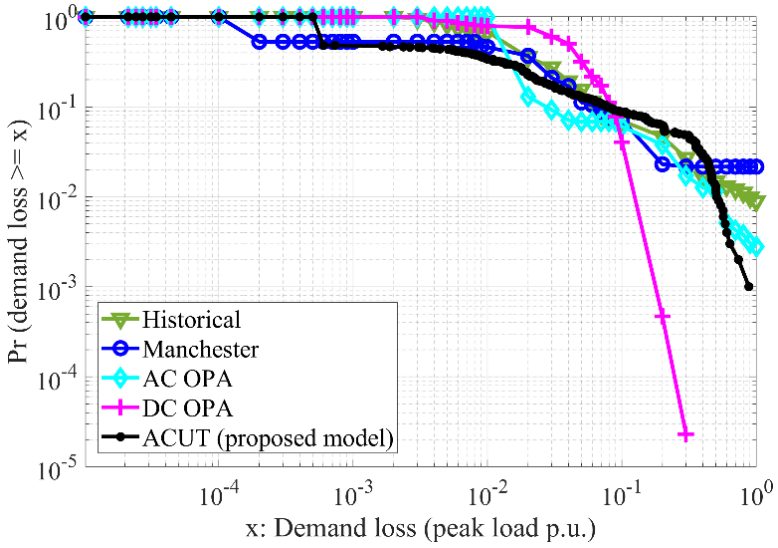


Figure 33. Distribution of load shed for different CF models

As can be seen in Figure 33, the probability of load shed for historical blackout data declines roughly linearly on a log-log plot which means that the corresponding distribution is a “heavy-tail” one. This implies that large blackouts are rarer than smaller blackouts, but not so rare that their risk is smaller. A similar linear pattern can be seen for other methodologies that employ ACPF including the proposed ACUT methodology. On the other hand, the DC OPA model shows a steep decline for the probability of large blackouts thus underestimating their probability of occurrence. This is because in the DC OPA methodology, the system does not have voltage stability issue, and loads can always be supplied by local generators, therefore the probability of a complete blackout is almost zero for this model.

Figure 34 shows the distribution of the number of line outages in the CF data for different methodologies in the form of the survival function. It shows the probability that the total number of line outages is larger than a given value, given that at least one line outage occurred in the system. Therefore, this probability would be 1 when the number of line outages is 1. However, this is not the case for the historical data and the CF simulations for the proposed methodology in this paper, because they are multiplied by the ratio of the number of lines in the RTS by the number of lines in the real system for the historical data and ACTIVSg500 for our simulations. This makes it possible to compare the statistics for systems of different sizes. Again, the log-log plot helps to show the frequency of the larger cascades that are likely to be more consequential.

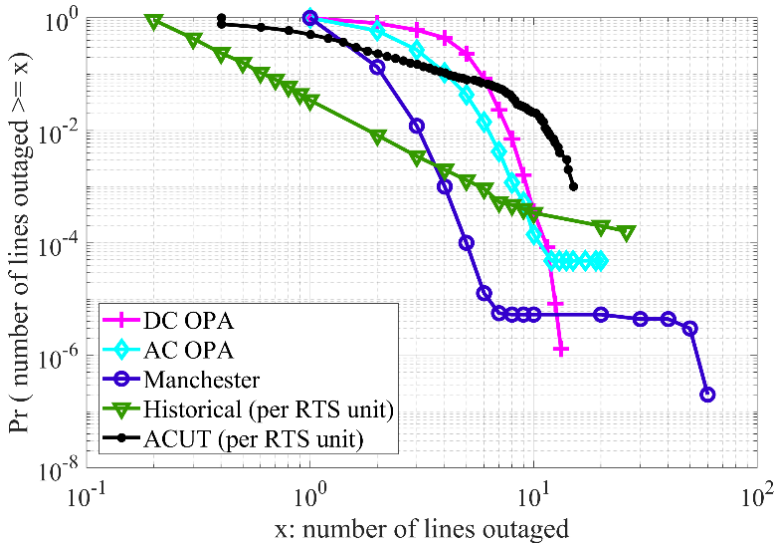


Figure 34. Distribution of the number of line outages for different CF models

In Figure 34 the distribution of the number of line outages for historical data shows a linear decline on the log-log plot similar to the load shed distribution. The distribution of the number of line outages for the proposed model shows a similar linear decline with a knee point for larger blackouts meaning that wide-spread outages are less likely. The distribution for other methodologies is steeper though with lower probabilities for large numbers of outages. The comparison is limited not only by historical data being collected from systems different than the RTS and ACTIVSg500 but also particularly by the larger size of real systems. The small size of the RTS could limit the cascading characteristics observed. This can also be observed by the fact that as the system size grows, the probability curve becomes more linear on the log-log plot. Note that the historical data is for the largest system and ACTIVSg500 with 597 branches is larger than the RTS-96 with 120 branches. Note that the proposed model is simulated on ACTIVSg500 original setting with no renewables integrated into it. This is important in order to have a fair comparison with other models as they are simulated CF for a traditional power system with no RE integration.

As illustrated in Figure 33 and Figure 34 the ACUT model shows a comparable statistical pattern of blackout size to the actual grid and other existing AC models. This means that the proposed model performs acceptably for traditional vulnerability studies of power system with no RE penetration while it also allows for simulating a large set of MC simulations of CF in the presence of intermittent renewables.

7 Overall conclusions and future work

7.1 Conclusions

This dissertation made efforts to develop tools and models for cascading failures (CF) in the power system with the presence of intermittent renewable energy (RE) resources, which reveals the impact of wind energy penetration on the line flow and voltage dynamics. A time-domain statistical model was adopted for uncertainty injected from erratic renewables and electrical loads. Utilizing a set of high-resolution generation and load data from utility companies in North America, the dissertation analyzed and characterized a number of uncertainty sources based on the size and dynamics of its uncertainty.

Next, using the adopted uncertainty model we developed a CF simulator based on DC approximation of power flows that allows preliminary vulnerability analysis considering RE penetration. This model uses a mixed OPF-stochastic method which enables us to see the impact of flow variation due to uncertainty injected from RE. It also incorporates the thermal relay operation mechanism to realistically simulate the cascade of line tripping events.

The limitation of DC power flow in the calculation of voltage profiles prevents us from the simulation of voltage-related failures during CF that could potentially underestimate the severity of blackout resulting from CF in power systems. Therefore, this concern motivated us to make several enhancements to the model to simulate CF under more realistic conditions. One enhancement was on the method with which we increased RE penetration to the grid. In our initial model, we randomly replace the existing conventional generators with RE generators without taking into account the actual potential integration location. We considered more realistic

assumptions in the modeling of wind power penetration using geographical information of grid topology and wind potential map for a given geographical area which indicates the most probable point of interconnection for RE units. The second and most critical enhancement was the incorporation of the full AC power flow (ACPF) model in the proposed CF simulator. This has enabled us to simulate voltage-related failures during the escalation phase of CF by accessing the voltage profiles. Due to the non-linear property of ACPF, we also employed unscented transform (UT) to estimate line flow process statistics from injected power statistics. We call the enhanced model ACUT CF model.

Finally, we have benchmarked ACUT model with historical blackout data as well as a number of existing CF models in the literature. The model generates comparable statistics of blackout size in terms of demand loss to existing methodologies and can successfully simulate under-voltage load shedding during the cascades. But the distribution of the number of line trips contains a big mismatch from historical data, especially in the heavy-tailed distribution. The comparison is limited not only by historical data being collected from systems different than the test cases used for CF simulations but also particularly by the larger size of real systems. The small size of the test system could limit the cascading characteristics observed.

The work on the statistical analysis of actual power system data resulted in the identification of a set of key characteristic metrics of the power system network. Furthermore, a set of statistical analysis tools in MATLAB GUI-based application called *GridStat Analysis Toolkit* is developed based on synthetic grid modeling study. The developed toolkit is capable of performing statistical analysis on realistic and standard power system test cases in four categories of 1) Topological analysis, 2) Grid parameter statistics, 3) Voltage interdependence and 4) Grid scaling properties.

7.2 Future work

This dissertation made preliminary efforts to develop tools and models for studying the impact of wind turbine (WT) penetration on grid vulnerability to CF. The next steps might be a) validation of the model's accuracy, b) MC simulations of CF with various levels of RE penetration to evaluate grid vulnerability, c) enhancing the model by considering generator ramp rates constraint in OPF calculations and d) enhancing ACUT tripping mechanism.

7.2.1 Validation of the model's accuracy

Although we made some efforts to validate the ACUT model by comparing the CF statistics in terms of demand loss and number of trips with other existing CF models in the literature and historical blackout data, the comparisons were limited. In order to fully validate the accuracy of the proposed model, it is necessary to consider other measures such as the line tripping rate, the number of resulted failures after each failure at different stages of CF, and the number of islands formed in the grid. Also, note that in order to have a fair comparison, it is necessary to simulate CF in a relatively large test case to match the size of the actual grid when comparing the statistics with those of historical data.

7.2.2 Monte-Carlo simulations of CF based on the ACUT model

In order to evaluate the impact of high penetration of RE on grid vulnerability, there is a need to run a large number of Monte-Carlo (MC) simulations of different N-1, N-1-1, and N-2 contingency to determine the scale of blackout resulting from CF. For this, it is recommended to replace conventional generators in a test system such as IEEE300, IEEE118, and a synthetic grid such as

ACTIVSg500 with RE generation units and trigger a CF by applying different contingency scenarios.

Critical parameters that need to be taken into consideration are the location of integration of RE units, the number of RE units, and the penetration ratio or the total capacities of RE units. Another important factor that has to be considered is the total generation capacity of the grid before and after RE installation since this will have a significant impact on overall grid vulnerability and resulting load shedding after CF simulation.

It is recommended to run MC simulations on different power system test cases with different characteristics. One particular factor is the loading level of a grid that could potentially have a huge impact when analyzing the grid vulnerability. Also, it is a good practice to simulate a given network under various loading level to take into account the seasonal effect on grid loading and make a comparison on the grid vulnerability under various scenarios.

7.2.3 Generator ramp rate constraints in OPF calculations

One critical assumption in the proposed model is that conventional generators in the grid are capable of absorbing the uncertainty injected by RE to the grid and maintain power balance at all time. In other words, we do not consider ramping constraints for conventional generators and this means that we assume infinite capability for the grid to absorb variation of injected power from RE resources. In the case of wind generators due to the actual inertia of wind turbines, a sudden big change of output power is not common. However, the solar PV farms are integrated to the grid via the use of power electronic devices such as inverters thus eliminating the traditional inertia of the generation. This means that the grid can experience a sudden big change of injected power

from big PV units due to cloud coverage for example. Therefore, it is necessary to consider the ramp rate of existing conventional generators of the system and design a new mechanism that can handle scenarios where the load power cannot be maintained due to high variation from RE. Table XIII shows the ramping capabilities for thermal units in three North American power systems.

Table XIII. Thermal generator ramping capabilities in MW/min for three actual power systems [118]

Measured Thermal Generation (MW/min)	CAISO	PJM	WAPA
Fastest unit MW/min ramp capacity (up/down)	8.6/-7.8	9.1/-8.9	2.4/-2.4
Average unit MW/min ramp capacity (up/down)	1.6/-1.6	0.8/-0.8	0.6/-0.7
Total capacity (up/down)	215/214	291/-306	17/-20
Total simultaneous capacity (up/down)	168/-175	160/-288	9/-19
Maximum used capability (up/down)	42/-66	54/-61	3/-6

Based on data from Table XIII, if the rate of output power from RE is bigger than the total simultaneous ramping capacity of thermal units, it means that the power balance cannot be kept with current dispatch and there is a need for load shedding (or generation curtailment). It also can cause frequency deviation due to the power imbalance that can cause frequency relays to operate which can potentially result in further tripping actions.

Therefore, in order to simulate a more accurate system behavior, it is recommended to incorporate generator ramp constraints into initial OPF as well as during cascading failure simulations. The challenge would be having a dynamic simulation step size instead of a fixed one accounting for fast dynamics of RE and slow response time of conventional generators.

7.2.4 Enhancing the ACUT tripping mechanism

The tripping mechanism in the proposed CF model is based on statistical estimation of the flow process for AC power flow and thermal relays operational scheme. In our simulations, we do not consider other protective schemes such as frequency relays and distance relays.

It is essential to thoroughly investigate the actual mechanism of power grid protection system to accurately simulate the line trips during CF. Particularly, the impact of flow process bandwidth on relay operation seems very interesting and critical in the evaluation of the true impact of RE uncertainty on grid vulnerability.

References

[1] W. Weisheng *et al.*, "On the Road to Wind Power: China's Experience at Managing Disturbances with High Penetrations of Wind Generation," *IEEE Power Energy Mag.*, vol. 14, no. 6, pp. 24–34, Nov. 2016.

[2] I. Dobson, B. A. Carreras, V. Lynch, and D. Newman, "An initial model for complex dynamics in electric power system blackouts," in *Hawaii International Conference On System Sciences HICSS*, 2001.

[3] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, "Critical points and transitions in an electric power transmission model for cascading failure blackouts," *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 12, no. 4, pp. 985–994, Dec. 2002.

[4] K. Ioannis and O. Konstantinos, "An analysis of blackouts for electric power transmission systems," *Trans. Eng. Comput. Technol*, vol. 12, pp. 289–292, 2006.

[5] "Analysis of cascading failure in electric grid based on power flow entropy," *Phys. Lett. A*, vol. 373, no. 34, pp. 3032–3040, Aug. 2009.

[6] Y. Cai, Y. Cao, Y. Li, T. Huang, and B. Zhou, "Cascading failure analysis considering interaction between power grids and communication networks," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 530–538, 2016.

[7] M. Parandehgheibi, E. Modiano, and D. Hay, "Mitigating cascading failures in interdependent power grids and communication networks," in *2014 IEEE International Conference on Smart Grid Communications, SmartGridComm 2014*, 2015, pp. 242–247.

[8] M. Rahnamay-Naeini and M. M. Hayat, "On the role of power-grid and communication-system interdependencies on cascading failures," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 527–530.

[9] Lv Xinyao, Li Huaqiang, Wang Yimiao, Liu Peiqing, and He Qiang, “Cascading failures forecasting research to power grid based on self-organized criticality,” in *2014 International Conference on Power System Technology*, 2014, pp. 820–825.

[10] V. Rampurkar, P. Pentayya, H. A. Mangalvedekar, and F. Kazi, “Cascading Failure Analysis for Indian Power Grid,” *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1951–1960, 2016.

[11] S. Shukla, A. Fung, and K. Raahemifar, “Transient stability optimization and analysis for transmission systems and generators to prevent cascade failure by coordinating contingency planning and load-shedding in power transmission grid,” in *Canadian Conference on Electrical and Computer Engineering*, 2015, vol. 2015–June, no. June, pp. 291–296.

[12] S. Poudel, Zhen Ni, T. M. Hansen, and R. Tonkoski, “Cascading failures and transient stability experiment analysis in power grid security,” in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2016, pp. 1–5.

[13] M. Rahnamay-Naeini and M. M. Hayat, “Impacts of operating characteristics on sensitivity of power grids to cascading failures,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2016–Novem, pp. 0–4, 2016.

[14] A. E. Motter and Y.-C. Lai, “Cascade-based attacks on complex networks,” *Phys. Rev. E*, vol. 66, no. 6, p. 65102, Dec. 2002.

[15] P. Holme and B. J. Kim, “Vertex overload breakdown in evolving networks,” *Phys. Rev. E*, vol. 65, no. 6, p. 66109, Jun. 2002.

[16] P. Crucitti, V. Latora, and M. Marchiori, “Model for cascading failures in complex networks,” *Phys. Rev. E*, vol. 69, no. 4, p. 45104, Apr. 2004.

[17] “Cascade-based attack vulnerability on the US power grid,” *Saf. Sci.*, vol. 47, no. 10, pp.

1332–1336, Dec. 2009.

[18] F. Wenli, L. Zhigang, H. Ping, and M. Shengwei, “Cascading failure model in power grids using the complex network theory,” *IET Gener. Transm. Distrib.*, vol. 10, no. 15, pp. 3940–3949, 2016.

[19] X. Wang, Y. Koc, R. E. Kooij, and P. Van Mieghem, “A network approach for power grid robustness against cascading failures,” in *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, 2015, pp. 208–214.

[20] Y. Zhu, J. Yan, Y. Sun, and H. He, “Revealing cascading failure vulnerability in power grids using risk-graph,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3274–3284, 2014.

[21] J. Yanbing, L. Ruiqiong, H. X. Shanxi, and W. Peng, “Risk assessment of cascading failures in power grid based on complex network theory,” in *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2016, vol. 2016, no. November, pp. 1–6.

[22] I. Dobson, B. Carreras, and D. Newman, “Branching Process Models for the Exponentially Increasing Portions of Cascading Failure Blackouts.,” in *Hawaii International Conference On System Sciences HICSS*, 2005.

[23] J. D. Glover, M. S. Sarma, and T. Overbye, *Power System Analysis & Design, SI Version*. Cengage Learning, 2012.

