DEVELOPMENT OF A RISK-INFORMED DECISION-MAKING CAPABILITY USING STANDARD ELECTRIC POWER INDUSTRY PLANNING TOOLS

PREPARED FOR THE U.S. DOE

OFFICE OF ELECTRICITY

DOE PROGRAM MANAGER: DR. ALIREZA GHASSEMIAN

PREPARED BY:

MENG YUE AND JUNPENG ZHAN

JULY 2019

TABLE OF CONTENTS

1. Introduction 1
2. Objective and Methodology 3
2.1 A Formal Statistical Test for Data Poolability 3
2.2 Modeling and Parameterization of Intermittency Induced Outages for Renewable Generation 3
2.3 Enhancing Probabilistic Contingency Analysis Capability of Existing Tools 6
2.4 Calculation of Probabilistic Reliability Indices 7
2.5 A Decision-making Process Using Probabilistic Reliability Metrics 8
3. Reliability Data Collection, Analysis, and Repository Development 10
3.1 Outage Parameters of Conventional Grid Components for PCA 10
3.1.1 Nomenclatures and Input Parameters for PCA 10
3.1.2 Raw Data Collection of Grid Component Outages 11
3.1.3 Determination of Data Poolability and Modeling Population Variability 16
3.1.3.1 A Formal Statistical Test 16
3.1.3.2 A Lognormal Distribution for PCA Parameters 17
3.2 An Investigation of Potential Intermittency Induced Outage Modes for Wind Generation 19
3.2.1 Outage Modes of Renewable Generation 20
3.2.2 Probabilistic Outage Models and Parameterization of Renewable Resources 20
3.2.2.1 Pearson' Correlation for Probabilistic Outage Models 20
3.2.2.2 Historical Wind Generation Data Cleaning 21
3.2.2.3 Extraction of Ramp-up and -down Ramping Events from Time Series Wind Generation Data 24
3.2.3 Parameters for Outage Models of Wind Generation Sites 25
3.2.3.1 Analyses of All Individual Rampings in Wind Generation 26
3.2.3.2 Analyses of Single Rampings in Wind Generation 27
3.2.3.3 Parameters for Common Mode Outage Models of Multiple Wind Generation Sites 35
4. Facilitation of Decision-making Process in Transmission Planning Using Probabilistic Reliability Metrics 45
4.1 Probabilistic Planning Criteria 45
4.1.1 Reliability Target 46
4.1.2 Probabilistic Cost 46
4.1.2.1 Total Cost Method 46
4.1.2.2 Benefit/Cost Ratio Method 48
4.1.2.3 Incremental Reliability Index 48

4.2 Combining Probabilistic and Deterministic Planning Criteria: The Well-being Approach 49
5. The Enhanced PSS/E Driven by Python ... 52
5.1 Python Code Architecture ... 52
5.2 Contingency Screening ... 53
5.3 IIO/CMO Processor ... 53
5.4 Sample Generator ... 54
5.5 Output Module ... 55
5.6 Flowchart of Python Code ... 55
6. Case Study ... 57
6.1 A Case Study Using a 23-bus System ... 57
6.2 A Case Study Using the WECC System ... 59
6.2.1 WECC System ... 59
6.2.2 Parameter Settings for Enhanced PCA Studies ... 60
6.2.3 Comparison between built-in PCA and enhanced PCA with and without IIOs for Wind Generation ... 60
6.2.4 Impacts on System Reliability of Different Wind Penetrations Considering IIOs/CMOs ... 66
6.2.5 Implementation of Well-being Approach and Case Study ... 67
6.2.6 Parallel vs. Sequential Execution of Contingency Analysis ... 68
7. Conclusions and Future Work ... 70
8. References ... 73
9. Appendix ... 75

List of Figures

Figure 2. 1: Example fast ramp-ups and ramp-downs in wind generation.................................. 4
Figure 2. 2: Example concurrent rampings of highly-correlated generation sites........................ 5
Figure 2. 3: A Scheme for Enhancing PCA Capability.. 7
Figure 3. 1: Illustration of continuously unchanged zero and non-zero data 22
Figure 3. 2: Scatter plot of wind power vs. wind speed for site 10. 22
Figure 3. 3: Scatter plot of wind power vs. wind speed for site 25. 23
Figure 3. 4: Scatter plot of wind power vs. wind speed for site 71. 23
Figure 3. 5: Example under- and over generation modes of a single generation site.................25
Figure 3. 6: Frequency distributions for individual rampings...26
Figure 3. 7: Frequency distributions for single rampings. 27
Figure 3. 8: Scatter diagram for initial generation level and deviation.......................................28
Figure 3. 9: Scatter diagram for initial generation level and deviation of single ramp-ups..........28
Figure 3. 10: Scatter diagram for initial generation level and deviation of single ramp-downs. ..29
Figure 3. 11: Distributions of initial generation levels $Geninit$ for single rampings. 30
Figure 3. 12: Distributions of deviation dev for single rampings...30
Figure 3. 13: Distributions of duration $tduration$ for single rampings. 31
Figure 3. 14: Frequency distributions for single rampings. 31
Figure 3. 15: Scatter diagram for initial generation level and deviation......................................32
Figure 3. 16: Scatter diagram for initial generation level and deviation of single ramp-up rampings...32
Figure 3. 17: Scatter diagram for initial generation level and deviation of single ramp-down rampings 2...33
Figure 3. 18: Distributions of initial generation levels $Geninit$ for single rampings. 34
Figure 3. 19: Distributions of deviation dev for single rampings...34
Figure 3. 20: Distributions of duration $tduration$ for single rampings. 35
Figure 3. 21: Concurrent under-generation modes among two and three generation sites........36
Figure 3. 22: Frequency distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0])..37
Figure 3. 23: Initial Generation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0])..38
Figure 3. 24: Generation deviation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0])...38
Figure 3. 25: Duration distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0])..38
Figure 3. 26: Frequency distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8])..39
Figure 3. 27: Initial generation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8])..39
Figure 3. 28: Generation deviation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8])...39
Figure 3. 29: Duration distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8])..40
Figure 3. 30: Frequency distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right). 41
Figure 3. 31: Initial Generation distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right)...41
Figure 3. 32: Deviation distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right). 41

Figure 3. 33: Duration distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right). 42
Figure 3. 34: Frequency distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right). 42
Figure 3. 35: Initial generation distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right)....42
Figure 3. 36: Deviation distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right). 43
Figure 3. 37: Deviation distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right). 43
Figure 4. 1: A Total cost method. 47
Figure 4. 2: System States for Grid Operation. 49
Figure 5. 1: Flowchart of the Python tool 52
Figure 5. 2: Contingency DBL_135_4710....53
Figure 5. 3: Illustration of a bin for the probability of a deviation level. 55
Figure 6. 1: Histogram and pdf of system problem frequencies. 58
Figure 6. 2: Histogram and pdf of system problem durations....58
Figure 6. 3: Histogram and pdf of system overloading frequencies. 59
Figure 6. 4: Histogram/distribution of overvoltage contingencies in Case 4: (a) frequency and (b) duration. 63
Figure 6. 5: Histogram/distribution of undervoltage contingencies in Case 4: (a) frequency and (b) duration. 63
Figure 6. 6: Histogram/distribution of system problem contingencies in Case 4: (a) frequency and (b) duration....63
Figure 6. 7: Histogram/distribution of overvoltage contingencies in Case 5: (a) frequency and (b) duration. 64
Figure 6. 8: Histogram/distribution of undervoltage contingencies in Case 5: (a) frequency and (b) duration. 64
Figure 6. 9: Histogram/distribution of system problem contingencies in Case 5: (a) frequency and (b) duration....64
Figure 6. 10: Histogram/distribution of overvoltage contingencies in Case 6: (a) frequency and (b) duration. 65
Figure 6. 11: Histogram/distribution of undervoltage contingencies in Case 6: (a) frequency and (b) duration. 65
Figure 6. 12: Histogram/distribution of system problem contingencies in Case 6: (a) frequency and (b) duration....65
Figure 6. 13: Schematic of N-1, N-2, N-3 contingencies and well-being 68
Figure A. 1: A snapshot of the Outages Statistics Data file (*.prb). 75
Figure A. 2: A snapshot of the Monitored Element Data file (*.mon). 76
Figure A. 3: A snapshot of the Contingency Description Data file (*.con). 76
Figure A. 4: A snapshot of the Contingency Description Data file (*_out.con). 77
Figure A. 5: A snapshot of the *.iio file. 78
Figure A. 6: A snapshot of the *.cmf2 file. 79
Figure A. 7: A snapshot of the *.cmf3 file. 80