[24] Z. Wang, A. Scaglione, and R. J. Thomas, “A Markov-Transition Model for Cascading Failures in Power Grids,” in *2012 45th Hawaii International Conference on System Sciences*, 2012, pp. 2115–2124.

[25] Dong Liu, X. Zhang, and C. K. Tse, “A stochastic model for cascading failures in smart

grid under cyber attack," in *2017 IEEE 3rd International Future Energy Electronics Conference and ECCE Asia (IFEEC 2017 - ECCE Asia)*, 2017, pp. 783–788.

[26] M. Rosas-Casals, S. Valverde, and R. V. Sole, "Topological Vulnerability of the European Power Grid Under Errors and Attacks," *Int. J. Bifurc. Chaos*, vol. 17, no. 7, pp. 2465–2475, Jul. 2007.

[27] R. V. Solé, M. Rosas-Casals, B. Corominas-Murtra, and S. Valverde, "Robustness of the European power grids under intentional attack," *Phys. Rev. E*, vol. 77, no. 2, p. 26102, Feb. 2008.

[28] Z. Wang, A. Scaglione, and R. J. Thomas, "The Node Degree Distribution in Power Grid and Its Topology Robustness under Random and Selective Node Removals," in *2010 IEEE International Conference on Communications Workshops*, 2010, pp. 1–5.

[29] H. Xiao and E. M. Yeh, "Cascading Link Failure in the Power Grid: A Percolation-Based Analysis," in *2011 IEEE International Conference on Communications Workshops (ICC)*, 2011, pp. 1–6.

[30] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Topological Models and Critical Slowing down: Two Approaches to Power System Blackout Risk Analysis," in *2011 44th Hawaii International Conference on System Sciences*, 2011, pp. 1–10.

[31] P. Dey, R. Mehra, F. Kazi, S. Wagh, and N. M. Singh, "Impact of Topology on the Propagation of Cascading Failure in Power Grid," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1970–1978, 2016.

[32] Vaiman *et al.*, "Risk Assessment of Cascading Outages: Methodologies and Challenges," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 631–641, May 2012.

[33] H. T. Ma, M. L. Crow, B. H. Chowdhury, and A. Lininger, "Cascading Line Outage

Prevention with Multiple UPFCs," in *2007 39th North American Power Symposium*, 2007, pp. 273–278.

[34] D. Fabozzi and T. Van Cutsem, "Simplified time-domain simulation of detailed long-term dynamic models," in *2009 IEEE Power & Energy Society General Meeting*, 2009, pp. 1–8.

[35] C. Parmer, E. Cotilla-Sanchez, H. K. Thornquist, and P. D. H. Hines, "Developing a dynamic model of cascading failure for high performance computing using trilinos," in *Proceedings of the first international workshop on High performance computing, networking and analytics for the power grid - HiPCNA-PG '11*, 2011, p. 25.

[36] J. Song, E. Cotilla-Sanchez, G. Ghanavati, and P. D. H. Hines, "Dynamic Modeling of Cascading Failure in Power Systems," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2085–2095, May 2016.

[37] J. P. Antoine and M. Stubbe, "EUROSTAG, software for the simulation of power system dynamics. Its application to the study of a voltage collapse scenario," in *Interactive Graphic Power System Analysis Programs, IEE Colloquium on*, 1992, pp. 1–5.

[38] N. Bhatt *et al.*, "Assessing vulnerability to cascading outages," in *2009 IEEE/PES Power Systems Conference and Exposition*, 2009, pp. 1–9.

[39] Shengwei Mei, Fei He, Xuemin Zhang, Shengyu Wu, and Gang Wang, "An Improved OPA Model and Blackout Risk Assessment," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 814–823, May 2009.

[40] M. J. Eppstein and P. D. H. Hines, "A 'Random Chemistry' Algorithm for Identifying Collections of Multiple Contingencies That Initiate Cascading Failure," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1698–1705, Aug. 2012.

[41] M. H. Athari and Z. Wang, "Studying Cascading Overload Failures under High Penetration

of Wind Generation," *2017 IEEE Power Energy Soc. Gen. Meet. (PES GM)*.

[42] M. H. Athari and Z. Wang, "Impacts of Wind Power Uncertainty on Grid Vulnerability to Cascading Overload Failures," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 128–137, Jan. 2018.

[43] Shengwei Mei, Yixin Ni, Gang Wang, and Shengyu Wu, "A Study of Self-Organized Criticality of Power System Under Cascading Failures Based on AC-OPF With Voltage Stability Margin," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1719–1726, Nov. 2008.

[44] D. P. Nedic, I. Dobson, D. S. Kirschen, B. A. Carreras, and V. E. Lynch, "Criticality in a cascading failure blackout model," *Int. J. Electr. Power Energy Syst.*, vol. 28, no. 9, pp. 627–633, Nov. 2006.

[45] M. P. Bhavaraju and N. E. Nour, "TRELSS: A computer program for transmission reliability evaluation of large-scale systems," Electric Power Research Inst., Palo Alto, CA (United States); Public Service Electric and Gas Co., Newark, NJ (United States), 1992.

[46] Q. Chen and L. Mili, "Composite Power System Vulnerability Evaluation to Cascading Failures Using Importance Sampling and Antithetic Variates," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2321–2330, Aug. 2013.

[47] W. Ju, K. Sun, and R. Yao, "Simulation of Cascading Outages Using a Power-Flow Model Considering Frequency," *IEEE Access*, vol. 6, pp. 37784–37795, 2018.

[48] B. Stott, J. Jardim, and O. Alsac, "DC Power Flow Revisited," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1290–1300, Aug. 2009.

[49] H. Cetinay, S. Soltan, F. A. Kuipers, G. Zussman, and P. Van Mieghem, "Comparing the Effects of Failures in Power Grids under the AC and DC Power Flow Models," *IEEE Trans. Netw. Sci. Eng.*, pp. 1–1, 2017.

[50] H. Cetinay, S. Soltan, F. A. Kuipers, G. Zussman, and P. Van Mieghem, “Analyzing Cascading Failures in Power Grids under the AC and DC Power Flow Models.,” *SIGMETRICS Perform. Eval. Rev.*, vol. 45, no. 3, pp. 198–203, 2017.

[51] Q. Chen and L. Mili, “Composite Power System Vulnerability Evaluation to Cascading Failures Using Importance Sampling and Antithetic Variates,” *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2321–2330, Aug. 2013.

[52] K. K. Yagnik and V. Ajjarapu, “Consideration of the wind and solar generation reactive power capability on grid voltage performance,” in *2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–7.

[53] P. Henneaux, P.-E. Labeau, and J.-C. Maun, “Blackout Probabilistic Risk Assessment and Thermal Effects: Impacts of Changes in Generation,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4722–4731, Nov. 2013.

[54] Zhai Xue, Xu Xialing, Lin Tao, Chen Rusi, and Ding Shaoqian, “Research on state vulnerability of power grid with large scale new energy sources based on short circuit capacity margin,” in *International Conference on Renewable Power Generation (RPG 2015)*, 2015, p. 5 .-5 .

[55] K. Jennett, C. Booth, and M. Lee, “Analysis of the sympathetic tripping problem for networks with high penetrations of Distributed Generation,” in *2011 International Conference on Advanced Power System Automation and Protection*, 2011, pp. 384–389.

[56] Jianan Mu, Hongbin Sun, Qinglai Guo, Wenchuan Wu, Fengda Xu, and Boming Zhang, “Design of an online intelligent alarming system for cascading failures of group of wind farms,” in *2013 IEEE Power & Energy Society General Meeting*, 2013, pp. 1–5.

[57] Xi Ye, Ying Qiao, and Zongxiang Lu, “Cascading tripping out of numerous wind turbines

in China: Fault evolution analysis and simulation study," in *2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–11.

[58] H. Khazaei and Y. Zhao, "Ex-post Stable and Fair Payoff Allocation for Renewable Energy Aggregation," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2017.

[59] Y. Zhao and H. Khazaei, "An incentive compatible profit allocation mechanism for renewable energy aggregation," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.

[60] A. Scala, S. Pahwa, and C. Scoglio, "Cascade failures and distributed generation in power grids," *Int. J. Crit. Infrastructures*, vol. 11, no. 1, pp. 27–35, 2015.

[61] L. Zhao, K. Park, and Y.-C. Lai, "Attack vulnerability of scale-free networks due to cascading breakdown," *Phys. Rev. E*, vol. 70, no. 3, p. 35101, Sep. 2004.

[62] I. Dobson, B. A. Carreras, and D. E. Newman, "A loading dependent model of probabilistic cascading failure," *Probab. Eng. Informational Sci.*, vol. 19, no. 1, pp. 15–32, Jan. 2005.

[63] I. Dobson, J. Kim, and K. R. Wierzbicki, "Testing Branching Process Estimators of Cascading Failure with Data from a Simulation of Transmission Line Outages," *Risk Anal.*, vol. 30, no. 4, pp. 650–662, Apr. 2010.

[64] Hui Ren, I. Dobson, and B. A. Carreras, "Long-Term Effect of the n-1 Criterion on Cascading Line Outages in an Evolving Power Transmission Grid," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1217–1225, Aug. 2008.

[65] M. A. Rios, D. S. Kirschen, D. Jayaweera, D. P. Nedic, and R. N. Allan, "Value of security: modeling time-dependent phenomena and weather conditions," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 543–548, Aug. 2002.

[66] M. Rahnamay-Naeini, Z. Wang, N. Ghani, A. Mammoli, and M. M. Hayat, "Stochastic Analysis of Cascading-Failure Dynamics in Power Grids," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1767–1779, Jul. 2014.

[67] M. Rahnamay-Naeini, Zhuoyao Wang, A. Mammoli, and M. M. Hayat, "A probabilistic model for the dynamics of cascading failures and blackouts in power grids," in *2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–8.

[68] M. Anghel, K. Werley, and A. Motter, "Stochastic Model for Power Grid Dynamics," in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 2007, pp. 113–113.

[69] B. Ayyub, *Applied research in uncertainty modeling and analysis*, vol. 20. Springer Science & Business Media, 2007.

[70] A. J. Conejo, M. Carrión, and J. M. Morales, *Decision making under uncertainty in electricity markets*, vol. 1. Springer, 2010.

[71] G. B. Dantzig, "Linear Programming Under Uncertainty," Springer, New York, NY, 2010, pp. 1–11.

[72] Y. Ben-Haim, *Information gap decision theory: decisions under severe uncertainty*. Academic Press, 2006.

[73] Y. Cao, Y. Liu, D. Zhang, W. Wang, and Z. Chen, "Wind power ultra-short-term forecasting method combined with pattern-matching and ARMA-model," *2013 IEEE Grenoble Conf. PowerTech, POWERTECH 2013*, 2013.

[74] S. Huang, S. Member, and K. Shih, "Short-Term Load Forecasting Via ARMA Model Identification Including Non-Gaussian," *Ieee Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, 2003.

[75] A. J. Wood and B. F. Wollenberg, “Power system generation, operation and control,” *John Wiley, New York*, 1984.

[76] Z. Wang, A. Scaglione, and R. Thomas, “Power grid vulnerability measures to cascading overload failures,” *Signal Inf. …*, vol. 1, pp. 0–4, 2012.

[77] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[78] S. A. Fulop and K. Fitz, “Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications,” *J. Acoust. Soc. Am.*, vol. 119, no. 1, pp. 360–371, Jan. 2006.

[79] F. Auger and P. Flandrin, “Improving the readability of time-frequency and time-scale representations by the reassignment method,” *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[80] W. Li, *Risk assessment of power systems : models, methods, and applications*. .

[81] M. Čepin, *Assessment of power system reliability: methods and applications*. Springer Science & Business Media, 2011.

[82] M. H. Athari and Z. Wang, “Modeling the uncertainties in renewable generation and smart grid loads for the study of the grid vulnerability,” in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2016, pp. 1–5.

[83] A. Field, *Discovering statistics using SPSS*. Sage publications, 2009.

[84] D. George, *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e*. Pearson Education India, 2003.

[85] J. Carneiro and L. Ferrarini, “Analysis and design of overload protections for HV lines with a probabilistic approach,” in *10th International Conference on Probabilistic Methods*

Applied to Power Systems PMAPS , 2008.

[86] S. E. Zocholl and G. Benmouyal, "On the Protection of Thermal Processes," *IEEE Trans. Power Deliv.*, vol. 20, no. 2, pp. 1240–1246, Apr. 2005.

[87] R. Thrash, *Overhead Conductor Manual 2nd Edition*. Southwire Company, 2007.

[88] P. Pinceti and M. Vanti, "An Algorithm for the Automatic Detection of Islanded Areas Inside an Active Network," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3020–3028, Nov. 2015.

[89] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[90] M. Drouineau, N. Maïzi, and V. Mazauric, "Impacts of intermittent sources on the quality of power supply: The key role of reliability indicators," *Appl. Energy*, vol. 116, pp. 333–343, Mar. 2014.

[91] N. K. Roy, M. J. Hossain, and H. R. Pota, "Voltage profile improvement for distributed wind generation using D-STATCOM," in *2011 IEEE Power and Energy Society General Meeting*, 2011, pp. 1–6.

[92] "WINDExchange: Wind Energy in South Carolina." [Online]. Available: https://windexchange.energy.gov/states/sc. [Accessed: 25-Jan-2018].

[93] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid Structural Characteristics as Validation Criteria for Synthetic Networks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3258–3265, Jul. 2017.

[94] M. Aien, M. Fotuhi-Firuzabad, and F. Aminifar, "Probabilistic load flow in correlated uncertain environment using unscented transformation," *IEEE Trans. Power Syst.*, vol. 27,

no. 4, pp. 2233–2241, 2012.

[95] A. K. Singh and B. C. Pal, "Decentralized Dynamic State Estimation in Power Systems Using Unscented Transformation," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 794–804, Mar. 2014.

[96] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[97] V. Ajjarapu and C. Christy, "The continuation power flow: a tool for steady state voltage stability analysis," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 416–423, 1992.

[98] U.-C. Force, "Final report on the august 14th blackout in the united states and canada," *Dep. Energy Natl. Resour. Canada*, 2004.

[99] G. A. Maas, M. Bial, and J. Fijalkowski, "Final report-system disturbance on 4 november 2006," *Union Coord. Transm. Electr. Eur. Tech. Rep*, 2007.

[100] N. Ferc, "Arizona-southern california outages on 8 September 2011: causes and recommendations," *FERC NERC*, 2012.

[101] J. Bialek *et al.*, "Benchmarking and Validation of Cascading Failure Analysis Tools," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4887–4900, Nov. 2016.

[102] P. Guide, "A guide to the project management body of knowledge," in *Project Management Institute*, 2004, vol. 3.

[103] W. L. Oberkampf and C. J. Roy, *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[104] D. N. Kosterev, C. W. Taylor, and W. A. Mittelstadt, "Model validation for the August 10, 1996 WSCC system outage," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 967–979, 1999.

[105] B. A. Carreras, D. E. Newman, I. Dobson, and A. B. Poole, "Evidence for self-organized

criticality in electric power system blackouts," in *34th Hawaii International Conference on System Sciences, Maui, Hawaii*, 2001.

[106] P. Hines, J. Apt, and S. Talukdar, "Large blackouts in North America: Historical trends and policy implications," *Energy Policy*, vol. 37, no. 12, pp. 5249–5259, 2009.

[107] I. Dobson, "Estimating the Propagation and Extent of Cascading Line Outages From Utility Data With a Branching Process," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2146–2155, Nov. 2012.

[108] B. A. Carreras, D. E. Newman, I. Dobson, and N. S. Degala, "Validating OPA with WECC Data," in *2013 46th Hawaii International Conference on System Sciences*, 2013, pp. 2197–2204.

[109] "Supplementary information for 'Large blackouts in North America: Historical trends and policy implications,'" 2009. .

[110] "Outage & Reliability Reports." [Online]. Available: https://transmission.bpa.gov/Business/Operations/Outages/. [Accessed: 11-Oct-2017].

[111] U. O. Handbook, "Union for the Co-ordination of Transmission of Electricity." June, 2004.

[112] P. Kundur, C. Taylor, and P. Pourbeik, "Blackout experiences and lessons, best practices for system dynamic performance, and the role of new technologies," *IEEE Task Force Rep.*, 2007.

[113] W. CIGRE, "of SC 38," Power System Reliability Analysis," *Appl. Guid. CIGRE, Paris, Fr.*, 1987.

[114] B. Pal and B. Chaudhuri, *Robust control in power systems*. Springer Science & Business Media, 2006.

[115] Y. Hou, C.-C. Liu, K. Sun, P. Zhang, S. Liu, and D. Mizumura, "Computation of milestones

for decision support during system restoration," in *2011 IEEE Power and Energy Society General Meeting*, 2011, pp. 1–10.

[116] B. A. Carreras, D. E. Newman, and I. Dobson, "North American Blackout Time Series Statistics and Implications for Blackout Risk," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4406–4414, Nov. 2016.

[117] P. Henneaux *et al.*, "Benchmarking Quasi-Steady State Cascading Outage Analysis Methodologies," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018, pp. 1–6.