List of Tables

Table E. 1: Mean Values of System Problem Frequency and Duration v
Table E. 2: Frequency, duration, and probability of the overvoltage, undervoltage, and system problems for Cases 3 - 6. v
Table E. 3: EENS for Cases 3 - 6. vi
Table 2. 1: Top Three 1-hour Changes in Wind Generation Output in 2017 (Source: ERCOT) .. 6
Table 3. 1: Raw Outage Data for AC Circuits. 13
Table 3. 2: Raw Outage Data for Transformers.. 15
Table 3. 3: Raw Outage Data for Conventional Generators.. 15
Table 3. 4: Parameters and Statistics for Lognormal Distributions of Different Outages 18
Table 3. 5: Mean Values for Poolable Outage Data 19
Table 3. 6: Correlation of $Geninit$, dev, and $tduration$. 29
Table 3. 7: Correlation of $Geninit$, dev, and $tduration$. 33
Table 3. 8: Correlation vs. concurrent modes between two wind generation sites (δ=30%)....... 37
Table 3. 9: Correlation vs. concurrent modes between two wind generation sites (δ=50%)....... 37
Table 3. 10: Correlation vs. concurrent modes between three wind generation sites (δ=30%) .. 40
Table 3. 11: Mean Values of Frequencies, initial generation, deviation, and durations (δ=0.5).. 43
Table 6. 1: Mean Values of System Problem Frequency and Duration.................................. 57
Table 6. 2. Parameter settings for Enhanced PCA 60
Table 6. 3: Number of single and double contingencies for Cases 3, 4, 5, and 6....................... 62
Table 6. 4: Frequency, duration, and probability of overvoltage, undervoltage, and system problems for Cases 3, 4, 5, and 6. 62
Table 6. 5: EENS for Cases 3, 4, 5, and 6.. 62
Table 6. 6: Number of single and double contingencies for Cases 3v2 - 6v2............................ 66
Table 6. 7: EENS for Cases 3v2 – 6v2. 66
Table 6. 8. Frequencies, durations, and probabilities of the overvoltage, undervoltage, and system problems for with different wind penetrations. 67
Table 6. 9: Frequency, duration, and probability of the system being in a marginal state with overvoltage, undervoltage, and system problems for Case 3. 67
Table 6. 10: Contingency Analysis Computational Time (Parallel vs. Sequential) 68

Executive Summary

Objectives

The objectives of this study are to address the following issues associated with current contingency analyses used for transmission system planning in the power grid:

- Adequacy of reliability data used for conventional grid component outages and emerging renewable related outage modes,
- Methods of including renewables in contingency analyses that are beyond the capability of existing tools in probabilistic contingency analysis (PCA), and
- Demonstration of the usefulness of probabilistic reliability metrics and the complementary enhancement of the deterministic counterpart considering renewable generation.

Background

Contingency Analysis (CA) is a major technique used in transmission planning to meet criteria mandated by NERC and FERC. Traditional CA calculates the potential for experiencing system problems, such as a loss of load, overloading, or voltage violations, due to equipment failure based on the existing system configuration and projected load profiles. At present, utilities or system operators need to follow NERC's mandated planning criteria, which are essentially deterministic. The potential for system problems is directly related to system reliability; therefore, CA results can produce a set of metrics to measure reliability in transmission planning. These metrics can also be used by utilities to develop a mitigation plan and determine remedial actions for reliability improvement, as needed. The traditional CA framework solely relies on a deterministic contingency analysis (DCA) approach without differentiating the likelihood of occurrences among different types of contingencies. The implied assumption of equally likely contingencies in DCA can lead to over-conservatism in decision-making for planning. In particular, it is very difficult to treat variable generation from renewables using the DCA framework. Therefore, DCA lacks the sophistication needed to deal with the greater uncertainty introduced by renewable generation, especially with high penetration levels. A better alternative is probabilistic CA (PCA).

Probabilistic planning, including PCA, has been in use for several decades. However, it has never been widely adopted in planning practices by utilities. There certainly exist several barriers to the successful implementation of probabilistic planning, which may include, but are not limited to availability of quality data and capable tools, and how to use the outcome of PCA for decision-making. Data collection and accumulation require following appropriate procedures and can be very time consuming. At the transmission level, the transmission data availability system (TADS) and generating availability system (GADS) that were developed and have been continuously supported by NERC are valuable reliability data sources for probabilistic planning, although the reliability data are generally provided in an aggregated form without many details.

Regarding the availability of probabilistic tools, the electric power industry has "standardized" on a relatively few software tools for planning and operation. Around these few *de facto* standardized tools, the industry has built a resource base of skilled personnel, required data, required data handling procedures, and required analysis capability. There is high confidence in this resource base and it represents a huge investment. It is foreseen that the industry will continue to use these tools until the next generation of grid analytical tools becomes available, mature and accepted by the industry. In the interim, there is a need to create a transparent means to make the transition to probabilistic planning techniques. One way is to integrate probabilistic risk assessment (PRA) capabilities into existing software tools or enhance the probabilistic capabilities already included in these tools, as needed. The key aspect is that there be no changes to the procedures the highly skilled planners currently use to make use of the tools. Decision-makers using the information provided by the tools will have the exact same information, but it will now include PRA defined results. With input from industry stakeholders, external add-on modules can be developed and integrated into these *de facto* tools. The enhanced tools will bring about speed and sophisticated analysis that will deal with uncertainty in a rigorous process.

There is also a need to develop new methods to better model uncertainties, especially those associated with renewables in probabilistic planning. The DCA or existing PCA focus on conventional grid components or equipment, such as conventional generators and transmission circuits. Other types of non-conventional uncertainties that also affect the contingency analyses, especially those associated with renewable generation from solar and/or wind in the grid, are either not included or they are included deterministically. The relevant uncertainties with renewables include not only the randomness of component failures, similar to conventional contingencies, but also uncertainties due to the uncontrollable variability of the generation. These uncertainties, especially the uncertainties due to generation intermittencies, will increase very quickly as increasing amounts of renewables are installed on the grid. Therefore, the contingencies associated with the renewable sources will very likely be a major concern in the near future and will have to be considered together with other contingencies in the evaluation of their impacts on reliability metrics in transmission planning.

Finally, it is critical to demonstrate the usefulness of the probabilistic metrics in the decision-making process, which has been a major barrier to the adoption of probabilistic planning by utilities. These issues are addressed in this study based on the technical approaches discussed below.

Technical Approaches

A Formal Statistical Test for Data Poolability

It is well-known that different types of components, as well as different sub-types of a particular type of components have different failure or outage rates. Outage frequencies of grid components (e.g., transmission lines) at different voltage levels vary and, accordingly, failure records from different sources can be combined to increase the population size. It is very common that the outage data of grid components that are relatively homogenous (e.g., 200–399kV lines) is aggregated to generate a constant outage frequency [1 - 6] and outage duration. These are the standard input parameters for performing PCA using tools built into existing grid simulation software programs, such as PSS/E and TRANSCARE [7].

The average values are the maximum likelihood estimates (MLE) of the outage frequencies or durations, assuming the same frequency or duration across different data sources for a Poisson process, describing outage occurrences [8]. The issue with using a constant outage frequency for a category of grid components (e.g., 200 – 399kV transmission circuits) is that it neglects the variability of outage occurrences among utilities or transmission owners (TOs). The factors that may have significant impacts on the outage frequency or the duration, i.e., the outage variability, may include environmental conditions and maintenance schedules that differ within utilities, which means outage frequencies or durations for the same category of components may be different from utility-to-utility, or from region-to-region. It is thus desirable to capture such variability in probabilistic analyses [8]. This is the so-called “data poolability” issue and was addressed by using a formal statistical test process to determine whether there is a need to model the population variation of data for a specific class of grid components [8].

Modeling and Parameterization of Intermittency Induced Outages for Renewable Generation

Existing contingency analysis focuses on contingencies or outages of conventional grid components or equipment without excluding renewable generators. A renewable generator such as a wind farm, however, is always modeled simply as a conventional generator with a pre-specified output that is determined by utilities from, e.g., a seasonal average generation. Following this practice, the contingencies of renewable generators are characterized by random failures of wind turbines, e.g., failure of mechanical parts, converters, controls, the collector networks, or other parts of the turbines. Although the uncertainties of such failures can be captured in PCA using outage frequencies and durations similar to a conventional generator, other types of uncertainties or non-conventional uncertainties that also affect the contingency analyses, especially those associated with intermittent renewable generation from solar and/or wind in the grid, are not included.

The major difference between conventional generator outages and renewable outages is that different outage modes for renewables must be considered and modeled, i.e., in addition to a complete loss of generation[1], under- or over-generation (relative to the generation levels prior to the fast ramping events) of renewable generators also must be explicitly modeled. This is because the causes of these outage modes may be very different.

The generation of a wind or solar farm can change dramatically within a short time period in terms of tens of minutes or even minutes, e.g., a loss of generation due to the shut-down of wind turbines when the wind speed exceeds the cut-out speed or reduced generation levels due to sudden changes in wind speed or direction. For the purpose of studying contingency analysis, sudden changes in generation levels are of interest. Such changes may be due to generator failures but may also be caused fast ramping events in wind farms. It is recognized that ramping events are not finished instantaneously but this may also be true for a loss of generation or de-rated generation of conventional generators. Upon the failure of a conventional generator, its generation level does not necessarily drop to zero instantaneously. Instead, the generation is likely to coast down to zero due to the inertia of the generator. Usually, a conventional generator failure may take a long time to repair [9]. However, after a ramping the wind generation may (or may not) recover to the pre-ramping level in a short time period. It should be noted that a sustained outage for a transmission circuit is defined as “an automatic outage that lasts for more than one minute”[2] in [10]. This indicates that any outages with durations longer than one minute can adversely impact system reliability in the current practices of PCA. As shown in this study, the duration of maintaining at an increased or reduced generation level after the ramp-up or -down event is usually short but may also last for hours. Therefore, fast wind ramping events can be considered as valid outages, i.e., intermittency induced outages (IIOs). These uncertainties due to generation intermittency will increase much more quickly as increasing amounts of renewables are installed. Therefore, the IIOs associated with renewable sources will very likely be an increasing concern in the near-term, and a major concern in the long-term. It is clear they must be considered together with other outages in the contingency analysis, which means that the probability of occurrence of different outage modes associated with these IIOs needs to be accounted for based on the cause of the outage modes in the PCA.

For conventional generators, dependent or common mode outages (CMOs) or failures are very common for co-located generation units. It is known that wind farms may be built along wind corridors, where wind is available all year long. Therefore, a concurrent change in wind generation can happen for adjacent wind farms and even for wind farms that are not physically very close to each other. As indicated in [11], the correlation[3] decreases with distance but can still exist and be relatively strong for wind farms that are 1,000 km away from each other. For under- or over-generation modes caused by sudden changes of wind speed, CMOs of multiple renewable sources are possible due to high geospatial correlation of wind farms that are close to each other. This poses a more severe threat to the grid and needs to be captured.

In summary, the major differences between conventional generator outages and renewable outages are [8]:

- In addition to a complete loss of generation, clearly under- or over-generation of renewable generators due to ramp-downs and -ups also must be explicitly modeled.
- For each of the outage modes, there can be associated common mode outages (CMOs) for different generation sites due to the reliance of generation on environmental conditions.

Probabilistic models and their parameterization for such contingencies need to be developed to provide input data for frequencies and duration times of different outage modes in the PCA framework. For this purpose, historical solar or wind generation data can be collected to analyze the statistics of these outage modes, e.g., how often the reduction in wind generation at a certain level occurs and how long this reduction lasts. This method is similar to the calculation of frequencies and probabilities of occurrences of different fast solar rampings using the solar irradiance data and electrical models of solar plants, as shown in [12]. In [12], fast solar rampings were captured by analyzing high-resolution irradiance measurement data and characterized in terms of how much the irradiance changes within a certain time period, which was used

[1] A conventional generator may also fail partially due to de-rated generation. However, data for such partial failures or outages are rare and usually unavailable.
[2] There is no such definition for generator outages.
[3] Correlations can be mathematically quantified using, e.g., Pearson’s correlation coefficient, from historical time series wind generation data.

for calculating the statistics of solar generation reduction. Such probabilistic models, once developed for wind or solar generators, are readily used as input and implemented in the developed tool for the enhanced PCA analysis.

A Scheme for Enhancing Probabilistic Contingency Analysis Capability of Existing Tools

As indicated above, existing tools that are capable of performing PCA, such as PSS/E and TRANSCARE, use input data comprised of averaged outage parameters only, i.e., outage data of grid components that are relatively homogenous are aggregated and averaged. If outage data from both conventional grid components and intermittencies of renewable generation cannot be aggregated and averaged and must be described by distributions (which are actually true for most of the grid components, as will be shown later), then the existing tools are incapable of calculating the probabilistic reliability indices and need to be enhanced. The mean values of the probabilistic reliability indices should not be calculated using the average or mean values of outage parameters for grid components. Based on the distributions of different input parameters, a generic scheme for improving the PCA capability of current contingency analysis tools was developed and is described in Chapter 2. The scheme uses a Monte Carlo simulation to calculate probabilistic indices based on a deterministic CA, which not only enables the retrieval of mean values but also other statistical properties of the reliability indices. The mean values calculated this way are the true means.

Facilitation of the Decision-making Process in Transmission Planning Using Probabilistic Reliability Metrics

Probabilistic planning can rigorously model and, thus, capture uncertainties. However, the major issue with its acceptance is that it is difficult for utilities to use probabilistic results in the decision-making process. Unlike the deterministic process that focuses on compliance with well-established transmission planning standards, there is a lack of probabilistic planning criteria for the transmission planner to adopt.

To facilitate the decision-making process using probabilistic metrics, probabilistic planning criteria are reviewed and illustrated herein to show how to use them together with deterministic criteria to facilitate and help utilities make planning decisions. The focus is on the discussion of probabilistic reliability metrics calculated from contingency analysis. In addition, a system “well-being” approach is discussed and implemented in the enhanced PCA tool developed in this study. The well-being analysis not only generates the probabilities of at-risk states but also that of healthy states.

Major Results

A significant amount of outage data has been collected from various data sources, including NERC (North American Electric Reliability Corporation), TADS (Transmission Availability Data System), and GADS (Generation Availability Data System) since they contain the outage data from different transmission owners or NERC regions. Some other publicly available data sources including some data from Canada were also explored and used in this study. The lumped outage data in [1 - 6] were extracted and treated as an independent data source. A formal statistical test was applied to the data from these sources to determine poolability. For those data that should not be pooled, lognormal distributions were developed to represent their variability across regions. The raw data and parameters for distributions are given in various tables in Chapter 3. The outage statistics in this data repository can be used by planners or researchers directly, while raw data are also available for users to develop their own outage data.

Real wind generation data with five-minute intervals were acquired and analyzed. Results of the generation variation analysis show that wind power from different wind farms can change dramatically within tens of minutes. Such dramatic changes, especially for concurrent changes in multiple wind farms that are highly correlated, may pose serious challenges to the system and can be treated as intermittency induced outages (IIOs) or contingencies. The issue will be more pronounced for very high penetration levels of renewable generation and there is a need to include these into the contingency analysis. To incorporate IIOs into the PCA, fast ramping events were extracted from the data and detailed statistical analyses were performed to extract the information needed. Special attention was paid to the concurrent IIOs (or common mode outages or CMOs) at multiple wind farms that are highly correlated. Statistics for the probabilistic models for IIOs and CMOs are provided in Chapter 4.

A number of case studies were performed based on the outage statistics developed in this study using both a small sample system (23-bus) and WECC (Western Electricity Coordinating Council) system. The performance of existing built-in PCA capabilities (using arithmetic mean frequencies and durations as input) and the enhanced PCA (using Monte Carlo simulation by sampling distributions of grid component outage frequencies and durations) developed in this study were compared in Cases 1 to 4. Cases 1 and 2 were based on the 23-bus system, while Cases 3 and 4 are based on the WECC system. The statistics of the system problem frequencies and durations for Cases 1 and 2 are shown in Table E.1.

Table E. 1: ***Mean Values of System Problem Frequency and Duration***

	Case 1: Built-in PCA	**Case 2: Enhanced PCA**
System Problem Frequency (per Year)	9.27	17.64
System Problem Duration (Hours)	47.1	34.15

1.

For Case 1, the mean values of the frequencies and duration were determined, i.e., 9.27 problem occurrences per year and a duration of 47.1 hours per system problem. Compared to the mean values calculated in Case 2, i.e., 17.64 occurrences per year and a duration of 34.15 hours per system problem, the differences between the two cases are significant, especially for the frequency. The means in Case 2 were calculated using Monte Carlo simulations, and, therefore, are the true mean values of the PCA outcomes. Very often, the mean values alone are used for decision making. Therefore, an accurate calculation of mean values is vital.

Cases 3 and 4, were performed in a similar manner to Cases 1 and 2 but used the WECC system instead. To investigate the effect of wind ramping on the PCA results, another two cases, i.e., Cases 5 and 6, were performed to study the WECC system with and without wind ramping and were compared. Case 5 used the enhanced PCA for the WECC system with 10% more wind generation than the base case, considering their IIOs and CMOs, while Case 6 used the enhanced PCA for the WECC system with 10% more wind generation than the base case, but not considering their IIOs and CMOs.