[118] B. Kirby and M. Milligan, "Method and case study for estimating the ramping capability of a control area or balancing authority and implications for moderate or high wind penetration," National Renewable Energy Lab., Golden, CO (US), 2005.

[119] M. H. Athari and M. M. Ardehali, "Operational performance of energy storage as function of electricity prices for on-grid hybrid renewable energy system by optimized fuzzy logic controller," *Renew. Energy*, 2016.

[120] H. Moradi, A. Abtahi, and M. Esfahanian, "Optimal energy management of a smart residential combined heat, cooling and power," *Int. J. Tech. Phys. Probl. Eng*, vol. 8, pp. 9–16, 2016.

[121] H. Moradi, A. Abtahi, and M. Esfahanian, "Optimal operation of a multi-source microgrid to achieve cost and emission targets," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, 2016, pp. 1–6.

[122] A. Dehghan Banadaki and A. Feliachi, "Voltage Control Using Eigen Value Decomposition of Fast Decoupled Load Flow Jacobian," in *49th North American Power Symposium (NAPS)*, 2017, pp. 1–6.

[123] A. Shokrollahi, H. Sangrody, M. Motalleb, M. Rezaeiahari, E. Foruzan, and F. Hassanzadeh, “Reliability Assessment of Distribution System Using Fuzzy Logic for Modelling of Transformer and Line Uncertainties,” in *49th North American Power Symposium (NAPS)*, 2017, pp. 1–6.

[124] M. H. Athari and Z. Wang, “Modeling the Uncertainties in Renewable Generation and Smart Grid Loads for the Study of the Grid Vulnerability,” *2016 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf.*, 2016.

[125] M. Parashar, J. S. Thorp, and C. E. Seyler, “Continuum Modeling of Electromechanical Dynamics in Large-Scale Power Systems,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 51, no. 9, pp. 1848–1858, Sep. 2004.

[126] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, “Critical points and transitions in an electric power transmission model for cascading failure blackouts,” *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 12, no. 4, p. 985, 2002.

[127] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[128] Z. Wang, R. J. Thomas, and A. Scaglione, “Generating Random Topology Power Grids,” in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, 2008, pp. 183–183.

[129] Z. Wang, A. Scaglione, and R. J. Thomas, “Generating Statistically Correct Random Topologies for Testing Smart Grid Communication and Control Networks,” *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 28–39, Jun. 2010.

[130] Z. Wang and R. J. Thomas, “On Bus Type Assignments in Random Topology Power Grid Models,” in *2015 48th Hawaii International Conference on System Sciences*, 2015, pp.

2671–2679.

[131] K. M. Gegner, A. B. Birchfield, Ti Xu, K. S. Shetye, and T. J. Overbye, "A methodology for the creation of geographically realistic synthetic power flow models," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, 2016, pp. 1–6.

[132] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical Considerations in the Creation of Realistic Synthetic Power Grids for Geomagnetic Disturbance Studies," *IEEE Trans. Power Syst.*, pp. 1–1, 2016.

[133] S. H. Elyas and Z. Wang, "Statistical analysis of transmission line capacities in electric power grids," in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2016, pp. 1–5.

[134] M. H. Athari and Z. Wang, "Statistically Characterizing the Electrical Parameters of the Grid Transformers and Transmission Lines," in *The 10th Bulk Power Systems Dynamics and Control Symposium – (IREP 2017)*, 2017, pp. 1–7.

[135] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[136] D. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[137] "Electric Grid Test Case Repository." [Online]. Available: https://electricgrids.engr.tamu.edu/. [Accessed: 24-Jul-2017].

[138] S. H. Elyas and Z. Wang, "Improved Synthetic Power Grid Modeling With Correlated Bus Type Assignments," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3391–3402, Sep. 2017.

[139] Z. Wang and S. H. Elyas, "On the Scaling Property of Power Grids," in *Hawaii International Conference On System Sciences HICSS*, 2017, pp. 1–8.

[140] E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numer. Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.

[141] S. H. Elyas, Z. Wang, and R. J. Thomas, “On the Statistical Settings of Generation and Load in a Synthetic Grid Modeling,” in *The 10th Bulk Power Systems Dynamics and Control Symposium - (IREP 2017)*, 2017, pp. 1–7.

List of publications

Journal publications:

[J4] M.H. Athari and Z. Wang, "Stochastic AC Cascading Failure Model under Uncertain Generation using Unscented Transform," *IEEE Tran. On Sustainable Energy,* Oct 2018 (passed 1st round of review).

[J3] M.H. Athari and Z. Wang, "Introducing Voltage-dependent Parameters to Synthetic Grids Electrical Topology," *IEEE Tran. On Smart Grids*, Jun. 2018, pp. 1–1.

[J2] A.B. Birchfield, E. Schweitzer, **M.H. Athari**, T. Xu, T.J. Overbye, A. Scaglione, and Z. Wang, "A Metric-Based Validation Process to Assess the Realism of Synthetic Power Grids," *Energies*, Aug. 2017, vol: 10(8), pp:1233.

[J1] M.H. Athari and Z. Wang, "Impacts of Wind Power Uncertainty on Grid Vulnerability to Cascading Overload Failures," *IEEE Tran. On Sustainable Energy*, Jan. 2018, vol. 9, no. 1, pp. 128–137.

Peer-reviewed conference papers:

[C10] M.H. Athari, K.D. Hannah, and Z. Wang, "Automated Enhancement of Transmission Planning Bus-Branch Model with EMS Node-Breaker Substation Models," *2018 CIGRE Grid of the Future (GOTF),* Oct. 2018 (accepted, to appear).

[C9] M.H. Athari and Z. Wang, "Grid Vulnerability Analysis based on Probabilistic Cascading Failure Model and Wind Power Installation," *52nd Hawaii International Conference on System Sciences (HICSS),* Jul. 2018 (accepted, to appear).

[C8] Z. Wang and **M.H. Athari**, "Statistically Analyzing Power System Network," *IEEE Power & Energy Society General Meeting, PESGM2018,* Aug. 2018, Portland, OR (accepted, to appear).

[C7] M.H. Athari and Z. Wang, "Interdependence of Transmission Branch Parameters on the Voltage Levels," *51st Hawaii International Conference on System Sciences (HICSS),* Jan. 2018, Waikoloa, HI.

[C6] M.H. Athari, C. Yang, and Z. Wang, "Sequential Optimal Placement of Distributed Photovoltaics using Downstream Power Index," *49th North American Power Symposium (NAPS),* Sep. 2017, Morgantown, WV.

[C5] M.H. Athari and Z. Wang, "Statistically Characterizing the Electrical Parameters of the Grid Transformers and Transmission Lines," *10th Bulk Power Systems Dynamics and Control Symposium, IREP2017,* Sep. 2017, Espinho, Portugal.

[C4] M.H. Athari and Z. Wang, "Studying Cascading Overload Failures under High Penetration of Wind Generation," *IEEE Power & Energy Society General Meeting, PESGM2017,* Jul. 2017, Chicago, IL.

[C3] H. Sadeghian, **M.H. Athari**, and Z. Wang, "Optimized Solar Photovoltaic Generation in a Real Local Distribution Network," *IEEE Innovative Smart Grid Technologies, ISGT2017,* Apr. 2017, Arlington, VA.

[C2] M.H. Athari and Z. Wang, "Time-Series Analysis of Photovoltaic Distributed Generation Impacts on a Local Distributed Network," *IEEE PowerTech2017,* Jun. 2017, Manchester, UK.

[C1] M.H. Athari and Z. Wang, "Modeling the Uncertainties in Renewable Generation and Smart Grid Loads for the Study of the Grid Vulnerability," *IEEE Innovative Smart Grid Technologies (ISGT)*, Sep. 2016, Minneapolis, MN.

Appendix A: Synthetic grid modeling for vulnerability studies

In this chapter, we perform a set of statistical analyses on transmission network parameters for the real-world power system to extract some validation and tuning metrics as well as some statistics to help us model the transmission lines accurately in grid vulnerability studies. This will also contribute to the "Synthetic Grid Modeling" project which is an ongoing effort to address the problem of limited access to actual grids in the research community.

1.1 Introduction

Synthetic power networks are emerging as a potential solution for the lack of test cases for performance evaluation in power system research and development. Generally, access to real data in critical infrastructure like power networks is limited due to confidentiality requirements. Utility companies and regulatory agencies don't share such data and strictly limit access to actual power systems data for the public and researchers due to their sensitivity. On the other hand, it is important that new concepts and algorithms developed by researchers be evaluated in relatively large and complex networks with the same characteristics as actual grids so that they can be reproducible by peers. For example, authors in [119]–[121] have developed a new storage management and energy management algorithms which enable a bidirectional power flow from Microgrids to power networks that need evaluation with realistic grid topology and in [122] the proposed voltage control algorithm needs to be tested and verified on several realistic power system test cases. Since Synthetic power networks are entirely fictitious but with the same characteristics as realistic networks, they can be freely published to the public to facilitate the advancement of new technologies in power systems.

Development of efficient synthetic power system models requires that their size, complexity, and electrical and topological characteristics match with those of real power grids. Power networks are complex infrastructures with various components. In addition to topological characteristics of power networks, they include several components with different electrical characteristics such as different types of transformers, switched shunt reactive power compensation, remote tap changing bus voltage regulation, etc. Development of synthetic power networks with the same complexity that can simulate the exact behavior of actual grids needs a comprehensive study of different components from both electrical and topological perspectives. For example, authors in [123] used historical data and probabilistic methods for reliability assessment of the distribution system. Also, the increasing level of renewable generation in power systems has introduced an unprecedented level of uncertainty into grids [124]. In the literature, many studies are dedicated for characterizing actual power networks mainly from topological perspectives such as ring-structured power grid developed in [125] and tree-structured power grid model to address the power system robustness [26], [126]. Small world approach described in [127] served as a reference for the works of [28], [128], [129] to develop an approach for generating truly synthetic transmission line topologies. A random topology power network model, called *RT-nestedSmallWorld*, is proposed in [129] based on comprehensive studies on the electrical topology of some real-world power grids. The impacts of different bus type assignments in synthetic power networks on grid vulnerability to cascading failures are investigated in [130].

In [93] the authors presented a substation placement method and transmission lines assignment from real energy and population data based on methodology introduced in [131], [132]. The proposed methodology employs a clustering technique to ensure that synthetic substations meet

realistic proportions of load and generation. However, the authors will continue to augment test cases by adding additional complexities such as transmission network electrical parameters assignment. In another study, the authors performed a statistical analysis on transmission line capacity regarding both topology and electrical parameters. However, all these studies focus mainly on topology-related parameters of transmission lines and ignore electrical parameters such as the impedance of transmission lines and transformers. For the validation purposes, [133] reported some initial study results on the statistics of transmission line parameters. Reference [134] studied the statistical properties of the transmission network and extract the empirical probability density functions (PDF) for some of its electrical parameters.

1.2 Characterizing electrical parameters of the transmission network

In this section, we mainly focus on the statistical analysis of transformers and transmission lines electrical parameters such as per unit impedance, nominal capacity and X/R ratio. The goal of this section is to a) provide a well-defined "rules" for transmission network parameters as potential validation metrics for existing synthetic grid models and b) to provide guidelines on how to accurately configure them in synthetic models for grid vulnerability studies applications. A very large sample of actual operating transformers and transmission lines from two real-world power systems is used to extract the statistical characteristics of their parameters.

A) *Grid transformers*

Generally, in power systems branches are referred to transmission lines or transformers between two buses in the network. Also, in some cases, shunts are considered in the branch category. In this paper, we first perform some statistical analysis on transformers electrical parameters

extracted from two real-world power systems. Next, transmission lines from the same networks will be studied to extract some statistics for their critical parameters.

1. *Per unit impedance using the system MVA base or transformer's power rating?*

In power system analysis the use of per unit system to express the system quantities as fractions of a defined base unit quantity is common. This is important especially for transformers as the voltage level is different for their terminals and per unit system simplifies transformer calculations. Another advantage for this expression is that similar types of apparatus like transformers will have the impedances lying within a narrow numerical range when expressed as a per-unit fraction of the equipment rating, even if the unit size varies widely. However, per unit impedances of power grid components are usually converted to new values using a common system-wide base for application in power system analysis like power flow or economic power flow calculations. This conversion depends on the reference voltage base for different zones in the system and a predefined unique power base for the entire system according to the following simple equation:

$$Z_{p.u.}^{rating} = Z_{p.u.}^{system} \times \left(\frac{S_{Base}^{rating}}{S_{Base}^{system}} \right) \tag{39}$$

where $Z_{p.u.}^{system}$ is the per unit impedance calculated using a system-wide common base S_{Base}^{system} and $Z_{p.u.}^{rating}$ is the new per unit impedance calculated using S_{Base}^{rating}. Note that, here the voltage bases are selected the same as the nominal voltage of transformer terminals for each zone to simplify the calculations.

In the power grids, the use of different voltage levels is a common practice to decrease the power loss through transmission lines. Thus there are transformers with different turn ratios to couple the

areas with different voltage levels. In this study, the transformers are grouped into different categories based on their high voltage terminals. This is because as the nominal voltage level increases the transformer size gets larger, so studying them in groups based on voltage level seems reasonable for extracting validation metrics. The purpose of statistical experiments in this study is to identify several validation metrics for transformers parameters including their impedances to help validate synthetic power networks. This would be even more helpful if the range for different parameters can be specified for typical power system components. The first experiment tries to find the relationship between the MVA rating of the transformer and its per unit impedance. These analyses are performed on both per unit values in system base and converted values to transformers own MVA ratings. The original power system data used in this study offer transformer impedance in per unit calculated based on the common base for the system. Figure 35 shows the scatter plot of transformers per unit reactance (X) and MVA rating for the original and converted per unit reactance of transformers. Note that although transformers with high voltage terminal of 115 kV are selected for this comparison, the results are fairly consistent for other voltage levels as shown in Figure 36.

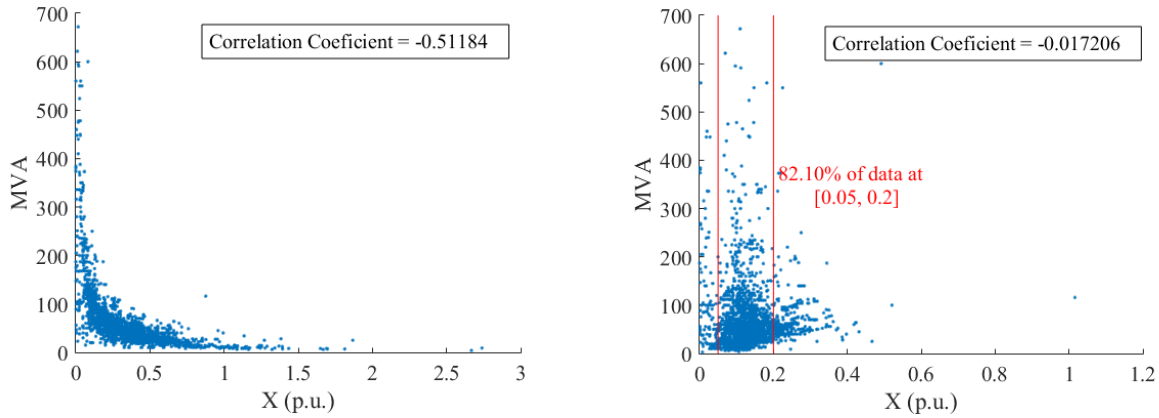


Figure 35. Scatter plot of per unit reactance versus MVA rating of the transformer for a) system common base and b) converted to transformer own MVA rating

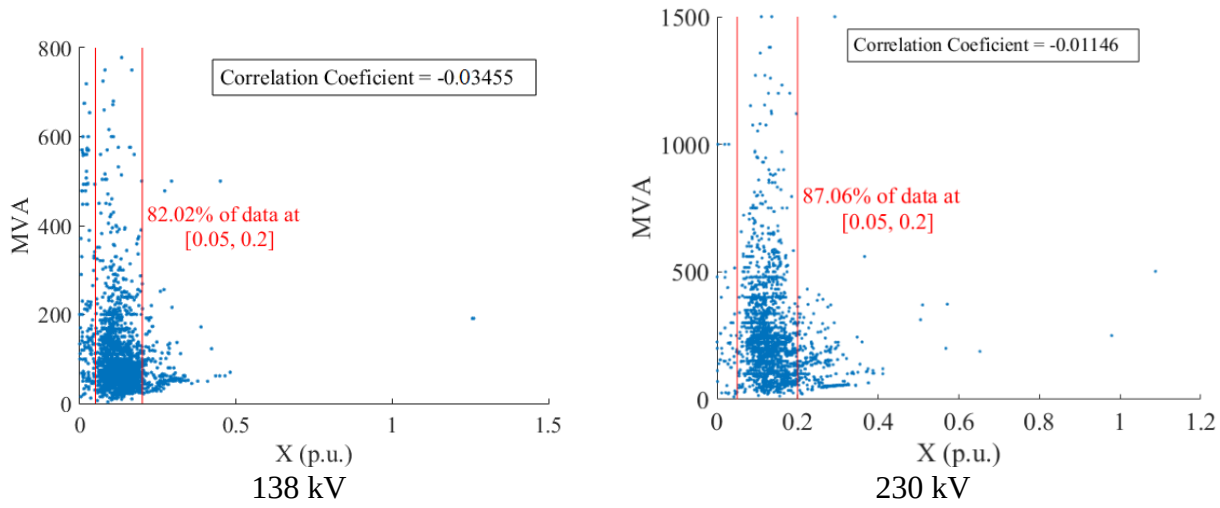


Figure 36. Scatter plot of per unit reactance versus MVA rating of the transformer for 138 and 230 kV transformers

The scatter plot for per unit reactance on system common base shows a descending trend as the size of transformer increases which means there is relatively large correlation coefficient between the two as shown in Figure 35 (a). In this case, the per unit reactance values span from nearly 0 to 2.75 p.u which is relatively a large range for this parameter. However, when we consider the same scatter plot for converted per unit reactance to transformer own MVA rating, this range narrows down to [0, 0.5] p.u putting at least 80% of them within even a narrower range of [0.05, 0.2] p.u. In addition, almost zero correlation coefficient means that this range is independent of transformer size and voltage level.