The mean and standard deviation values for frequencies and durations of the overvoltage, undervoltage, and system problems for Cases 3 - 6 are tabulated in Table E.2. The Expected Energy Not Supplied (EENS) for each of Cases 3 - 6 is given in Table E.3. It was found that the system problem indices for Case 3 are not very different from Case 4, unlike the large differences for Cases 1 and 2, but they are still significant, especially for EENS. Table E.2 shows a large difference in EENS between Cases 3 and 4. The reason is that the mean values of frequency and duration in Case 3 are similar to that in Case 4; but the frequency/duration of some contingencies that cause load shedding in Case 3 are quite different from Case 4, which leads to the large difference between Cases 3 and 4. Also, using the enhanced PCA method, as used in Case 4, more information (e.g., the distribution) can be obtained about the EENS of the system than with the built-in PCA method, as used in Case 3. This indicates the necessity of using the enhanced PCA instead of the built-in PCA.

Table E. 2: ***Frequency, duration, and probability of the overvoltage, undervoltage, and system problems for Cases 3 - 6.***

Problems	Cases	Frequency (no. per year)		Duration (hours)		Probability (hours)
		Mean	Standard deviation	Mean	Standard deviation	
Overvoltage	3	642.6	N/A	4.1	N/A	2,634.7
	4	655.3	227.8	4.4	2.6	2,712.9
	5	993.1	389.7	2.8	1.8	2,572
	6	680.6	273.5	4.7	4.4	2,889.8
Undervoltage	3	70.4	N/A	3.5	N/A	246.4
	4	72.3	27.0	4.2	4.8	277.4
	5	71.0	30.1	2.6	3.4	170.9
	6	75.8	33.2	3.5	6.1	227.9
System	3	649.8	N/A	4.2	N/A	2,729.2
	4	663.0	230.9.5	4.4	2.7	2,741.5
	5	994.7	392.0	2.8	1.8	2,582.1
	6	681.1	274.2	4.7	4.4	2,901.2

*Table E. 3: **EENS for Cases 3 - 6.***

Cases	EENS (MW)	
	Mean	Standard deviation
3	81.66	N/A
4	22.69	29.44
5	8.8e-4	3.9e-3
6	2.14e-3	8.68e-3

A comparison of Cases 5 and 6 shows that the frequency of overvoltage problems and system problems in Case 5 is larger than Case 6, while the duration in Case 5 is smaller than Case 6. The reasons are as follows. Wind ramping contingencies, including the IIOs and CMOs, are considered in Case 5 but not Case 6. Hence, contingency frequency in Case 5 is larger than Case 6. The duration of a wind ramping event is usually several hours while conventional contingencies, such as generation and branch failures, can last for weeks. This explains why the duration in Case 5 is smaller than that in Case 6. Table E.3 indicates that consideration of IIOs/CMOs will significantly increase the system problem frequencies but reduce the system problem durations.

Tables E.3 – E-6 also show that voltage problems are an issue for the example case study. One of the reasons is that many buses had voltage issues in the base case, as discussed in Chapter 6. Another major reason is the rare event approximation used in the calculation, as discussed in future work below.

Additional information can also be extracted from the PCA output generated by the enhanced PCA. For example, the standard deviations calculated in Case 2 indicate how the estimates of system problem frequencies and duration deviate from the mean. While the mean values of frequencies and duration provide much information about the system problems, it is also interesting to see the shapes of their PDFs, as shown in the case study results of Chapter 6.

Contributions

The major contributions of this study include the following:

1. Outage data were collected from disparate data sources, including NERC's TADS and GADS and some other sources, and a reliability data repository including both raw data and statistics was developed for various grid components including transmission circuits, transformers, and conventional generators;
2. The potential poolability issue of grid component outage data that impact the probabilistic contingency analysis was identified and a solution was provided to resolve this issue by using formal statistical testing and characterizing the statistics of outages using distributions;
3. The fast ramping events of renewable generation caused by intermittencies were identified, and modeling such events as generator outrages (i.e., IIOs and CMOs) based on their actual impacts was proposed;
4. A method was developed for extracting ramping events from real wind generation data, and probabilistic models and parameterizations of the models for IIOs and CMOs were completed such that renewable outages can be naturally included in the PCA framework;
5. A Monte Carlo simulation was implemented using samples from distributions of grid component outages (related to both conventional grid components and renewables) in the add-on Python modules to enhance and drive the PCA capability of PSS/E;
6. A "well-being" method was implemented in the enhanced PCA capabilities to facilitate the decision-making process;
7. Case studies were performed using both example and utility scale systems, and numerical results were produced to assess and confirm the impacts of data poolability, as well as IIOs and CMOs associated with renewable generation.

Conclusions

This study provides a number of important insights related to probabilistic contingency analyses, including the following:

1. The study shows that data poolability is indeed an issue that needs to be resolved first before performing a probabilistic contingency analysis (PCA). As shown in this study, outage data for many different types of grid components cannot be pooled and their statistics need to be described by distributions.
2. The results also show that a majority of outage data from conventional generators can be pooled. For the rest of conventional generators, the distributions of their outage statistics are relatively narrow compared to other grid components. These results also are an indicator that environmental conditions have significant impacts on grid component outages.
3. Based on the analyses of actual renewable generation data, the results show that generation of a single wind farm can ramp up and down within a very short time period (i.e., tens of minutes). In addition, this type of ramping event can happen concurrently for multiple wind farms and can cause more significant impacts on grid operation. The impacts can be similar to conventional generator outages, and such fast ramps should be considered as outages.
4. Case studies performed in this project show that Monte Carlo simulations implemented in the enhanced PCA are capable of computing the true mean values and providing statistics of system problem indices. This will provide valuable information for developing mitigation actions in the planning process.
5. Increasing penetration levels of wind generation will definitely have increasingly larger, but different impacts on grid reliability. Due to the nature of outages related to intermittency of renewable generation, the system problem frequencies will increase and the durations will decrease.
6. The well-being approach can be implemented in the enhanced PCA framework and additional information can be provided to facilitate decision-making.
7. The performance of parallel processing cannot be justified and needs to be further investigated.

Future Work

As a follow-up to the current project, additional studies are proposed to further refine the tool and transfer the technologies to industry. The following studies are identified for future potential activities:

1. To supplement the data collected in the study, outage data from other sources will be sought, such as Canadian experience, and analyzed to update and expand the reliability data repository.
2. Solar PV generation is another important source of renewable generation. It will be included in the PCA by collecting and analyzing solar generation related data.
3. The interval of the time series wind generation data used in this study is 5 minutes. The algorithm for extracting the fast ramping events is relatively simple, i.e., it considers the monotonic increase or decrease in the ramping events. There exist time series data of different temporal resolutions (e.g., 2 second intervals) for wind generation or wind speed. Therefore, fluctuations of wind generation may occur during a ramping event and the event may still qualify as an outage. A generic algorithm will be developed to extract fast ramping events from data of different temporal resolutions to support the PCA study using the enhanced PCA tool.
4. The PCA built in existing software including PSS/E calculates probabilistic indices of system problems approximately based on the rare event approximation. When the probabilities of contingencies are relatively large, the rare event approximation can no longer be used, and calculated probabilities may deviate significantly from the real values, or even be larger than 1.0. This is especially true for higher order contingencies since they are not mutually exclusive. Theoretically, the total probability can only be calculated exactly using the inclusion-exclusion principle, which may be difficult because large computational efforts are required when the number

of contingencies is large. A new quantification scheme is needed and will be developed to more precisely calculate the probabilistic indices in the PCA.

5. We will reach out to more utilities for exercising and refining the enhanced PCA tool and demonstrating the capabilities of the tool.

Acronyms

CA	Contingency Analysis
CDD	Contingency Description Data
CDF	Cumulative Density Function
CMO	Common Mode Outage
CMY	Circuit-Mile-Year
CY	Circuit-Year
DCA	Deterministic Contingency Analysis
EENS	Expected Energy Not Supplied
FERC	Federal Energy Regulatory Commission
GADS	Generating Availability Data System
HBM	Hierarchical Bayesian Method
IIO	Intermittency Induced Outage
IRI	Incremental Reliability Index
KDE	Kernel Density Estimation
LOLE	Loss of Load Expectation
MLE	Maximum Likelihood Estimation
NERC	North American Electric Reliability Corporation
NPP	Nuclear Power Plant
OMA	Operation, Maintenance, and Administration
OSD	Outage Statistics Data
PCA	Probabilistic Contingency Analysis
PDF	Probability Density Function
PRA	Probabilistic Risk Assessment
SAIDI	System Average Interruption Duration Index
TADS	Transmission Availability Data System
TO	Transmission Owner
TY	Terminal-Year
UIC	Unit Interruption Cost
WECC	Western Electricity Coordinating Council

Acknowledgement

The authors are grateful to the Drs. Sun Wook Kang, Chunlian Jin, and Julia Matevosjana from ERCOT and Dr. Milorad Papic from Idaho Power Company for their help with renewable generation and system data as well as the time and efforts they dedicated to providing invaluable comments and feedbacks on this study.

1. Introduction

As a whole, the electric power industry has “standardized” on a relatively few software tools for planning and operation. Around these few *de facto* standardized tools, individual utilities have built a resource base of skilled personnel, required data and data handling procedures, required analysis capability, and confidence in the resource base, which is a huge investment. It is foreseen that the industry will continue to use these tools until the next generation of grid analytical tools becomes available, mature, and accepted by the industry. While these tools perform their intended function, they are fast becoming outdated and inefficient, and would benefit from new advances being made in probabilistic grid analysis. There is a need to create a transparent means to make existing software tools computationally more efficient and also include additional capabilities to address the emerging challenges in grid planning and operation, such as probabilistic risk assessment (PRA) capabilities. Industrial users are extremely cautious about adopting new tools or even upgrading existing tools. With input from industry stakeholders, external add-on modules, if developed and integrated with these *de facto* tools properly, can enhance existing tools and be accepted more easily by the power industry since the core and usage of the tools are untouched. This can be considered a short-term and yet economic and feasible solution. The enhanced tools bring about new features and additional capabilities that are specifically designed to deal with new issues, such as the increasing uncertainties in the grid.

Uncertainties always exist in grid planning and operation and have been mostly dealt with deterministically, e.g., by performing assessments based on multiple scenarios or worst-case scenarios, such as summer peak load conditions. A growing concern of the power industry is the ever-increasing uncertainties due to the high penetration level of renewable generation. It is recognized that addressing such concerns is increasingly difficult for the majority of existing tools that were mainly built based on the traditional deterministic approaches.

Contingency analysis (CA) is a major technique and practice in transmission planning for compliance with criteria mandated by NERC (North American Reliability Corporation) and FERC (Federal Energy Regulatory Commission). Traditional contingency analysis determines whether the system may experience certain problems or violations, such as a loss of load, overloading, or voltage violations, due to specific equipment failures or a combination of equipment failures (i.e., a contingency) for the given system configuration and projected load profiles. Typically, a CA is deterministic and used to evaluate system violations due to, e.g., the failures of major components in the system, such as a single transmission line or a single transformer (i.e., N-1 contingency). The potential for system problems is directly related to system reliability; therefore, CA results can be used to calculate a set of metrics to measure reliability in transmission planning. The CA results can also be used offline and online to identify contingencies that cause system problems, which in turn are used by utilities to develop a mitigation plan and determine remedial actions for reliability improvement, as needed. This traditional framework solely relies on deterministic contingency analyses (DCA), i.e., they consider only the occurrence of contingencies without differentiating the likelihood of occurrence among different types of contingencies. The implied assumption of equally likely contingencies in DCA can lead to over conservativeness in decision-making for planning activities. More importantly, it is much more difficult for the DCA to treat variable generation from renewables. Therefore, DCA lacks the sophistication needed to deal with the greater uncertainty introduced by increasing renewable generation. The capabilities to address this uncertainty now exist and need to be integrated into the current tools.

A probabilistic contingency analysis (PCA) is thus needed. Probabilistic planning including PCA has been in use for several decades. It, however, has never been widely adopted in planning practices by utilities. There exist several barriers to the successful implementation of probabilistic planning, which include but are not limited to availability of quality data and capable tools, and how to use the outcome of PCA for decision-making. Data collection and accumulation take a long time following the appropriate procedures. At the transmission level, the transmission data availability system (TADS) and generating availability system (GADS) that were developed and have been continuously supported by NERC are valuable reliability data sources for probabilistic planning, although the reliability data are generally provided in an aggregated form without many details.

Another major issue is that the traditional DCA (and the existing PCA built into some tools such as PSS/E and TRANSCARE) focuses on contingencies associated with conventional grid components or equipment, such as conventional generators or transmission circuits. A wind farm or solar plant is simply modeled exactly like a conventional generator. Therefore, outages of renewable generators are the same as the ones for conventional generators, i.e., a loss of generator that is mainly due to random failures of generator parts or components. Other types of non-conventional contingencies and their associated uncertainties that also affect the contingency analyses, especially those related to renewable generation from solar and/or wind in the grid, are either not included or included deterministically. The relevant uncertainties with renewables include not only the randomness of component failures, similar to the conventional contingencies, but also the uncertainties due to the uncontrollable variability and intermittency. Therefore, the contingencies associated with renewable sources will very likely be a major concern in the near future and will have to be considered together with other contingencies in evaluation of their impacts on reliability metrics in transmission planning.

As an extension of the previous project [13], the objective of this study is to, based on the developed tool, further enhance it by (1) completing an outage data repository with different categories of conventional generators; (2) including renewable variations in probabilistic contingency analysis, and building that into the add-on Python modules for PSS/E to further enhance the PCA capabilities of PSS/E; (3) applying the enhanced PSS/E tool to a large scale utility system with different penetration levels of various renewable sources; and (4) performing a demonstration of the usefulness of the probabilistic reliability metrics calculated to complement the deterministic metrics being employed by utilities. The corresponding functional modules have been developed using Python and integrated with PSS/E, and can be readily used by the researchers and utilities.

BNL has been working closely with ERCOT and Idaho Power Company to conduct a use-case, based on the WECC planning model, to compare and demonstrate the strength and weakness of both deterministic and probabilistic reliability metrics and how the two types of metrics can be used to complement each other in transmission planning.

2. Objective and Methodology

The overall objective of this study is to model and parameterize the uncertainties associated with grid contingencies such that these uncertainties can be accommodated in the probabilistic contingency analysis (PCA) framework to further enhance the PCA capabilities of an existing planning tool. For the purpose of developing a complete reliability data repository in a single document, the results from a previous study [13] are also summarized and included here. In addition to the issue of "data poolability" addressed in the previous study, the issue of how to model the intermittency and variability of renewable generation in the PCA with a focus of including the outage models of renewable generation in the PCA will be studied in detail, and more importantly a demonstration study will be performed using a utility-scale planning model.

2.1 A Formal Statistical Test for Data Poolability

It is well-known that different types of components, as well as different sub-types of a particular type of component have different failure or outage rates. Outage frequencies of grid components (e.g., transmission lines) at different voltage levels vary and accordingly, failure records from different sources can be combined to increase the population size. It is very common that the outage data of grid components that are relatively homogenous (e.g., 200–399kV lines) is aggregated to generate a constant outage frequency [1 - 6] and duration of the outages, which are the standard input parameters for performing PCA built into existing tools, such as PSS/E and TRANSCARE [7].

The average values are the maximum likelihood estimates (MLE) of the outage frequencies or durations assuming the same frequency or duration across different data sources for a Poisson process describing outage occurrences [8]. The issue of using a constant outage frequency for a category of grid components (e.g., 200 – 399kV transmission circuits) is that it neglects the variability of outage occurrence among different utilities or transmission owners (TOs). The factors that may have significant impacts on the outage frequency or duration, i.e., the outage variability, may include environmental conditions and maintenance schedules that differ between utilities. That means outage frequencies or durations for the same category of components may be different from utility-to-utility or from region-to-region. It is thus desirable to capture such variability in probabilistic analyses [8]. This is the so-called "data poolability" issue and was addressed by using a formal statistical test process to determine whether there is a need to model the population variation of data for a specific class of grid components [8].

2.2 Modeling and Parameterization of Intermittency Induced Outages for Renewable Generation

Existing contingency analysis focuses on new operating point of the system upon the contingencies or outages of conventional grid components or equipment by ignoring the transient dynamics or the process of equipment failure. A renewable generator such as a wind farm is modeled but usually as a conventional generator with a pre-specified capacity that is determined by utilities from, e.g., a seasonable average generation. Following this practice, the contingencies of renewable generators are characterized by the random failures of the wind turbines and other equipment, e.g., failure of mechanical parts, the converters, controls, etc. Although the uncertainties of such failures can be captured in PCA using outage frequencies and durations similar to a conventional generator, other types of uncertainties or non-conventional uncertainties that also affect the contingency analyses, especially those associated with intermittent renewable generation from solar and/or wind in the grid, are not included.