The same scatter plots for converted values of per unit reactance versus MVA rating of transformers for other voltage levels are depicted in Figure 36. It is found that per unit reactance of transformers in power systems regardless of their size lie within a narrow range when calculated on their own power base and statistics reflect what is known from engineering practice. This can be a potential validation metric for synthetic power networks transformers along with other statistical measures such as their probability distribution.

2. *Transformer parameter distribution*

Transformer parameters statistics are derived using over 30000 actual power transformers. The database includes different types of transformers such as fixed step-down and step-up transformers, three winding transformers, On-load Tap Changer (OLTC) transformers, and autotransformers. A negative impedance often occurs in the star modeling of a three winding transformer due to how the leakage reactance is measured/modeled [23]. Also, Network equivalencing methods can create negative impedances which can affect the statistics of transmission network parameters. To avoid such a scenario, data are filtered by $R > 0, X > 0$ to exclude abnormal transformer parameters from samples. Also, due to lack of detailed information on some transformers, their MVA ratings are reported with either very large or zero values. These transformers too are excluded from samples to have accurate statistics.

The probability distribution of transformer parameters is another measure that can be used along with parameter range as a validation metric in synthetic power networks. The probability distribution of a random variable, say transformer per unit reactance, is a function that describes how likely we can obtain the different possible values of the random variables. Using the database of real transformer data, we can get the empirical cumulative density function (CDF) of each parameter that can give us the empirical probability density function (PDF). Next, to provide a more systematic approach for generating synthetic models, we try to fit approximated distribution functions to empirical PDFs. The goodness of this fit can be measured with Kullback-Leibler divergence.

3. *Kullback-Leibler Divergence*

In probability theory and information theory, the Kullback–Leibler (KL) divergence, also called discrimination information, is a measure of the difference between two probability distributions P and Q. It is not symmetric in P and Q. In applications, P typically represents the "true" distribution of data, observations, or a precisely calculated theoretical distribution, while Q typically accounts for a theory, model, description, or approximation of P [135]. Specifically, the KL divergence from Q to P denoted $D_{KL}(P \parallel Q)$, is the amount of information lost when Q is used to approximate P. For discrete probability distributions P and Q, the KL divergence from Q to P is defined to be [136]

$$D_{KL}(P \parallel Q) = \sum_{i} P(i) log \frac{P(i)}{Q(i)} \tag{40}$$

In words, it is the expectation of the logarithmic difference between the probabilities P and Q, where the expectation is taken using the probabilities P. Therefore, smaller values for the divergence represents a more accurate fit for the empirical PDF of transformer parameters.

4. *Transformer per unit reactance*

Figure 37 shows the empirical PDF and the normal fit distribution of transformers per unit reactance for select voltage levels. The goodness of this fit is measured with Kullback-Leibler divergence.

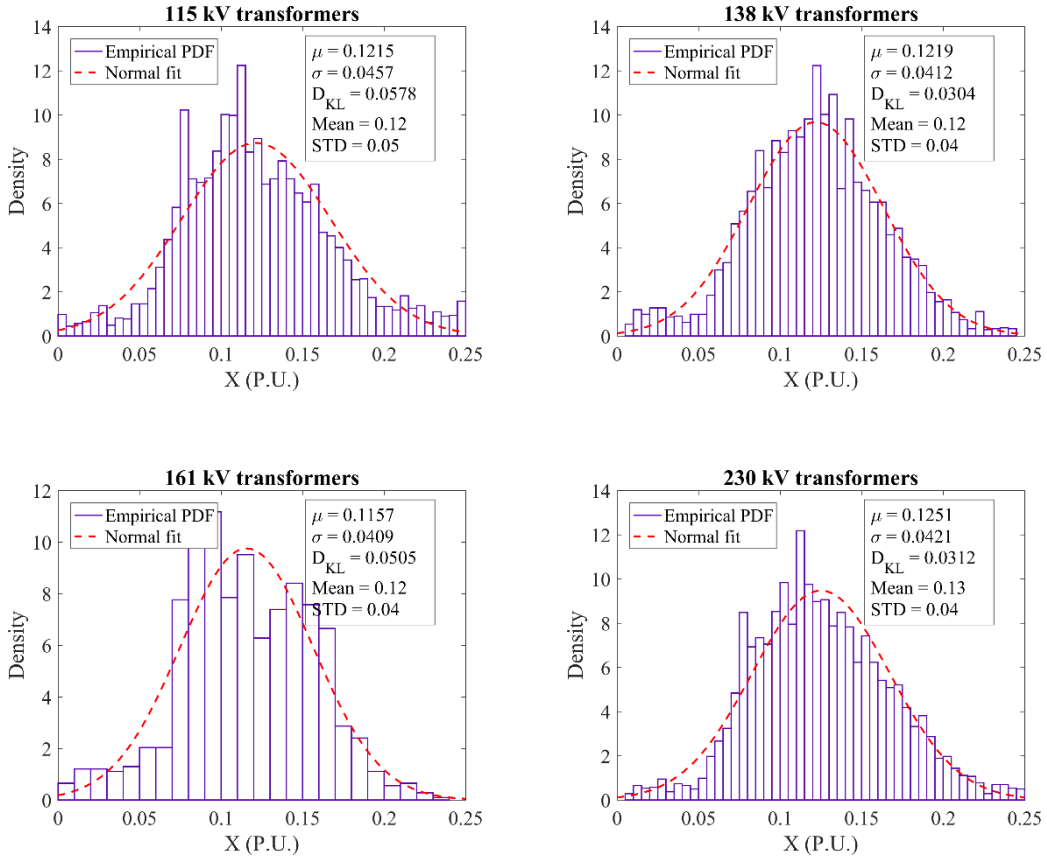


Figure 37. Scatter plot of per unit reactance versus MVA rating of the transformer for 138 and 230 kV transformers

5. *Transformer capacity distribution*

Another key parameter of a transformer especially in gird vulnerability studies is its capacity or MVA rating. For the set of data from real-world power grids, there are transformers with different sizes from couple MVA to +1000 MVA. Also, due to the lack of detailed information in some cases, the MVA rating of some transformers are set to very large or small values. To exclude such cases, in addition to identifying the full range of transformer MVA rating, an 80% range centered at the median is defined to get rid of "extreme values" on both upper and lower bounds. This will give us a more useful range where most transformers fall in. Table XIV shows the median, mean, minimum and maximum range, and 80% range for transformers MVA ratings.

Table XIV. MVA rating statistics for 115, 138, and 230 kV transformers

	Transformer MVA rating			
Voltage Levels (kV)	**Median**	**Mean**	**Range**	**80% range**
115	53	46.68	[3, 683]	[22, 140]
138	83	51.72	[3.3, 782]	[39, 239]
230	203	145.17	[10, 1610]	[62.5, 470]

Figure 38 depicts the empirical PDF of transformers MVA rating and the approximated fit distribution for 115 kV transformers. Note that the results for 138 kV and 230 kV transformers will be presented later in a table. According to the KL divergence, transformers capacity is approximated with Generalized Extreme Value (GEV) distribution with the minimum D_{KL} value where its CDF is represented by (41)

$$F(x|\zeta,\mu,\sigma) = exp\left(-\left(1+\zeta\frac{(x-\mu)}{\sigma}\right)^{\frac{-1}{\zeta}}\right) \tag{41}$$

where μ is location parameter, σ is scale parameter, and $\zeta \neq 0$ is shape parameter. Using this mathematical distribution, one can generate reasonable values for transformer capacities in a given synthetic grid model.

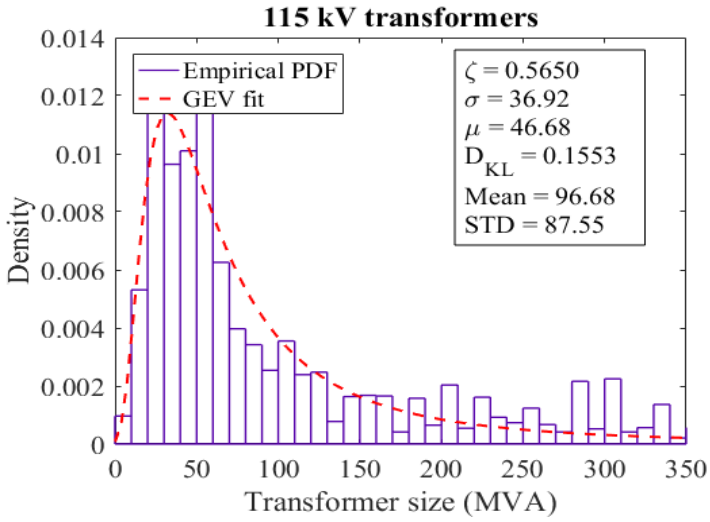


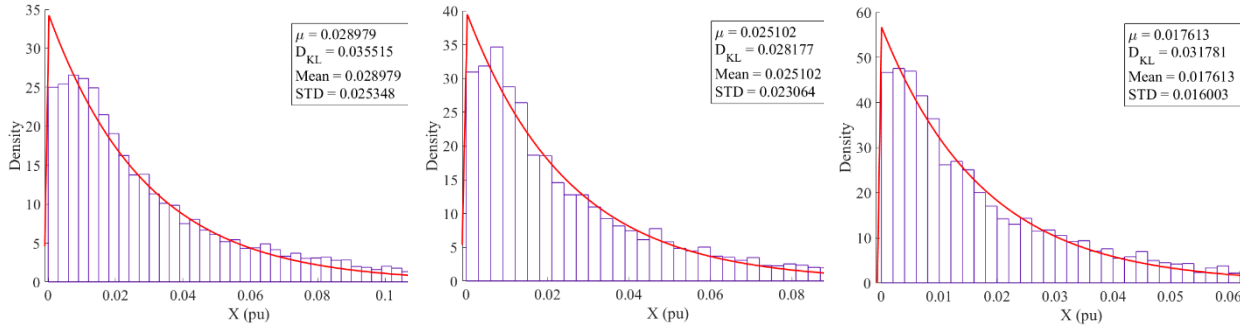
Figure 38. Empirical PDF and GEV-fit of MVA rating for 115 kV transformers

B) *Transmission lines*

Transmission line parameters statistics are derived using over 50000 lines from real power systems. Transmission lines are categorized based on their nominal voltage level which ranges from 0.6 to 765 kV. Here we study lines with nominal voltage levels of 115, 138, 161, and 230 kV. We studied per unit reactance, X/R ratio, and line capacities as three critical parameters of transmission lines to provide several validation metrics and guidelines for synthetic grid modeling.

1. *Transmission line per unit reactance distribution*

Figure 39 shows the empirical PDF of the transmission line per unit reactance and the approximated fit distribution for different voltage levels.



115 kV 138 kV 230 kV

Figure 39. Empirical PDF and Exponential-fit of per unit reactance for 115, 138, and 230 kV transmission lines

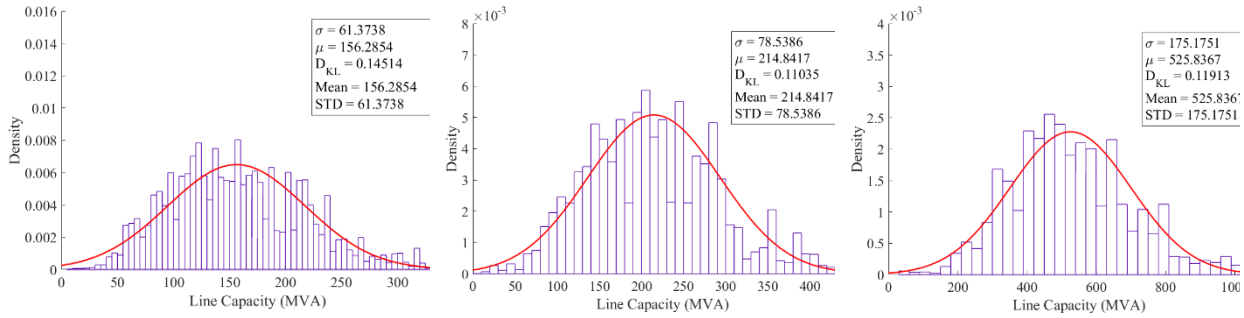
It is found that for all three voltage levels, per unit reactance is mostly less than 0.02 p.u. and the density drops exponentially as reactance increases. According to the KL divergence, transmission line reactance is approximated with Exponential distribution with the minimum D_{KL} value where its PDF is represented by (42)

$$f(x|\mu) = \frac{1}{\mu} e^{\frac{-x}{\mu}} \tag{42}$$

Using this mathematical distribution, one can generate reasonable values for the transmission line per unit reactance in a given synthetic grid model. Note that, the distribution of per unit reactance for transmission lines is very different from Normal distribution for those of transformers. This is because of per unit conversion for transformers and implies that in order to have a more stabilized range for lines reactance, it is better to study their actual distributed reactance (Ω/km). This will be presented in our next comprehensive study.

2. *Transmission line capacity distribution*

Transmission line capacity is a critical parameter in various analysis such as optimal power flow (OPF) analysis, contingency analysis, and power grid expansion planning. Therefore, here we studied the distribution of line capacity for different voltage levels to identify a useful guideline and range for actual capacities in the real grids. Figure 40 shows the empirical PDF of transmission line capacity and the approximated Normal distribution with best-estimate parameters based on D_{KL} for three different voltage levels. Note that, unlike transformers, the distribution of MVA rating for transmission lines is approximated with normal distribution with higher mean values for each voltage level.



115 kV 138 kV 230 kV

Figure 40. Empirical PDF and Normal-fit of line capacity for 115, 138, and 230 kV transmission lines

1.2.1 Overall statistics of electrical parameters

The distribution fitting results for all other parameters and variables of the transmission network is presented in Table XV.

Table XV. Distribution fitting results for branch parameters

Parameter name	Fit distribution properties											
	Distribution type				**The goodness of fit (D_{KL})**				**Distribution parameters**			
Voltage level (kV)	115	138	161	230	115	138	161	230	115	138	161	230
Transformer reactance (p.u.)	Normal	Normal	Normal	Normal	0.057	0.041	0.050	0.031	μ=0.121 σ=0.045	μ=0.121 σ=0.041	μ=0.115 σ=0.040	μ=0.125 σ=0.042
Line length (km)	GEV*	GEV	GEV	GEV	0.055	0.061	0.055	0.075	ζ=0.276 μ=5.68 σ=4.37	ζ=0.299 μ=6.63 σ=5.12	ζ=0.226 μ=9.80 σ=7.19	ζ=0.397 μ=10.16 σ=8.95
Line per reactance (p.u.)	Exp.	Exp.	Exp.	Exp.	0.038	0.027	0.068	0.034	μ=0.029	μ=0.023	μ=0.025	μ=0.015
Line distributed resistance ($\boldsymbol{\Omega/km}$)	GEV	GEV	GEV	GEV	0.043	0.039	0.033	0.035	ζ=0.249 μ=0.07 σ=0.06	ζ=0.275 μ=0.06 σ=0.05	ζ=0.175 μ=0.06 σ=0.04	ζ=0-0.11 μ=0.05 σ=0.03
Transformer capacity (MVA)	GEV	GEV	GEV	GEV	0.155	0.087	0.067	0.088	ζ=0.565 μ=46.68 σ=36.92	ζ=0.372 μ=51.72 σ=36.79	ζ=0.219 μ=80.39 σ=49.54	ζ=0.029 μ=145.1 σ=95.55
Line capacity (MVA)	Normal	Normal	Normal	Normal	0.145	0.110	0.262	0.119	μ=156.2 σ=61.37	μ=214.8 σ=78.53	μ=264.0 σ=83.74	μ=525.8 σ=175.1
Transformer X/R ratio	GEV	GEV	GEV	GEV	0.052	0.077	0.074	0.045	ζ=-0.008 μ=25.12 σ=11.54	ζ=-0.028 μ=26.18 σ=10.67	ζ=-0.024 μ=27.42 σ=12.90	ζ=-0.099 μ=37.70 σ=16.41
Line X/R ratio	Normal	Normal	Normal	Normal	0.117	0.103	0.134	0.102	μ=4.472 σ=2.022	μ=5.385 σ=2.330	μ=6.115 σ=2.019	μ=8.689 σ=2.462

*Generalized Extreme Value

1.3 The interdependence of transmission network parameters on voltage level

The use of multiple voltage levels in power systems is a common practice to decrease the energy loss in the transmission network [23]. The multi voltage-level structure of the power grid may cause grid components such as transmission lines and transformers to have various voltage dependence. Studying the interdependence of transmission branch parameters and variables on

voltage levels may provide useful insights as well as multiple validation metrics and tuning criteria for developing synthetic power networks.

The literature review on synthetic grid modeling development and validation studies suggests that there is a need for a comprehensive statistical study on the voltage dependence of transmission network electrical and non-electrical parameters and variables. This study will be useful in providing both validation metrics for existing grid models and generating and tuning new synthetic grid cases.

A) *Data filtering method*

Access to real-world power system data is restricted to researchers due to security reasons. Many publicly available test cases were modified from their original settings to provide an abstract version for testing new algorithms and methodologies. Many of these equivalencing cause critical parameters of the grid to be altered and thus not reflect the actual structural and operational features of the realistic grids. Therefore, here we used a large sample of real-world power system data to extract the statistics of different parameters. These data are collected by FERC and provide a range of parameters and variables for two real-world power networks in North America.

We mainly focus on transmission network that consists of transmission lines and transformers to study the interdependence of various branch parameters and variables on voltage level. Initial observations on the data indicate that for any examined parameter most of its values concentrate within a recognizable region. However, there always exist a small number of *outlier* values that may span an extraordinarily wide range. For instance, a negative impedance often occurs in the star modeling of a three winding transformer due to how the leakage reactance is

measured/modeled [23]. Also, network equivalization methods may create negative impedances which can affect the statistics of transmission network parameters. In order to appropriately deal with such data and avoid erroneous disturbance on statistical analysis, raw grid data are filtered to exclude *outliers* based on boxplot method a useful graphical display for describing the behavior of the data in the middle as well as at the tails of the distributions. Following is the description of a standard box plot presentation of the data:

- *lower quartile (Q1)*: the 25^{th} percentile of the data set
- *upper quartile (Q3)*: the 75^{th} percentile of the data set
- *interquartile range (IQ)*: the upper and lower quartile difference (Q1-Q3)
- *lower inner fence*: defined as Q1-1.5×IQ
- *upper inner fence*: defined as Q3+1.5×IQ
- *lower outer fence*: defined as Q1-3×IQ
- *upper outer fence*: defined as Q3+3×IQ
- *mild outlier*: a point beyond an inner fence (either side)
- *extreme outlier*: a point beyond an outer fence (either side)

The box plot uses the median and the lower and upper quartiles. A box plot is constructed by drawing a box between the upper and lower quartiles with a solid line drawn across the box to locate the median. For this study, we excluded the extreme outliers from the data set to remove their impacts on derived statistics.

Our statistical study on transmission network electrical parameters in the previous section mainly focused on categorizing their empirical probability density functions (PDFs). In this section, we will focus on the interdependence of several parameters and variables of transmission network on

the nominal voltage level. The parameters include transmission lines distributed reactance and resistance (Ω/km), transmission line length (km), transformers per unit reactance ($p.u.$), and transformers and transmission lines capacity (MVA) and X/R ratio and the operation variables that are examined include transmission line real power flow, current, voltage drop, and real power loss. We will examine both interdependence of parameters on voltage level and their PDFs. In addition, we will compare the statistics of three synthetic grid models called ACTIVSg cases that are published in [93] and available in [137]. This is to show a potential application of the statistics presented in this paper and shows how we can tune the critical parameters of a synthetic grid so that they become consistent with real-world networks.

1. *Transformer per unit reactance*

In this study, the transformers are grouped based on their high voltage side into eight categories from 69 to 735 kV. The original data acquired from FERC were reported in per unit values based on the system-wide common base. As found in [134], the transformer per unit X calculated based on its own MVA rating falls within a narrow range that is consistent for all voltage levels. In other words, there exists no interdependence between per unit reactance and the transformers voltage level after this conversion. Figure 41 shows the interdependence of transformer per unit X on voltage level where black dots are average reactance for each voltage level and dashed blue line is the average of all data points. The box plot of the data is also shown in the figure where the range of the data and outliers can be recognized. In order to validate the statistics of ACTIVSg cases, the average transformer per unit reactance for these cases are shown in Figure 41. It is found that all three ACTIVSg cases are within the scope and present independent values from voltage level which is consistent with what is found from FERC data.

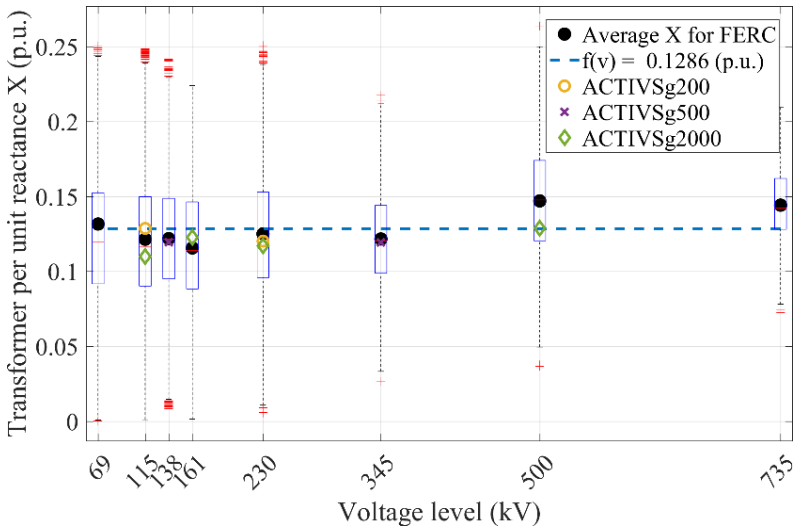


Figure 41. The interdependence of transformer per unit reactance on the voltage level

2. *Transmission line length*

The line length (*km*) data reported from FERC is approximate evaluation using Geographical Information System (GIS) data and great circle method. While this approximation may not exactly reflect the line length, the data can be used to examine the interdependence of average line length on voltage level. This interdependence is shown in Figure 42 where black dots represent the average line length for each voltage level in *km* and blue dashed curve shows a power function that is fit to the data according to minimum Root Mean Squared Error (RMSE) criteria and is formulated as:

$$l(V_B) = 0.001521 \times {V_B}^{1.738} \tag{43}$$

This helps us identify the relationship of line length and voltage level in real-world power systems that can be used in tuning procedure for synthetic grids. Note that power function is selected for curve fitting to simplify the result validation. As shown in Figure 42, all three synthetic cases show

a similar trend with close values to those from real data except 500 kV lines for ACTIVSg2000 where they seem a bit shorter than those are in real networks.

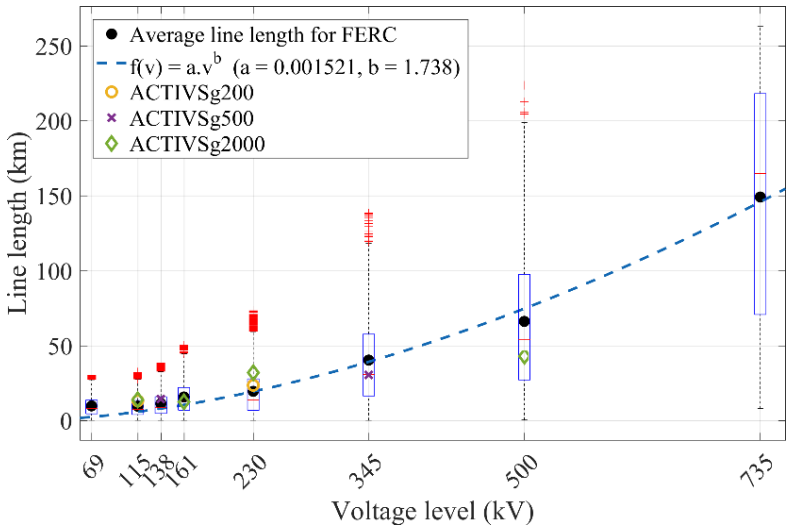


Figure 42. The interdependence of transmission line length on the voltage level

3. *Transmission line distributed reactance*

The transmission line distributed reactance (Ω/km) which is calculated using per unit reactance and line length is another independent parameter from voltage level similar to transformer reactance. The distributed reactance is calculated based on

$$X(\Omega/km) = \frac{X_{pu}.{V_B}^2}{l.S_B} \tag{44}$$

in which using system common base S_B and voltage base V_B for each transmission line the actual reactance in *ohms* is first calculated; then using the approximated line length l in *km*, the distributed reactance in Ω/km is derived. In Fig. 24, similar to transformer reactance, the blue dashed line is the average of all black dots that represent mean distributed reactance for each voltage level. There

is no visible interdependence between distributed reactance and the nominal voltage level of lines meaning that power function representation of distributed reactance is in the form of

$$X_d(V_B) = 0.4174 \times {V_B}^0 \tag{45}$$

It is also shown that for transmission line distributed reactance, the ACTIVSg500, and ACTIVSg2000 cases show comparable values with those of FERC data and there is no visible trend in the data, while the ACTIVSg200 case seems to have some extraordinary large exceptions.

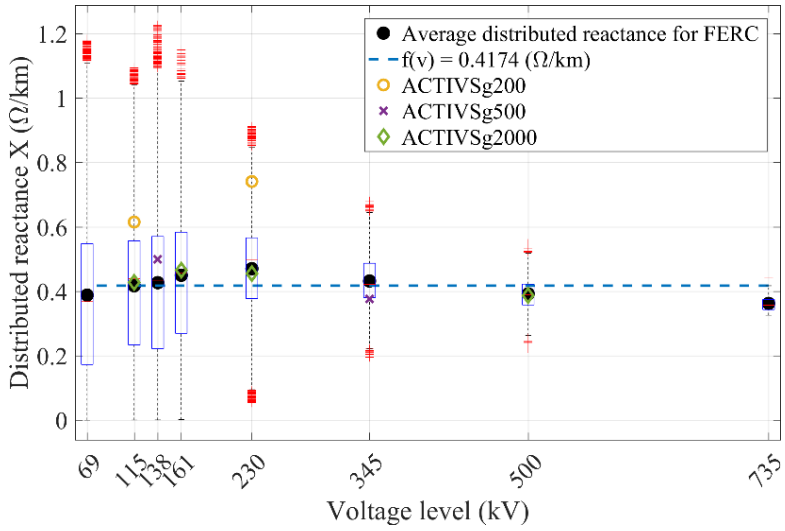


Figure 43. The interdependence of transmission line distributed reactance on the voltage level

4. *Transformer and transmission line capacity*

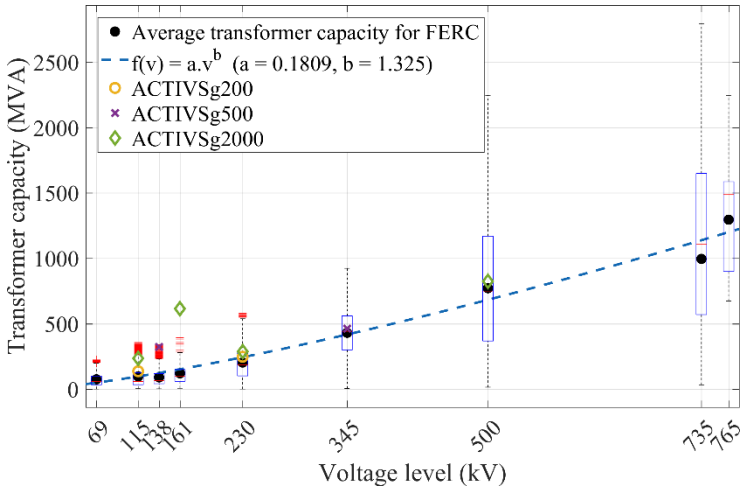
Figure 44 shows the interdependence of transformers and transmission lines capacity on voltage level. For both parameters, there is a visible trend in their capacity and that is the higher the voltage level the bigger the capacity. This is consistent with the common engineering practice in the power systems. The fitted curves based on power function for these parameters are shown in Figure 44 (a) and (b) with mathematical expressions in Eqs. (46) and (47), respectively.

$$S_{trans.}^{rating}(V_B) = 0.1809 \times {V_B}^{1.325} \quad (46)$$

$$S_{line}^{rating}(V_B) = 0.1295 \times {V_B}^{1.565} \quad (47)$$

This can be useful in the capacity assignment for transmission network in the synthetic grid creation. Comparison of transformer and transmission lines capacity in the synthetic cases shows that for transformers although they have comparable capacities, for some voltage levels (161 in ACTIVSg2000 and 138 in ACTIVSg500) we can see oversized transformers. While the capacities of transmission lines in all three synthetic cases are perfectly matching with statistics from real data in terms of both interdependence on voltage level and average capacities.

(a)



(b)

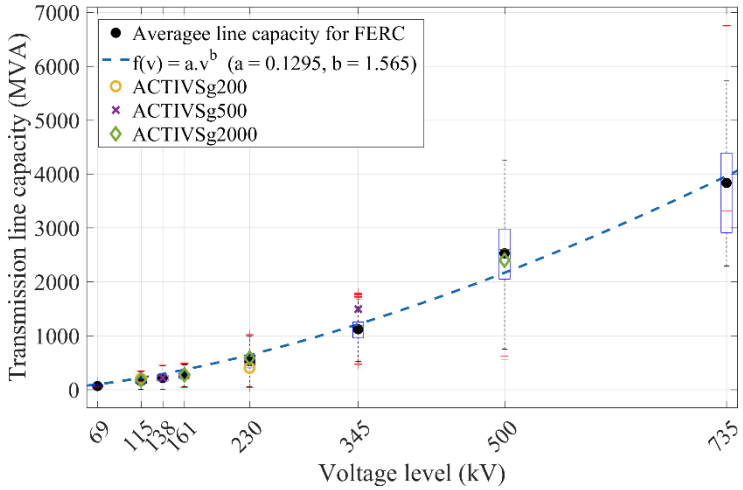


Figure 44. The interdependence of transformer and transmission line capacity on the voltage level

1.4 Introducing voltage-level dependent parameters to synthetic grid electrical topology

The random-topology synthetic grid network proposed in [129], called RT-*nestedSmallWorld*, models the admittance matrix (Y_{bus}) in per unit form and assigns the bus types of the network based on the statistical properties of actual power systems featuring the same kind of small-world *electrical* topology as the real power grids.

However, transformers and transmission lines cannot be distinguished in per unit format while according to [134] these two elements exhibit different electrical characteristics. Also, our voltage-level dependence analyses on the actual power grids reveal a strong correlation between parameters and geographical information. Therefore, the study of transmission network electrical and non-electrical parameters dependence on the nominal voltage could help enhance our random-topology grid model by adding geographical information into the electrical topology of the synthetic grid. A similar approach has been employed in the synthetic grid modeling in [93] to

design the grid topology based on geographical location of population centers and generation and load proportions.

In this section, we introduce a new framework to enhance the RT-*nestedSmallWorld* synthetic grid model by integrating voltage-level dependent parameters to the electrical topology of the developed synthetic network. The proposed methodology consists of two phases as illustrated in Figure 45.

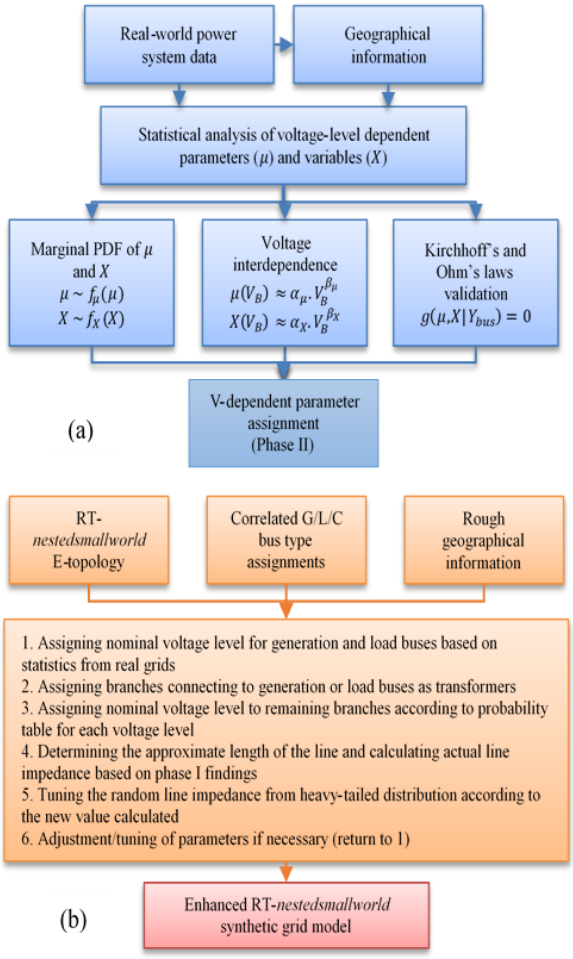


Figure 45. The flowchart of the proposed algorithm for enhanced E-topology: (a) phase I and (b) phase II

A) Phase I of the proposed methodology

A set of statistical analyses will be performed on a large set of real-world power system data to extract critical characteristics of the power systems for enhancing the electrical topology of the synthetic grid. This phase includes the following three steps to extract the necessary information for the next phase as shown in Figure 45 (a).