There is no issue with modeling renewable sources as generators in the system. Since the majority of renewable sources are installed in a distributed manner, individual sources may have negligible or very limited impact on the transmission system. Therefore, generator models can be built by properly aggregating the distributed renewable sources that are geologically close to each other for obvious reasons, as needed. The utility-scale wind farms or solar farms can be individually modeled as conventional generators in the planning model. The major difference between conventional generator outages and renewable outages is that different outage modes for renewables must be considered and modeled, i.e., in addition to a complete loss of generation[4], under- or over-generation of renewable generators due to generation ramping events also must be explicitly modeled. This is because the causes of these outage modes may be very different.

A solar farm or wind farm may fail due to different reasons and the generation level may drop to zero quickly and stay at zero before it is repaired and placed back online. A similar effect can be produced by a fast ramp-downs for a wind turbine or solar array. The wind speed or direction may change suddenly, which subsequently can reduce or even increase the wind generation quickly from the generation level prior to the change, and the post-ramping generation level may be kept at the reduced or increased level for some time, i.e., the duration of the ramping. In other words, a fast wind ramping can behave like a failure or partial failure of a wind turbine and can have the same impacts on the system, i.e., changes in the generation level. Therefore, a complete loss of wind turbine can be due to not only equipment failure (e.g., failure of a step-up transformer), but also very low wind speed (less than the cut-in speed) or very high wind speed (higher than the cut-out speed).

In addition, the prediction of instantaneous changes in wind speed or direction is very difficult if not impossible even in the very short term, let alone in the time horizon of transmission planning. Fast wind rampings can be considered random events. Therefore, the uncertainties of renewable generation not only include the randomness of component failures but also the randomness due to the uncontrollable intermittency of the generation.

The generation of a wind or solar farm can change dramatically within a short time period in terms of tens of minutes or even minutes. For the purpose of studying contingency analysis, sudden changes in generation levels are of interest. Such changes may be due to generator failures but may also be caused fast ramping events in wind farms. It is recognized that ramping events are not finished instantaneously but this may also be true for a loss of generation or de-rated generation of conventional generators. Upon the failure of a conventional generator, its generation level does not necessarily drop to zero instantaneously. Instead, the generation is likely to coast down to zero due to the inertia of the generator. Usually, a generator failure may take a long time to repair [9]. After a ramping the wind generation may (or may not) recover to the pre-ramping level in a short time period. However, a sustained outage for a transmission circuit is short, and is defined as "an automatic outage that lasts for more than one minute"[5] in [10]. This indicates that any outages of durations longer than one minute degrade system reliability in the current practices of PCA. As shown below, the durations of the wind rampings is usually short but may also last for hours. Therefore, fast wind rampings can be considered valid outages, i.e., intermittency induced outages (IIOs) [14].

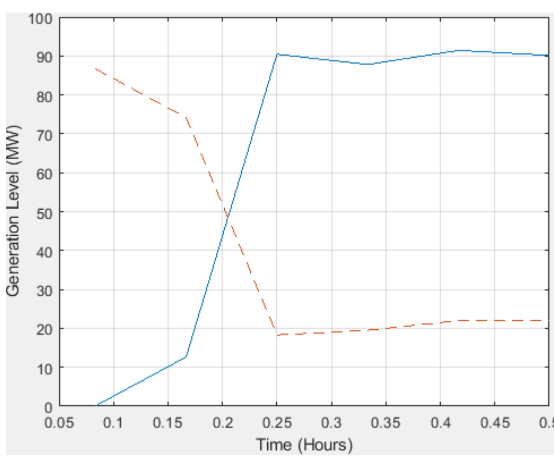

Figure 2. 1: ***Example fast ramp-ups and ramp-downs in wind generation.***

[4] A conventional generator may also fail partially with a de-rated generation. However, data for such partial failures or outages are rare and usually unavailable.
[5] There is no such a definition for generator outages.

Figure 2.1 shows an example of two fast rampings extracted from real generation data of two wind farms. Note that these two rampings did not happen simultaneously and are only plotted together. An up-ramp (or over-generation represented in solid line) is completed within around 20 minutes and the generation level increases from zero to 90MW. The dashed curve indicates a down-ramp (or under-generation) that causes a reduction of almost 70MW within 20 minutes. In addition, the post-ramping generation levels for both cases are maintained for more than 30 minutes. The examples shown in Figure 2.1 are a clear indication that these "intermittency induced rampings" can have significant impacts in terms of quick changes in generation level with different durations. Therefore, such rampings have a similar impact on the grid as outages of conventional generators. Figure 2.1 also indicates that wind generation possesses unique outage modes that differ from the conventional generators, i.e., increased generation.

These uncertainties due to generation intermittency will increase much more quickly as more and more renewables are installed. Therefore, the IIOs associated with renewable sources will very likely be an increasing concern in the near future and a major concern in the future and have to be considered together with other outages in the contingency analysis. That means the probability of occurrence of different outage modes needs to be accounted for based on the cause of the outage modes in the PCA.

For conventional generators, dependent failures or common mode failures are very common for co-located generation units. It is known that wind farms may be built along wind corridors, where wind is available all year long. Therefore, a concurrent change in wind generation can happen for adjacent wind farms and even for wind farms that are not physically very close to each other. As indicated in [11], the correlation[6] decreases with distance, but can still exist and be relatively strong for wind farms that are 1,000 km away from each other. For the under- or over-generation modes caused by sudden changes of wind speed, the CMOs of multiple renewable sources are due to high geospatial correlation of wind farms that are close to each other. This poses a more severe threat to the grid and needs to be captured.

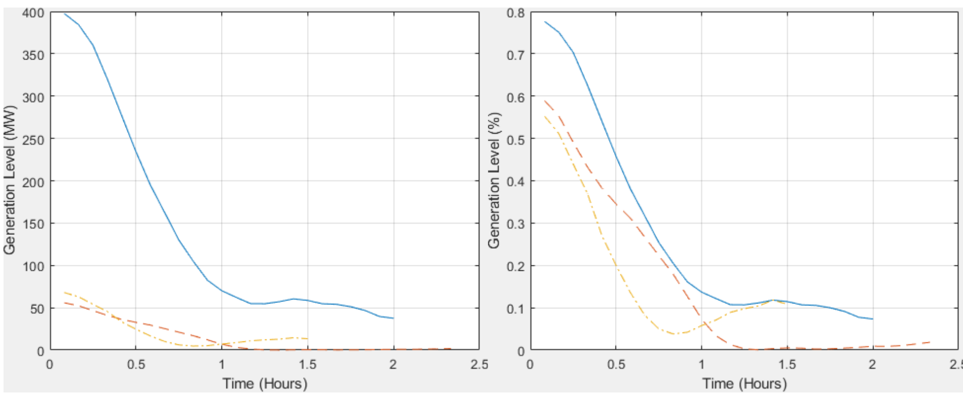

*Figure 2. 2: **Example concurrent rampings of highly-correlated generation sites.***

An example of concurrent ramping or CMO is shown in Figure 2.2. Three ramp–down events from three highly correlated (based on Pearson's correlation coefficients, as will be discussed in Chapter 3) wind farms occurred approximately at the same time. The rated capacities of three wind farms are very different, as seen in the ramping curves in the left of Figure 2.2. Normalized ramping curves shown in the right indicate they are very similar to each other in terms of percentage of initial generation (around 60% - 80% of rated capacities) and reduction (to less than 10%) in generation (normalized using rated capacities of individual wind farms). This CMO caused a ~500MW drop in generation in less than one hour and lasted up to two hours, and therefore, can have a significant impact in contingency analysis.

In fact, high wind ramps have been frequently observed in ERCOT. Table 2.1 summarizes the top three examples in up and down ramps of wind generation, respectively. There were a total number of 193 events (106 up ramps and 87 down ramps) involving 1-hour changes that are more than 2,000MW among the total 8,760 data points in 2017. Such frequent and significant hourly variations or intermittencies in wind generation may add to higher reserve requirements and more frequent reserve deployments. This is more challenging under light load conditions, when wind generation is high since it is supplying a higher

[6]Correlation can be mathematically quantified using, e.g., Pearson's correlation coefficient, from historical time series wind generation data.

percentage of the load. For the economic operation of the system, more conventional generators may be shut down, which may cause grid inertia to be further reduced and can potentially result in deterioration of frequency performance during grid disturbances. All the data shown are from the ERCOT public website.