- *Step 1*: The empirical marginal probability density function (PDF) for individual transmission network parameters and variables will be examined and a standard distribution function will be fitted to the empirical PDF. This is used to assign random parameter values to synthetic grids so that they exhibit consistent statistics with actual grids.

- *Step 2*: Parameters and variables of the transmission network are categorized based on their nominal voltage level. The transformers are categorized based on their nominal high-voltage. Then, the average parameter/variable value is studied to mathematically formulate how it scales with voltage level. The relationship is modeled based on curve fitting using power function in the form of $f(V_B) = \alpha . {V_B}^{\beta}$, where V_B is the nominal base voltage in *kV*.

- *Step 3*: The physical constraints by *Kirchhoff's* and *Ohm's* laws are used to verify the voltage/parameter relationships extracted from empirical data in step 2. This helps make sure that the mathematical representation of voltage dependence is consistent with the actual constraints that are unique to the power grids.

B) Phase II of the proposed methodology

The implementation of the proposed methodology is the next step to enhance the E-topology of the developed synthetic grid in [138]. This can be achieved by integrating rough geographical data in the form of line lengths and use the findings of the phase I to tune the electrical properties of

the synthetic grid topology for an accurate representation of the power system. We start with the random topology network generated based on the algorithm proposed in [129]. The initial line impedances are randomly generated from a specified heavy-tailed distribution, and then sorted by magnitude and group into local links, rewire links, and lattice connection links according to corresponding portions derived from statistical analysis of actual grids. The line impedances in each group are assigned at random to the corresponding group of links in the topology. Next, based on the following algorithm, we tune the line impedances and at the same time introduce additional details to our synthetic grid in the form of network nominal voltage levels and line lengths.

- *Step 1*: The statistical studies on FERC data shows that in power grid networks, generation and load (G/L) buses tend to have low nominal voltage levels that are because of engineering practices in generators design. Table XVI shows the top six nominal voltage levels for G/L buses in FERC data and their respective shares in the entire North American network. The nominal voltage levels for G/L buses are randomly assigned based on each level's probability.

Table XVI. Nominal voltage levels and their respective probability for FERC data

Generation buses		**Load buses**	
Nominal voltage (kV)	**Probability**	**Nominal voltage (kV)**	**Probability**
13.8	0.40	69	0.27
18	0.09	115	0.21
69	0.05	138	0.12
13.2	0.03	13.8	0.05
34.5	0.03	34.5	0.04
115	0.03	161	0.04

- *Step 2*: According to the analysis of actual grid data and common engineering practice in power systems, each branch connecting either generation or load buses to connection buses is, in fact, a transformer. Therefore, the transformers are placed and accordingly their per unit reactance is assigned based on the stable range identified in phase I (shown in Figure 41).

- *Step 3*: The remaining branches represent the high voltage transmission network. Figure 46 shows the per unit line impedance (Z_{pr}) and approximated length for FERC data. It is found that the scatter plot shows a high correlation between the nominal voltage level of the line and its Z_{pr} so that each voltage level represents a cluster of data. The range of line impedance can be divided into three zones representing small, medium and large values. Then, for each branch in our developed synthetic network, the nominal voltage level is selected according to the probability of each level shown in Table XVII. The impedance zone to choose from is determined based on the Z_{pr} value generated randomly from the heavy-tailed distribution in our initial synthetic grid. Note that Table XVII is derived from actual grid data and zonal analysis shown in Figure 46.

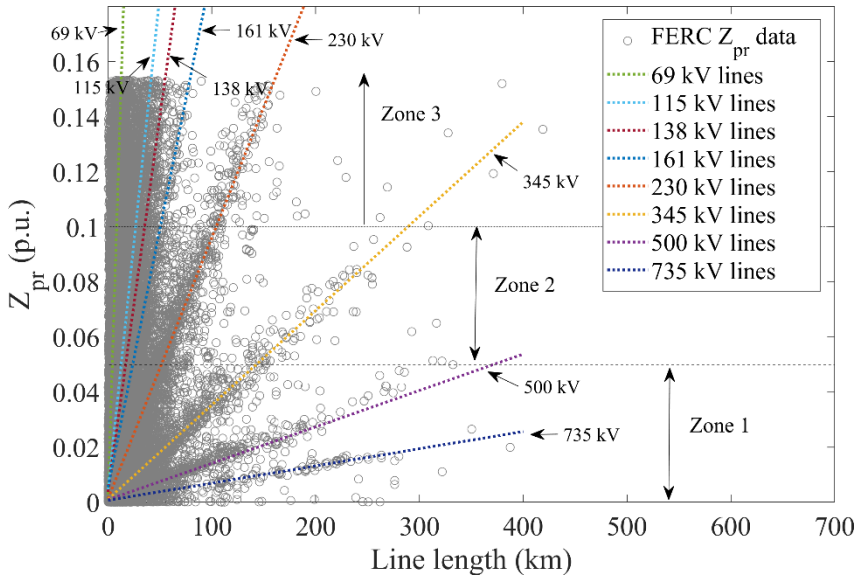


Figure 46. The scatter plot of per unit impedance and approximate line length for transmission lines of North American power network from FERC data

Table XVII. Nominal voltage level selection probabilities for each impedance zone

	Probability of selection		
Nominal voltage level (kV)	**Zone 1**	**Zone 2**	**Zone 3**
69	0.155	0.341	0.454
115	0.281	0.240	0.183
138	0.202	0.135	0.081
161	0.060	0.052	0.023
230	0.132	0.056	0.036
345	0.036	0.008	0.001
500	0.018	0.0002	0
735	0.002	0	0

- *Step 4*: Given the nominal voltage level of the transmission line, one can estimate an approximate length for the line according to the linear regression model fitted to each cluster of impedance data in Figure 46. Next, the analysis conducted in phase I helps to calculate the actual line impedance using the average distributed reactance and resistance (Figure 43) and the approximate length as $Z_{actual} = l \times Z_d$, where l is the approximate line length (km) and Z_d is the average distributed impedance. Note that in transmission networks usually, the line impedance is dominant, hence the resistance can be neglected in these calculations.

- *Step 5*: The new Z_{pr} is then calculated by converting the actual impedance to the per unit using impedance base $Z_B = \frac{V_B^2}{S_B}$ with S_B as the system common base.

The proposed approach allows for the integration of geographical data to the synthetic grids in the form of line approximate length and at the same time, it helps assign nominal voltage levels to the transmission network. This will enhance the E-topology of the synthetic grid since a strong correlation has been found between the nominal voltage level of the network and various electrical and non-electrical parameters. A similar approach can be employed to tune/adjust other parameters of the network such as branch capacities once the voltage levels are known. Finally, the validation

process based on DC power flow solutions will follow and if necessary tuning/adjustment of the parameters will be performed to ensure having a feasible power flow solution and similar characteristics for the transmission network as real grids.

1.5 GridStat Analysis Toolkit

The popularity of statistical analysis on power system along with a diverse range of topological and electrical features of the power grid motivated us to develop a framework that offers a systematic grid statistical analysis toolkit. The toolkit provides an interactive input-output structure where the user can perform a wide range of statistical analyses on a power grid and compare its statistics with a reference grid. In this paper, we introduce our *GridStat Analysis Toolkit* developed using MATLAB Graphical User Interface (GUI), present its wide range of functions, and then demonstrate its usability by running analysis on three different categories of power system networks. We then discuss and identify critical metrics of power system networks which are consistent for various networks and will serve as the minimum requirement to pass the validity test for synthetic grids.

GridStat Analysis Toolkit is designed to perform a wide range of statistical analyses on power system networks. The tool receives network data for the input grid and optionally for a reference grid in the standard power system data format. Following discusses data format, critical analytics incorporated into the developed toolkit, and an overview of the toolkit.

1.5.1 Data format

The data files used by GridStat Analysis Toolkit are Matlab M-files or MAT-files which define and return a single Matlab struct. Matlab M-files are plain text and thus can be edited by any

standard text editor. The power system cases that the toolkit accepts are MATPOWER cases. MATPOWER is an open-source MATLAB-based power system simulation package that provides a high-level set of power flow, optimal power flow (OPF), and other tools targeted toward researchers, educators, and students [89]. In this format, the fields of Matlab struct are bus, branch, and gen which are matrices and baseMVA which is a scalar. The rows of the matrices correspond to a single bus, branch, or a generator while the columns correspond to various parameters similar to the standard IEEE CDF and PTI formats. Note that, any power system case can be easily reformatted to MATPOWER data format to input to the toolkit.

1.5.2 Distribution fitting

One of the important functionalities that the toolkit provides is plotting the empirical distribution of a wide range of electrical and non-electrical parameters of the grid and then fit a standard distribution function to the empirical data to mathematically model the distribution. The type of the distribution to fit is determined based on the goodness of fit meter which indicates how closely a particular distribution form represents the empirical data. There are several statistical methods to measure the goodness of the fit such as Kolmogorov-Smirnov test and Jarque-Bera test for normality check and the Kullback-Leibler (KL) divergence criteria to measure the similarities of the fit distribution and the empirical data. In our toolkit, the goodness of fit is measured by KL divergence and based on the best score (i.e. the distribution with the smallest KL divergence value) the best fit is selected [134].

1.5.3 GridStat Analysis toolkit overview

The power grid has unique features impacted by geographical and demographic characteristics of the society it is built in. In general, the topology of a power grid network can be represented as an undirected graph with *small-world* properties [127]. Statistical analysis of both topological and electrical metrics of a grid can reveal insightful information about the unique structure of these networks. Next step would be to mathematically model this statistical knowledge to build a set of unique, special, and characteristic set of metrics that can be used for validating various models and algorithms. Branch impedance, capacity, generation, and load settings, and loading status of transmission lines are among the properties with identified patterns. Also, different parameters in an actual transmission network have a correlation with other parameters. The connecting point for many of these correlations is the nominal voltage level of the network.

This toolkit is designed with four main functions to statistically characterize any given power system and identify their distinguished patterns for better insights into unique features of power systems. These four functions (tabs) include the *topological analysis, grid parameter statistics, nominal voltage interdependence*, and *grid scaling properties*. Below we briefly discuss the functionalities of each tab in the toolkit.

A) Topological analysis

The power grid networks are generally a graph with known *small-world* properties and sparse connection. They have shorter average path length (in hops) and a much higher clustering coefficient, compared to similar size random graph networks [127]. There exist two main distinctions between power grids and *small-world* networks though: first, they have a very small average node degree ($\langle k \rangle$ =2~5) due to their sparsity which does not scale with network size. Second, the power grids exhibit a very special scaling properties in terms of their connectivity

metrics. In addition, the nodes in a power grid unlike other networks have types. They can be either a generation, a load, or a connection bus that cannot be interchanged with each other. In [139] a new topological metric is defined based on this special characteristic called *bus type entropy* that shows a unique statistical property.

Figure 47 shows the Input/Topology tab of the developed toolkit. The statistical analysis in the GridStat Toolkit starts with Input/Topology tab where the user loads the input grid for analysis. The user also can optionally load a reference grid (e.g. an actual grid with similar size or a standard case) to compare the statistics of the two grids. Note that the reference grid will serve as the comparison reference in the other tabs. The "Analyze" button will perform a set of statistical and topological analyses to extract the characteristic topological metrics of each grid. Because the calculation of "Average path length" can be lengthy for larger network sizes based on Dijkstra's algorithm [140], by default the toolkit will not calculate this metric unless the user selects "Average path length" checkbox provided. As an example, on a machine with Intel Core i7 @ 3.60 GHz CPU it takes 1.25 seconds to calculate all topological metrics except the "Average path length" for PEGASE2869 with 4582 links while if the checkbox is selected, it takes 30 minutes. Fig 1 shows the topological comparison of two similar size networks. ACTIVSg2000 is a synthetic grid and case2383 is an actual power system from MATPOWER's database.

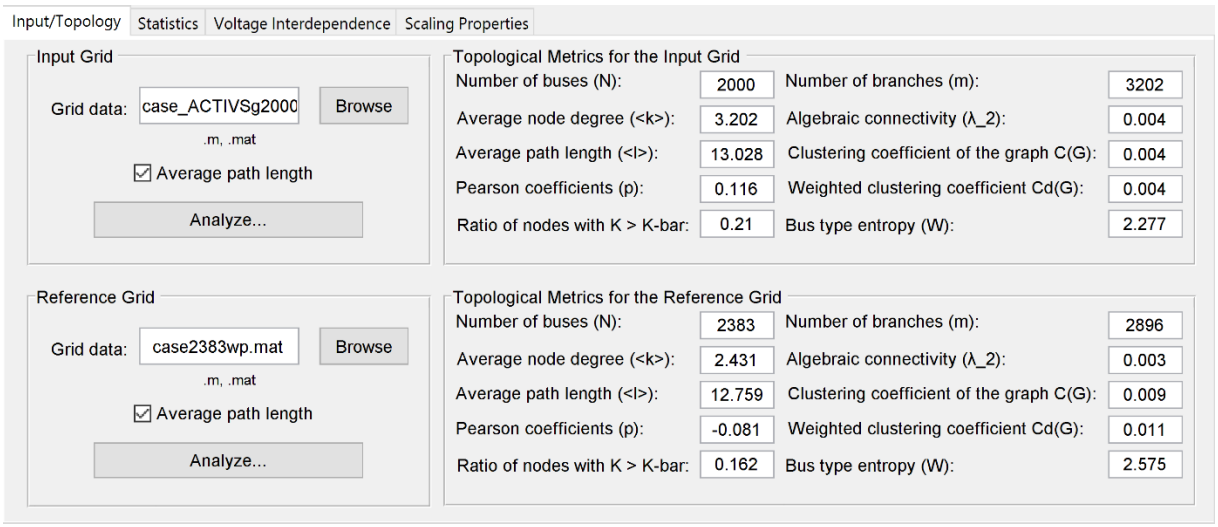


Figure 47. Topological analysis tab of the developed toolkit

Next, the topological parameters that can be analyzed by the toolkit will be briefly discussed.

1) Average node degree $\langle k \rangle$: The node degree of bus *i* in a grid equals the total number of branches it connects and can be obtained from the *i*th diagonal entry of the Laplacian matrix, i.e., $k_i = L(i,i)$ where $L = A^T A$ is the Laplacian matrix and $A \coloneqq (A_{l,k})_{m,N}$ is the branch-node incidence matrix, arbitrarily oriented and defined as: $A_{l,i} = 1;\ A_{l,j} = -1$, if the *l*th branch is from node *i* to node j and $A_{l,k} = 0, k \neq i,j$. Then the average nodal degree of the grid is calculated as $\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} L(i,i)$ [129].

2) Average path length $\langle l \rangle$: Given the connecting topology of a grid, we can run the Dijkstra's algorithm to calculate the shortest path length measured in hops between any two buses *i* and *j*, i.e., l_{ij}. Then the average shortest path length of a grid is $\langle l \rangle = \frac{2\sum_{i,j} l_{ij}}{N(N-1)}$ [129].

3) The ratio of buses with $k_i > \langle k \rangle$ *(*κ*):* This ratio shows how sparse a power grid network is and it is defined as $\kappa = \frac{\sum \sigma_j}{N}$ where

$$\sigma_j = \begin{cases} 1 & if\ k_j > \langle k \rangle \\ 0 & otherwise \end{cases} \tag{48}$$

4) Algebraic connectivity (λ_2): A topology measure that is the second smallest eigenvalue of the Laplacian matrix, $\lambda_2(L)$, with $[\lambda_1, \lambda_2, \dots, \lambda_N] = \text{Eigen}(L)$ [129] and reflects the overall connectivity of a network. Note that the smallest eigenvalue of the Laplacian matrix is always zero and the number of zero eigenvalues in the Laplacian determines the total number of islanded areas in the network.

5) Clustering coefficient $C(G)$: The clustering coefficient is defined by Watts and Strogatz [127] as the average of the clustering coefficient for each node, i.e. $C(G) = \frac{1}{N}\sum_{i=1}^{N} C_i$ where $C_i = \frac{\lambda_G(i)}{\tau_G(i)}$, $\lambda_G(i)$ is the number of edges between the neighbors of node *i* and $\tau_G(i)$ the total number of edges that could possibly exit among the neighbors of node i. For undirected graphs, obviously $\tau_G(i) = k_i(k_i - 1)/2$ given k_i as the node degree.

6) Bus type entropy $W(\mathbb{T})$: It is a characteristic feature showing the bus types and link types in a power system as a scalar measure. It is defined as [139]

$$W(\mathbb{T}) = -\sum_{i=1}^{N} \log\left(r_{\mathbb{T}_i}\right) - \sum_{j=1}^{m} \log(R_{\mathbb{L}_j}) \tag{49}$$

where $r_{\mathbb{T}_i} = n_{\mathbb{T}_i}/N$ represent the bus type ratio of bus *i* and $R_{R_{\mathbb{L}_j}} = m_{R_{\mathbb{L}_j}}/m$ the corresponding link type ratio of the *j*th line; n_k and m_k representing the total number of buses and lines of different types in the grid that have some specified types respectively. The *bus type vector* is $\mathbb{T}_i \in \{1,2,3\}$ where values assigned to generation, load, and connection buses, respectively. Similarly,

$R_{\mathbb{L}_j} \in \{1,2,\dots,6\}$ assuming each transmission branch is one of the following link types, i.e. {*GG, GL, GC, LL, LC, CC*}, respectively.

7) Normalized bus type entropy distance $d_W(\mathbb{T}^*, \widetilde{\mathbb{T}})$: It is the normalized distance of the bus type entropy for correlated bus type assignments $\{\mathbb{T}^*\}$ and other randomized assignments based on permutation $\widetilde{\mathbb{T}} = \wp(\mathbb{T}^*)$. By the central limit theorem, the randomized entropy values may assume a normal (Gaussian) distribution. With the extracted distribution parameters (μ, σ), a normalized distance can be defined to measure the difference between $\mathbb{T}^*$ and $\widetilde{\mathbb{T}}$ as:

$$d_W(\mathbb{T}^*, \widetilde{\mathbb{T}}) = \frac{W(\mathbb{T}^*) - \mu}{\sigma} \tag{50}$$

B) Grid parameters statistics

Figure 48 shows the second tab of the developed toolkit that is designed to individually analyze various electrical and non-electrical parameters of the input and reference grids. Here the "Analyze" button will calculate and prepare the empirical distribution of each parameter. Next, the user can select a parameter or a variable to plot its distribution. The list of examined parameters and variables with well-identified statistical patterns is given as follows: node degree, bus type entropy for randomly permuted G/L/C assignments, load and generation capacity, SIL fraction, MVA/SIL ratio, loading factor, line and transformer capacity (MVA), branch flow (MW), line power loss (MW), line and transformer parameters such as X and R in distributed and per unit forms, X/R ratios for transmission lines and transformers, branch angle difference, line length, and line voltage drop.

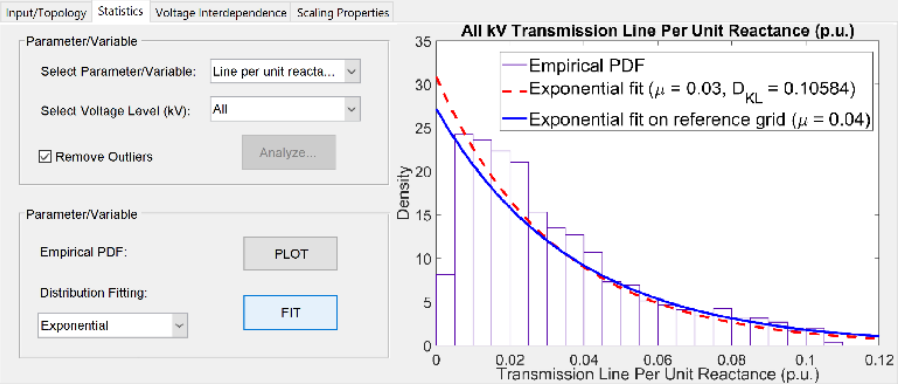


Figure 48. Parameter statistics tab of the developed toolkit

If the loaded power system cases provide the nominal voltage level information, then the toolkit also has the ability to perform the statistical analysis on a certain voltage level network. The second popup menu lists all the nominal voltage levels of the input grid. Note that it is possible that a grid test case may not include all the data of listed parameters for analysis. In that case, the corresponding analysis function will be disabled in the toolkit. Sometimes, network equivalisation or record inaccuracy may introduce erroneous values for some parameters (e.g. impedance) which can affect the statistical evaluation of those parameters. In order to appropriately deal with data with abnormal outliers which stay far away from recognized "normal range" and avoid erroneous disturbance on statistical analysis the checkbox "Remove Outliers" is provided in the toolkit. To remove outliers, the *interquartile range* (*IQ=Q1-Q3*) is defined where *Q1* is the 25th percentile of the dataset and *Q3* is the 75th percentile. Then thresholds *Q1-3IQ* and *Q3+3IQ* are defined as *lower* and *upper outer fences*, respectively. Finally, a point beyond an outer fence (either side) is identified as an *extreme outlier* and removed from the dataset. This will help find a better standard PDF function for the empirical distribution. Plotting the PDF of the selected parameter/variable will activate the "FIT" button which allows the user to select a standard fitting function and display

it over the empirical PDF. Below we discuss select parameters and variables that can be studied for their empirical distributions.

1) Node degree distribution: The analysis shows that node degree distribution has an exponential tail similar to that of the geometric distribution. However, in [129] it is found that for small node degrees (e.g. $k \leq 3$), the empirical probability mass function (PMF) curve clearly deviates from that of a geometric distribution. Therefore, characterizing the node degree distribution based on node type seems more reasonable. The analysis on two North American power grids suggests that the node degree distribution can be expressed as a sum of a truncated geometric random variable and an irregular discrete variable.

2) Bus type entropy for random assignment: The bus type entropy can be calculated for a given network with known topology and generation/load settings according to (10). We then can generate a synthetic power grid network with random bus type assignments that has the same bus type and link type ratios. If we plot the empirical PDF of bus type entropy for a large number of random assignments, we will see that the target entropy (entropy of the actual grid) locates on the far left side of average bus type entropy for random bus type assignments with the Normal distribution. This implies that the probability of generating a random bus type assignment as T^* that has an entropy of W^*, tends to be very small and in fact, realistic power grids assume some "special" or correlated bus type assignment instead of a random one [139].

3) Load and generation capacity: The statistical analysis on generation capacity and demand within actual power grids like NYISO-2935 and WECC-16994 in [141] suggests that 99+% of the generation units (also loads) exhibit exponential distribution with less than 1% showing extremely large capacities (or demands) that are outliers.

4) SIL fraction: The surge impedance loading (SIL) fraction $\gamma = \frac{F_{ij}}{SIL_{ij}}$ is the ratio between the actual power flow of each line and the SIL of the line calculated as $SIL = S_b\sqrt{\frac{B\ [p.u.]}{X\ [p.u.]}}$, where S_b is the system base power, B is line per unit susceptance, and X is line per unit reactance. This variable can show the usage of the transmission line with respect to its SIL. It is found that the empirical PDF of SIL fraction can be best fit by exponential distribution based on actual data from FERC.

C) Nominal voltage interdependence

Earlier we investigated the interdependence of various branch electrical and non-electrical parameter on the nominal voltage level of the network and found that some parameters exhibit a strong correlation with the nominal voltage level while others are almost independent of the nominal voltage level. Recently, we introduced voltage level dependent parameters into our *nestedSmallWorld* random topology synthetic grid model in which by mathematically modeling the voltage interdependence we can more accurately model power system parameters. This also helps us incorporate geographical information of the grid to our synthetic networks since we found a strong correlation between geographical parameters and nominal voltage levels.

Figure 49 shows the "Voltage interdependence" tab of the developed toolkit where one can analyze various parameters in terms of their relationship with the underlying nominal voltage level, given that the nominal voltage information is present in the input grid. It gives two options to the user: the box plot of the selected parameter/variable and interdependence on the nominal voltage level. The box plot is a useful graphical display for describing the behavior of the data in the middle as well as at the ends of the distributions. On each box, the central red mark indicates the median, and the bottom and top edges of the box indicate the 25th (*Q1*) and 75th (*Q3*) percentiles,

respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the red '+' symbol. The “Box Plot” button displays the box plot of the selected parameter to visualize the data and detect the range, median, and outliers in the original data of the input grid. The interdependence button will calculate the average values of the selected parameter for each nominal voltage level and then fit a power function in the form of $f(v) = a.v^b$ to the data to mathematically model them. In Figure 49 the interdependence of line capacity on the nominal voltage level and the fitted curve is shown for “ACTIVSg2k” synthetic case.

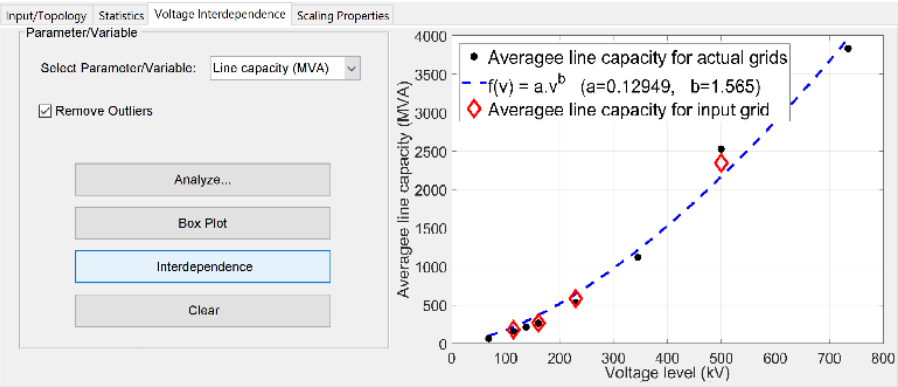


Figure 49. Nominal voltage interdependence tab of the developed toolkit

D) Grid scaling properties

As mentioned earlier, one can identify special *small-world* properties and electrical parameter settings for power systems. In [139] a number of real-world power system networks are studied to identify the scaling properties of topological and electrical parameters of the grid. Among several parameters and variables, some exhibit very clear scaling properties including the average node degree, the average path length, algebraic connectivity, generation capacity, total demand, and transmission network capacity. As found in [129] the average node degree of a typical power grid

does not scale with the network size but remains within a very strict range. The algebraic connectivity of a grid also exhibits some special scaling property. Figure 50 shows the Scaling properties tab of the developed toolkit with the plot showing the algebraic connectivity of "ACTIVSg2k" with respect to that of real grids.

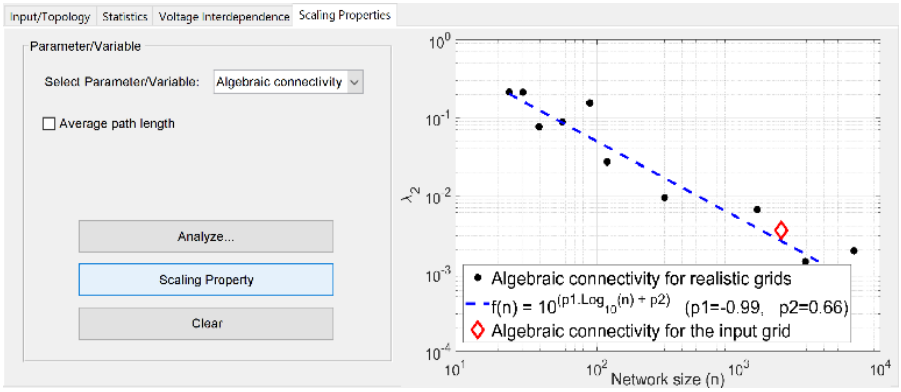


Figure 50. Grid scaling properties tab of the developed toolkit

The scaling properties of the input grid is analyzed and validated using a set of actual grids including the IEEE test cases, PEGASE systems that represent some European nation's grid at different levels of network reduction, the North American power grid (FERC data), and the RTE system which is an equivalent of the French Grid. All of these cases are available in MATPOWER database except FERC data which are not publicly available cases.

1.6 Case studies and discussions

In this section, several case studies including IEEE standard test cases, real North American and European power grids, and synthetic networks are analyzed using the toolkit to extract featured metrics and showcase some of the functionalities of the toolkit. Then, we present a list of key characteristic metrics for power system networks that can serve as the minimum requirement to validate the synthetic grids.

1.6.1 Statistical analysis on sample networks using the toolkit

Table XVIII lists the topological metrics for the select grids in the three categories. The topological metrics can contain critical information about the power system network. Some of these metrics are critical in validating the artificial and random topology power grid networks developed recently. They also may contain insightful information about the underlying geographical parameters of each grid and are unique. In general, the scaling properties of power system grids can be clearly identified in topological metrics. This is another critical piece of information that can be utilized in tuning bigger synthetic cases. According to Table 1, the following observations can be made:

- Average node degree has a weak dependence on network size and does not scale. For the standard test cases, the average node degree is with 2.7 to 3.1 range. The IEEE-145 is an outlier with an average node degree of 6.24.
- Algebraic connectivity shows a strong scaling property for all three categories. The algebraic connectivity of the network decreases as the size of the network increases. IEEE-145 shows an outlier for this metric which can be explained by its large number of branches.
- Clustering coefficient and weighted clustering coefficient both show scaling properties with network size for all three categories. The two topological parameters decrease as the network size scales up. Again, IEEE-145 is an outlier here with a larger coefficient due to its higher number of branches.
- The bus type entropy is the most stable topological metric. It does not exhibit scaling property with network size and is within a narrow band for networks of different sizes. The bus type entropy is a unique feature for power system networks and can be utilized for validation

purposes.

Table XVIII. Comparison of topological metrics of different power system cases

	Topological Metrics									
	Case	N	m	$\langle k \rangle$	λ_2	$C(G)$	$C_d(G)$	ρ	$W(\mathbb{T})$	κ
Standard test cases	IEEE-14	14	20	2.857	0.458	0.367	0.354	-0.074	2.22	0.357
	IEEE-30	30	41	2.733	0.212	0.235	0.219	-0.087	2.503	0.233
	IEEE-57	57	80	2.807	0.092	0.122	0.135	0.19	2.445	0.228
	IEEE-118	118	186	3.153	0.028	0.165	0.157	-0.052	2.387	0.195
	IEEE-145	145	453	6.248	0.107	0.537	0.529	0.106	2.692	0.179
	IEEE-300	300	411	2.74	0.009	0.086	0.097	-0.214	2.667	0.247
Synthetic grids	ACTIVSg200	200	245	2.45	0.023	0.037	0.053	-0.395	2.475	0.21
	ACTIVSg500	500	597	2.388	0.008	0.017	0.025	-0.214	2.106	0.176
	ACTIVSg2000	2000	3202	3.202	0.004	0.004	0.004	0.116	2.277	0.21
	ACTIVSg10k	10000	12706	2.541	0	0.015	0.022	0.139	2.493	0.201
	ACTIVSg25k	25000	32230	2.578	0	0.019	0.03	0.24	2.186	0.216
Real grids	PEGASE89	89	210	4.719	0.154	0.192	0.334	0.405	1.873	0.202
	PEGASE1354	1354	1991	2.941	0.007	0.056	0.06	-0.044	2.697	0.143
	PEGASE2869	2869	4582	3.194	0.001	0.087	0.094	0.031	2.632	0.155
	WECC	20131	25156	2.499	N/A	0.027	0.035	0.102	2.4	0.152
	EI	62605	80595	2.575	N/A	0.029	0.035	0.028	2.353	0.154

Table XIX shows the average value of select parameters in different nominal voltage levels for the cases that the nominal value of voltages is available. These results are generated by the voltage interdependence tab of the toolkit and extracted to the table. According to the table, the following observations can be made:

- The average branch flow in MW is larger in higher voltage networks. This is consistent with the engineering design of power grid networks where higher voltage networks are intended for the transmission of a large quantity of power while lower voltage networks are more for distribution of the power. This pattern is fairly consistent for all three categories except the IEEE-300 where the 138 kV network has the highest flow average among all cases.
- The average power loss shows an increasing trend with respect to the nominal voltage level for the standard cases and actual grids. This can be explained by the fact that higher voltage networks tend to carry greater currents thus assuming similar material in transmission lines, the

active power loss would be higher. The “ACTIVSg10k” and “ACTIVSg25k” do not show a similar trend and they are lightly loaded in 138, 161, and 230 kV networks.

- While the average angle difference across the network branches shows an increasing trend for actual grids (WECC and EI), we cannot clearly identify a similar pattern for other cases.
- The average parameter values for each individual voltage level exhibits a large range meaning that these operational parameters are heavily dependent on underlying geographical and social factors associated with each grid.

The above statistical analyses demonstrate examples of the type of studies one can perform with the developed toolkit and identify distinguished patterns among different power systems, check the consistency and validity of synthetic grids, and possibly tune their electrical and topological parameters to match with actual grids. These analyses also provide insightful information about the complex nature of power system networks that differentiate them from other networks such as communication and traffic networks. Next, we present our list of key characteristic metrics of power system networks identified by extensive statistical analyses on a large set of real and artificial power system networks.