*Table 2. 1**: Top Three 1-hour Changes in Wind Generation Output in 2017 (Source: ERCOT)***

Ramp Types		Up ramps	Down ramps
1	Change (MW)	+4,562	-3,975
	Start output (MW)	5,054	9,037
	End output (MW)	9,616	5,061
	Percentage of variation	90.3%	44%
2	Change (MW)	3,584	-3,430
	Start output (MW)	9,616	9,844
	End output (MW)	13,200	6,594
	Percentage of variation	37.3%	33%
3	Change (MW)	3,437	3,250
	Start output (MW)	3,855	12,365
	End output (MW)	7,292	8,935
	Percentage of variation	89.2%	27.7%

In summary, the major differences between conventional generator outages and renewable generation outages are [8]:

- In addition to a complete loss of generation, clearly under- or over-generation of renewable generators due to ramp-downs and -ups occur and must be explicitly modeled.
- For each of the outage modes, there can be associated common mode outages (CMOs) even for outages from different generation sites.

The probabilistic models and the parameterization of the models for such contingencies need to be developed to provide input data for frequencies and duration times of different outage modes in the PCA framework. For this purpose, the historical solar or wind generation data can be collected to analyze the statistics of these outage modes, e.g., how often the reduction in wind generation at a certain level occurs and how long this reduction lasts[7]. This method is similar to the calculation of frequencies and probabilities of occurrences of different fast solar ramping events using the solar irradiance data and electrical models of solar plants, as shown in [12]. In [12], fast solar rampings were captured by analyzing high-resolution irradiance measurement data and characterized in terms of how much the irradiance changes within a certain time period, which was used for calculating the statistics of solar generation reduction. Such probabilistic models, once developed for wind or solar generators, are readily used as input and implemented in the developed tool for the enhanced PCA analysis.

2.3 Enhancing Probabilistic Contingency Analysis Capability of Existing Tools

As indicated above, existing tools capable of performing PCA, such as PSS/E and TRANSCARE, take as input data only averaged outage parameters, i.e., outage data of grid components that are relatively homogenous are aggregated and averaged. The outage data for both conventional grid components and intermittencies of renewable generation, if they cannot be aggregated and averaged, must be described by distributions (which is actually true for most of the grid components, as will be shown later), then the existing tools are incapable of calculating the probabilistic reliability indices and need to be enhanced. The mean values of the probabilistic reliability indices should not be calculated using the average or mean values of outage parameters for grid components. Based on the distributions of different input parameters, a generic scheme for improving PCA capability of the current contingency analysis tools is shown in Figure 2.3 [13]. The scheme is to use a Monte Carlo simulation to calculate the probabilistic indices based on the deterministic CA, which not only enables the retrieval of the mean values, but also other statistical properties of the reliability indices. The mean values calculated this way are the true means.

[7] Strictly speaking, any sudden changes in the system should be include in contingency analysis. If the system load changes suddenly due to, e.g., a demand response, such a change can also be considered as a contingency and included in the contingency analysis.

The essence of the solution is to implement the proposed scheme by developing interfaces around the current tools without making changes to the tools themselves due to the cautiousness and conservativeness of utility users. The interfaces will generate samples from the statistical distributions of the different outage parameters as input and extract the risk measures and metrics from the output for decision-making. In the existing PCA capability in PSS/E, the point estimates are directly used to specify the statistics of a set of grid components (e.g., all 230kV lines) or a particular complement (e.g., a single 230kV line as specified) in the Outage Statistics Data (OSD) file and/or Contingency Description Data (CDD) file [7].

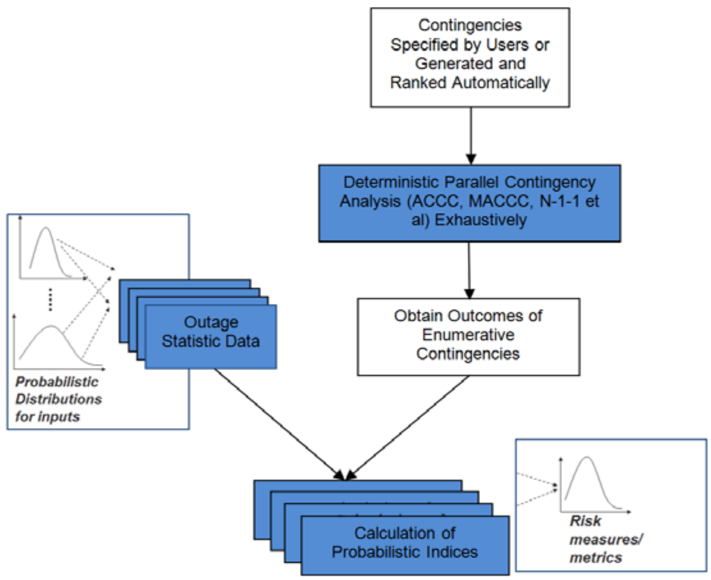

*Figure 2. 3: **A Scheme for Enhancing PCA Capability***

In summary, the enhancement scheme does the following:

1) instead of a single OSD or CDD file, it generates a number of OSD/CDD files containing variates for frequencies and durations sampled from the distributions for different types of component outages;
2) repeatedly runs the PCA in PSS/E using an OSD/CDD file at each run; and
3) post-processes the output OSD files to develop the distributions for the probabilistic indices.

This entire process has been automated by implementing it via a set of Python scripts that interface with PSS/E to perform a complete PCA study.

2.4 Calculation of Probabilistic Reliability Indices

The deterministic CA focuses on the assessment of impacts (i.e., the system problems) of specified contingencies or sequences of contingencies. System problems caused by certain contingencies usually include overloading or voltage violation (under- and over-voltage) for the entire system or a subsystem. PCA not only considers the impacts, but also the frequencies (or probabilities) and duration times of such impacts, i.e., the outcomes of PCA are the probabilistic reliability indices. To calculate the probabilistic indices (e.g., frequency or duration time) for system problems, each component is modeled with two states, i.e., in-service and outage or failure [52]. For a single component, the frequency of transition from in-service to an outage and the frequency of transition f from an outage to in-service are the same:

$$f = P_{in-service} \times \lambda = P_{outage} \times \mu$$

where P represents the probability of being in an in-service or outage state, λ is the failure rate (number of failures per unit time), and μ the repair rate (number of repairs per unit time). Therefore, the probability for the component to stay in the outage state within duration time d is given as:

$$P_{outage} = f \times d \tag{2.1}$$

The calculation of a transmission line outage frequency or duration needs to account for both the element-initiated and non-element-initiated (or terminal-initiated) outages, i.e.,

$$f = length \times fmt \times ft$$

$$d = \frac{fmt \times dmt \times length + ft \times dt}{f}$$

Where $length$ represents the length of a single line. For transformer or generator outages, the frequencies and duration times are derived from the collected outage data, as will be shown in Section 3.1.3. A double or triple contingency represents simultaneous occurrence of two or three single outages. If the occurrence of two or three outages is dependent, e.g., a common mode failure or dependent failure, collection of such outage mode data needs to be performed to calculate the frequency or duration of such contingencies in the same way as for a single contingency. More common is the independent occurrences of multiple single contingency events. The frequency and duration can be calculated using the data for single contingency outage data. The frequency of a double contingency consisting of two independent single contingencies, A and B, is given in Equation (2.2*)*,

$$P_{AB} = P_A \times P_B = \frac{(f_A \times d_A)(f_B \times d_B)}{8760^2}$$

$$f_{AB} = f_A \times P_A + f_B \times P_B \tag{2.2}$$

The mean duration time that the two components stay in the outage state is given using Equation (6.1), i.e.,

$$d_{AB} = \frac{P_{AB}}{f_{AB}}$$

The calculation of the probabilistic reliability indices is straightforward. The frequency of a system overloading condition is usually calculated based on a rare-event approximation [9].[8] by the sum of frequencies of all contingencies that caused the system overload, and the duration of such overloading condition is the average of duration times of all relevant contingencies.

The existing PCA in tools such as PSS/E is accomplished by performing the deterministic contingency analyses followed by a post-processing to calculate probabilistic indices using the average values of different contingencies. This practice, however, simply neglects the difference among various outage data sources and the calculation of probabilistic indices may become problematic.

2.5 A Decision-making Process Using Probabilistic Reliability Metrics

Decision-making based on deterministic metrics from contingency analysis is very straightforward and has been used for decades by the power industry. On the other hand, probabilistic planning can rigorously model and, thus, capture uncertainties that deterministic techniques cannot address. However, the major issue is that it is difficult for a utility to use probabilistic results in the decision-making process. Unlike the deterministic process that focuses on the compliance of existing metrics with well-established transmission planning standards, there is a lack of probabilistic planning criteria for the transmission planner to adopt.

Chapter 4 reviews the probabilistic planning criteria and illustrates how to use the probabilistic criteria together with the deterministic ones to facilitate and help the utilities in making a planning decision. The focus of this study is on the discussion of probabilistic reliability metrics calculated from PCA, which are not presented but can be readily done by using an enhanced PCA tool. In addition, a well-being approach that

---

[8] The rare-event approximation is only applicable when the outage frequencies are small.

can provide more information about the system states and its implementation is discussed in detail using a case study.