Table XIX. Comparison of select parameter voltage interdependence for different power system cases

Select parameter voltage interdependence																		
Parameter	**Average branch flow (MW)**						**Average power loss (MW)**						**Average angle difference (º)**					
Nominal kV	115	138	161	230	345	500	115	138	161	230	345	500	115	138	161	230	345	500
IEEE-118	-	33	-	-	164	-	-	0.41	-	-	1.87	-	-	2.11	-	-	3.52	-
IEEE-145	-	-	-	-	-	160	-	-	-	-	-	1.12	-	-	-	-	-	1.58
IEEE-300	46	168	-	154	247	-	0.57	0.47	-	1.23	1.39	-	2.89	1.99	-	3.71	2.47	-
ACTIVSg200	14	-	-	62	-	-	0.03	-	-	0.12	-	-	0.46	-	-	1.23	-	-
ACTIVSg500	-	45	-	-	145	-	-	0.05	-	-	0.18	-	-	0.54	-	-	1.32	-
ACTIVSg2k	63	-	91	140	-	324	0.30	-	0.31	0.77	-	1.92	1.33	-	1.00	2.87	-	1.80
ACTIVSg10k	41	44	45	93	156	245	0.08	0.06	0.06	0.31	0.29	0.42	0.45	0.34	0.40	2.18	1.34	1.18
ACTIVSg25k	31	52	34	95	202	229	0.03	0.14	0.02	0.10	0.78	0.20	0.24	1.14	0.13	1.16	1.72	0.92
WECC	26	20	30	96	186	530	0.05	0.06	0.14	0.27	0.66	1.17	0.50	1.27	1.61	2.01	2.42	3.44
EI	31	48	53	120	239	401	0.07	0.09	0.12	0.31	0.79	1.07	0.86	1.05	1.13	2.13	2.49	2.98

1.6.2 Key characteristic features of the power system networks

Here we discuss a list of key characteristic metrics of power systems that are unique and consistent for every power system network. They are found by analyzing a large set of different power system networks. These key features exhibit distinctive statistical patterns and scaling properties which can be used to benchmark synthetic grids against actual grids.

Average node degree: As shown in Table XVIII, the average node degree is independent of the network size, and range from 2.5 to 3.2 for most real networks. This is a special characteristic topological parameter for power system networks and what makes them different from *SmallWorld* networks. Figure 51 shows the average node degree for real and synthetic grids which indicates this parameter is independent of network size.

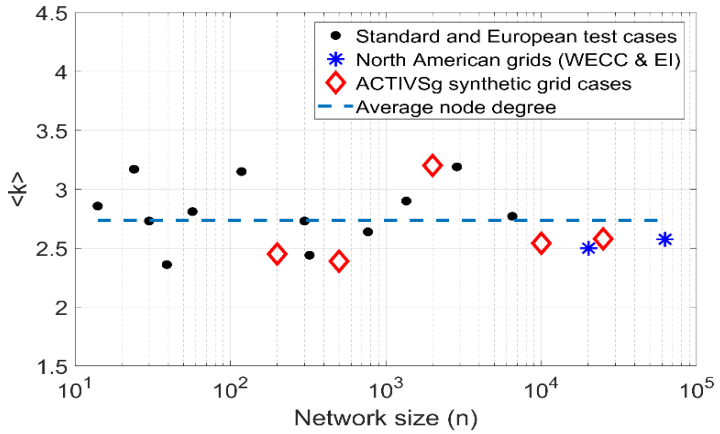


Figure 51. Average node degree for real and synthetic grids

The relative distance of correlated bus type assignment: The relative distance of bus type entropy in correlated assignments and randomized permutation for power grid networks exhibit a strong dependence on the network size that can be mathematically modeled. Figure 52 shows the distance for actual grids and synthetic grids. The approximate scaling function is derived as [139]:

$$d_W(n) = \begin{cases} -1.39 \ln n + 6.79, & \ln n \le 8 \\ -1.25 \times 10^{-13} (\ln n)^{15.1} + 0.43, & \ln n > 8 \end{cases} \quad (51)$$

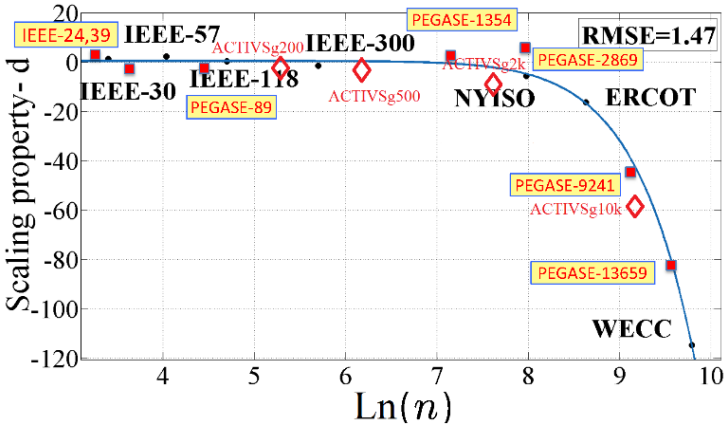


Figure 52. Scaling function of the relative distance of bus type entropy in random assignment for different network sizes

Transformer per unit reactance: The transformer per unit reactance calculated based on its own MVA rating falls within a narrow range of [0, 0.25] p.u with 80% of them inside [0.05, 0.2] range.

The average transformer p.u. reactance is independent of the voltage level and network size as shown in Figure 53 for North American networks and ACTIVSg cases.

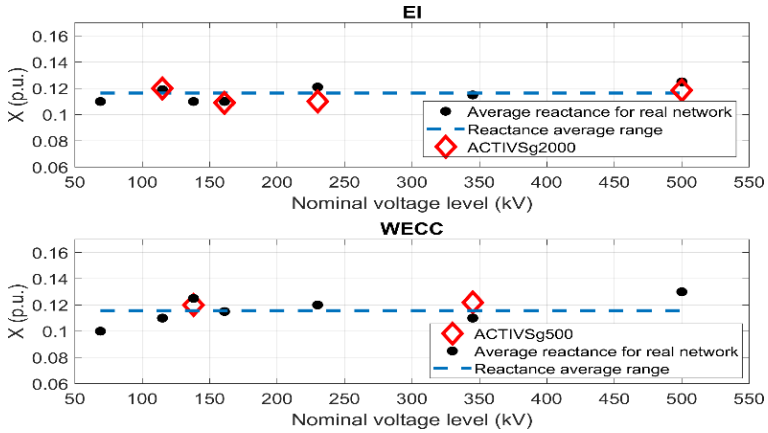


Figure 53. Average transformer p.u. reactance in its own MVA for EI, WECC, ACTIVSg500, and ACTIVSg2000

Transmission line X/R ratio: Our statistical analyses on a set of real-world power system data indicate that the ratio of transmission line reactance over its resistance (X/R) shows a heavy dependency on the nominal voltage level which can be estimated by a linear relationship. This is also consistent with engineering practice where higher voltage networks are designed to have lower resistance to minimize active power loss.

The above list provides a set of critical parameters and metrics that exhibit unique characteristics of power system networks. Since the observed statistical ranges and patterns are consistent for many real-world and standard power system cases, therefore, this list provides a standard set of metrics to evaluate and validate the synthetic grids. In other words, if any artificial power system network does not meet any of these metrics, it lacks the minimum requirements to be considered a realistic test case.

Appendix B: Vita

Mir Hadi Athari

(804) 269-6059 | atharih@vcu.edu | LinkedIn

EDUCATION

Virginia Commonwealth University, Richmond, VA, USA, fall 2015 – spring 2019
Ph.D. Candidate, Electrical Engineering (Power systems)
Research area: Power system vulnerability analysis, renewable energy integration, power system modeling
Advisor: Dr. Zhifang Wang
GPA*:* 4.0

Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, 2014
M.Sc. in Electrical Engineering (Energy management)
Thesis: Optimal Fuzzy Control of On-grid Hybrid Renewable Energy System
Advisor: Dr. Morteza Mohammadi Ardehali
GPA: 3.91

University of Tabriz, Tabriz, Iran, 2012
B.Sc. in Electrical Engineering (majoring in power systems)
Thesis: Installation and implementation of Schering Bridge in High Voltage Laboratory of University of Tabriz
Advisor: Dr. Heresh Seyyedi

GPA: 3.60

RESEARCH INTERESTS

- Power system vulnerability analysis
- Power system operation and control
- Integration of renewable energy resources
- Transmission Planning
- Application of intelligent methods in power systems
- Statistical modeling of renewable generation and smart grid loads

WORK EXPERIENCE

- **Graduate Teaching Assistant**, Virginia Commonwealth University, Richmond, VA ,Aug. 2018 - Present
- **Graduate Intern at Transmission Planning**, Dominion Energy, Richmond, VA, May 2018 - Aug. 2018
- **Graduate Research Assistant**, Virginia Commonwealth University, Richmond, VA, Aug. 2015 - Present
- **Academic Projects Consultant**, Amirkabir University of Technology, Tehran, Iran, Sep. 2012 - Jun. 2015
- **Graduate Research Assistant**, Amirkabir University of Technology, Tehran, Iran, Sep. 2012 - Aug. 2014
- **Intern**, Tabriz Electricity Distribution Company, Tabriz, Iran, Jul. 2012. - Sep. 2012

PROJECTS

- **Node-Breaker modeling in transmission planning**, summer 2018 (*Sponsored by Dominion Energy*)
  (Developing a fully automated process to implement Node-Breaker substation model in transmission planning cases)
- **Validation metrics for synthetic power system modeling**, fall 2016 (*Sponsored by U.S. Department of Energy*).
  (A comprehensive statistical study on real-world power system data to extract validation metrics for artificial power

systems, ongoing project).

- **PVDG impacts on distribution network**, summer 2016 (*Sponsored by U.S. Department of Energy*).
  (Simulation of high penetration of Photovoltaic Distributed Generation (PVDG) units to examine their impacts on distribution network).
- **Impacts of renewable generation on grid vulnerability**, fall 2015 (*Sponsored by U.S. Department of Energy*).
  (Simulation of cascading failures in transmission networks with high penetration of renewable energy sources to examine their adverse impacts on grid vulnerability to cascading failure, ongoing project).
- **Optimal operation of storage unit for hybrid renewable energy system**, fall 2013.
  (Optimal design of a Fuzzy controller for storage system management in a hybrid renewable energy system).
- **Energy auditing and performance evaluation of a cement factory**, winter 2012.
  (Auditing energy process of a cement factory to provide recommendations for better efficiency).

PROFESSIONAL/EXTRACURRICULAR ACTIVITIES

- Vice President, VCU Engineering Graduate Student Association (EGSA), Apr. 2018 – Present.
- Treasurer, VCU Engineering Graduate Student Association (EGSA), Apr. 2017 – Apr. 2018.
- Mentor, VCU Dean's Early Research Initiative (DERI) program, Jul. 2016 – Jun. 2017.
- Webmaster, EPES lab website, Aug. 2016 – Mar. 2018.
- EPES lab tour guide for high school students, Sep. 2015 – Mar. 2018.
- Reviewer for journals and IEEE sponsored conferences
- IEEE Power and Energy Society (PES) student membership, since Sep. 2014.

Professional talks:

- Paper presentation at CIGRE Grid of the Future (GOTF) conference, Oct. 2018.
- Paper presentation at 49th North American Power Symposium (NAPS2017), Sep. 2017.
- Poster presentation at 2017 IEEE PES General Meeting (PESGM2017), Jul. 2017.
- Invited talk in VCU department of Electrical Engineering Graduate Seminar, Feb. 2017.
- Paper presentation at 2016 Innovative Smart Grid Technologies (ISGT2016), Sep. 2016.
- Invited research presentation for Virginia Dominion Power experts, Apr. 2016.

HONORS & AWARDS

- VCU ECE department Outstanding Graduate Research Assistant, Jun. 2017.
- VCU Dean's Early Research Initiative (DERI) program award, Sep. 2016.
- VCU Graduate School travel award, Sep. 2016 & Jun. 2017.
- VCU ECE department travel award, Sep. 2016 & Jun. 2017.
- Ranked 1st in 2014 class of Energy Management students, Amirkabir University of Technology, Apr. 2014.
- Ranked 142nd in Iran's Nationwide University Entrance Exam among more than 230,000 participants, Sep. 2012.

PUBLICATIONS

Journal Publications:

- **M.H. Athari** and Z. Wang, "Stochastic AC Cascading Failure Model under Uncertain Generation using Unscented Transform," *IEEE Tran. On Sustainable Energy*, Oct 2018 (passed 1st round of review).
- **M.H. Athari** and Z. Wang, "Introducing Voltage-dependent Parameters to Synthetic Grids Electrical Topology," *IEEE Tran. On Smart Grids*, Jun. 2018, pp. 1–1.

- A.B. Birchfield, E. Schweitzer, **M.H. Athari**, T. Xu, T.J. Overbye, A. Scaglione, and Z. Wang, "A Metric-Based Validation Process to Assess the Realism of Synthetic Power Grids," *Energies*, Aug. 2017, vol: 10(8), pp:1233.
- **M.H. Athari** and Z. Wang, "Impacts of Wind Power Uncertainty on Grid Vulnerability to Cascading Overload Failures," *IEEE Tran. On Sustainable Energy*, Jan. 2018, vol. 9, no. 1, pp. 128–137.
- **M.H. Athari** and M.M. Ardehali, "Operational performance of energy storage as function of electricity prices for on-grid hybrid renewable energy system by optimized fuzzy logic controller," *Renewable Energy, Elsevier,* Jan. 2016, 85:892-902.

Peer-reviewed Conference Papers:

- **M.H. Athari**, K.D. Hannah, and Z. Wang, "Automated Enhancement of Transmission Planning Bus-Branch Model with EMS Node-Breaker Substation Models," *2018 CIGRE Grid of the Future (GOTF)*, Oct. 2018 (accepted, to appear).
- **M.H. Athari** and Z. Wang, "Grid Vulnerability Analysis based on Probabilistic Cascading Failure Model and Wind Power Installation," *52nd Hawaii International Conference on System Sciences (HICSS)*, Jul. 2018 (accepted, to appear).
- Z. Wang and **M.H. Athari**, "Statistically Analyzing Power System Network," *IEEE Power & Energy Society General Meeting, PESGM2018,* Aug. 2018, Portland, OR (accepted, to appear).
- **M.H. Athari** and Z. Wang, "Interdependence of Transmission Branch Parameters on the Voltage Levels," *51st Hawaii International Conference on System Sciences (HICSS),* Jan. 2018, Waikoloa, HI.
- **M.H. Athari**, C. Yang, and Z. Wang, "Sequential Optimal Placement of Distributed Photovoltaics using Downstream Power Index," *49th North American Power Symposium (NAPS),* Sep. 2017, Morgantown, WV.
- **M.H. Athari** and Z. Wang, "Statistically Characterizing the Electrical Parameters of the Grid Transformers and Transmission Lines," *10th Bulk Power Systems Dynamics and Control Symposium, IREP2017,* Sep. 2017, Espinho, Portugal.
- **M.H. Athari** and Z. Wang, "Studying Cascading Overload Failures under High Penetration of Wind Generation," *IEEE Power & Energy Society General Meeting, PESGM2017*, Jul. 2017, Chicago, IL.
- H. Sadeghian, **M.H. Athari**, and Z. Wang, "Optimized Solar Photovoltaic Generation in a Real Local Distribution Network," *IEEE Innovative Smart Grid Technologies, ISGT2017*, Apr. 2017, Arlington, VA.
- **M.H. Athari** and Z. Wang, "Time-Series Analysis of Photovoltaic Distributed Generation Impacts on a Local Distributed Network," *IEEE PowerTech2017*, Jun. 2017, Manchester, UK.
- **M.H. Athari** and Z. Wang, "Modeling the Uncertainties in Renewable Generation and Smart Grid Loads for the Study of the Grid Vulnerability," *IEEE Innovative Smart Grid Technologies (ISGT)*, Sep. 2016, Minneapolis, MN.
- **M.H. Athari**, G.B. Gharehpetian, and H. Sadeghian, "Optimized Fuzzy Controller for Charging Algorithms of Plug-in Hybrid Electric Vehicles" *23rd Iranian Conference on Electrical Engineering (ICEE 2015), Sharif University of Technology*, Tehran, Iran, Feb. 2015.
- H. Sadeghian, G.B. Gharehpetian, and **M.H. Athari** "Improved Multi-agent System for Intelligent Energy Management of Microgrids in Presence of PHEVs," *23rd Iranian Conference on Electrical Engineering (ICEE 2015), Sharif University of Technology, Tehran, Iran*, Feb. 2015.
- **M.H. Athari** and M.M. Ardehali, "Performance Evaluation of a Renewable Energy System in Grid-connected Mode and Optimal FLC Using Particle Swarm Optimization Algorithm," *22nd Iranian Conference on Electrical Engineering (ICEE 2014), Shahid Beheshti University, Tehran, Iran,*, Mar. 2014.

SKILLS

- **Computer:** Programming Languages: MATLAB, C++, Python
  Engineering Software Expertise: PSS/E, PSCAD/RSCAD, PowerWorld, GAMS, MATLAB Simulink, MATLAB Control and Optimization Toolboxes
  Text editors: MS Word, Latex
  General software: MS Excel, Photoshop
- **Data analysis and interpretation**

- **Machine learning and deep learning**

RELEVANT COURSEWORK

- Power system analysis
- Energy storage
- Substation design
- Power plant
- Sustainable and eff. energy systems
- Power system operation and control
- Economy and energy management
- Power system protection and relays
- Renewable energy resources
- Micro-grids and smart grids
- VAR control in power systems
- Electrical machines

LINKS

- Google Scholar
- Researchgate
- LinkedIn