3. Reliability Data Collection, Analysis, and Repository Development

Recognizing the need for probabilistic contingency analysis (PCA), raw outage data for different grid components, such as transmission circuits, transformers, and generators, have been collected, categorized, and analyzed statistically for decades by different parties. However, the common issue is that there is a lack of details about the outage data (e.g., the grid conditions when an outage occurs) and raw outage data are considered proprietary and not available. Instead, only lumped and averaged outage data are given in most studies to derive outage parameters, e.g., in [1 - 6]. These average values, e.g., frequencies or durations of grid component outage, are the maximum likelihood estimation (MLE) of these parameters, assuming the same mean value across different data sources for a Poisson process, that outages actually occur [15]. However, this assumption of the same means across different data sources (or data from different entities or regions) has not been examined or validated yet in the literature.

On the other hand, factors that may have significant impacts on outage frequencies or repairs can be very region- or utility-specific, especially weather conditions (e.g., humidity and temperature) and policies and procedures for replacing or repairing damaged equipment. The environmental impacts can be huge for grid components, especially for those that are located outdoors, such as transmission circuits. For obvious reasons, if environmental conditions are so different, then the differences of impacts on outages can no longer be neglected, and simply lumping and averaging the outage parameters from different utilities and regions[9] become problematic, which is the so-called "data poolability" issue [16]. To address this issue, a formal statistical test process was adopted, and a distribution was used to model the population variation of data, as needed [8]. One can always calculate the mean values and other statistics from the distributions, but not vice versa. A Monte Carlo scheme can be introduced to accommodate the distributions for outage parameters, as shown in [17].

It is always beneficial to have a generic and publicly available repository for (1) raw outage data, such that individual practitioners can perform their own statistical analysis, and (2) the distributions of outage parameters for grid components that did not exist and are needed in the PCA studies. In this study, recognizing the importance of reliability data, a repository was created with an additional collection of different generators. The report expands the studies in [8] and [17] with generator outage data collection and analysis, as well as intermittencies of renewable generation. The outage repository development provides a compendious and expandable repository for outage data from many disparate sources, as well as variability of renewables by addressing the poolability issue and developing a practical solution using statistical distributions that can be used by existing tools.

3.1 Outage Parameters of Conventional Grid Components for PCA

3.1.1 Nomenclatures and Input Parameters for PCA

The following nomenclature is most relevant to this study.

AC Circuit: A set of AC overhead or underground three-phase conductors that are bounded by AC substations [6].

Circuit mile: One mile of either a set of AC three-phase conductors in an overhead or underground AC circuit, or one pole of a DC circuit. A one-mile-long AC circuit tower line that carries two three-phase circuits (i.e., a double-circuit tower line) would equate to two circuit miles [6].

A Terminal: A transmission line end or cable end which is equipped with primary protection [1]. Most circuits have two terminals. However, multi-terminal circuits also exist.

Mile years (*Terminal years*): The product of number of line lengths in miles (terminals) and the line's (terminal's) exposure time in years [3].

[9] One can argue that a specific utility may have an abundant amount of reliability data and the average of this utility-specific data can be used. While this is possible, this is usually rare, and it is inappropriate to make such a claim. This is not often exercised since the scarcity of data is always an issue in probabilistic risk assessment. Complementary data from other utilities/regions are always beneficial to better capture the uncertainties. In addition, we can easily add more weight to utility-specific data mathematically, e.g., by using a two-stage Bayesian approach, as discussed in [40].

Element-initiated outages: Outages that were initiated on or within an element. For AC circuits, this means that the outage is line-related, and a non-element-initiated outage means a terminal-caused line outage [6].

Refer to [1] and [3 - 6] for more details about other terminologies.

Contingencies are outages of grid components, such as transmission lines, transformers, generators et al. The major input parameters to a PCA include both the outage frequency and duration of various components [7]. To obtain these parameters, two types of data are needed, i.e., counts of outage events X_i and the exposure data O_i, from the i^{th} data categories or sources.

Existing studies all assume (implicitly or explicitly) that the occurrence of outages follows a homogeneous Poisson distribution (i.e., a constant outage rate or frequency λ), and the probability of observing x outages in time t is

$$Pr(X = x) = e^{-\lambda t}(\lambda t)^x/x!$$

The most commonly used estimator of x/t for λ is the one that will maximize the likelihood of the Poisson distribution [8]. Therefore, for the given exposure data and outage event counts, the mean outage frequency is simply given in Equation (3.1), as used in [1 - 6]

$$E = \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} O_i} \tag{3.1}$$

Another parameter, the average duration time per outage, is an arithmetic mean. It is calculated using X_i for total duration time and O_i for the total outage number but has a different physical meaning than the frequency. The calculation of probabilistic reliability indices in PCA requires the input of frequencies and duration time of individual component outages. The following parameters for transmission line outages are used [7]:

fmt: Frequency for single-circuit outages per mile, i.e., the number of line-caused outages per circuit-mile-year.

dmt: Duration of single-circuit outages, i.e., the duration or the repair time of per line-caused circuit outage.

ft: Frequency for terminal-caused single-circuit outages, i.e., number of terminal-caused circuit outages per circuit-year.

dt: Duration for terminal-caused single-circuit outage, i.e., the duration of each terminal-caused circuit outage.

For transformer and generator outages, two parameters, f the outage frequency and d the duration per outage, are needed. Parameters for other types of outages (e.g., circuit breakers) can be found in [7].

3.1.2 Raw Data Collection of Grid Component Outages

The focus of this data collection effort is on outages for three types of grid components that are most relevant to contingency analysis, i.e., transmission lines, transformers, and generators. The major data sources are NERC (North American Electric Reliability Corporation) TADS (Transmission Availability Data System) and GADS (Generation Availability Data System) since they contain outage data from different transmission owners or NERC regions. Some other publicly available data sources were also explored and used in this study. The lumped outage data in [1 - 6] were extracted and treated as an independent data source. The raw outage data for transmission lines, transformers, and generators are summarized in Tables 3.1, 3.2, and 3.3, where "-" means "no data". The transmission lines and transformers are categorized in terms of voltage classes while the MW rating is used to classify generating units.

The following rules were used for the data collection:

1. Momentary outages (i.e., outages lasting less than one minute [10]) and outage types or failure modes (e.g., three-phase or single-phase failures) are excluded.

2. Outage data from the same NERC region or transmission owners (TOs) in different years were combined based on the assumption that the component outages of the same class are relatively homogeneous. This assumption was determined by the level of detail available for the outage data in the sources identified in this study. This assumption, however, is also considered reasonable because we do not expect significant differences in terms of environmental conditions and operation and maintenance practices among utilities in the same region or TO.

3. Data for DC circuits, converters, and circuit breakers in TADS are not sufficient to perform the proposed analysis. Also, these components are not the major sources for the contingencies, and therefore, the data for these components are not presented here.

4. The generator outage data were extracted from a NERC developed software platform, pc-GAR, which is a database for GADS [18] containing data from 1982 to 2014 (33 years).

Table 3. 1: ***Raw Outage Data for AC Circuits.***

Voltage Classes	Element-Initiated		Non-Element-Initiated		AC Circuit-mile-years (CMYs)	Circuits-years (CYs) or Terminal-years (TYs)	Time Period	References
	Outage Number	Duration (Hours)	Outage Number	Duration (Hours)				
Up to 109 kV	1,717	20,148	542	20,291	37,086.4	2,904.0 (TYs)	-	[1]
110 – 149 kV	2,581	80,521	1,360	6,085	142,104.4	10,764.0 (TYs)	-	[1]
	4.38	-	-	-	-	1,267 (CYs)	11/1972–10/1979	[2]
150 – 199 kV	74	4,794	17	252	7057.8	829.0 (TYs)	-	[1]
200-299kV	210	1,858.3	81	3,510.9	11,689	773.4 (CYs)	2008 – 2009	[19][20]
	173	1,112.4	99	469.3	21,423	468.4 (CYs)	2008 – 2009	[21][22]
	161	2,033.6	123	17,495.3	25,466	734.1 (CYs)	2008 – 2009	[23][24]
	69	2,313.7	117	5,446.2	12,999	1,109.2 (CYs)	2008 – 2009	[25][26]
	372	16,982.1	140	675.8	40,988	2,368.6 (CYs)	2008 – 2009	[27][28]
	72	1,273.1	28	166.6	5,464	220.0 (CYs)	2008 – 2009	[29][30]
	798	17,742.5	490	14,235.4	87,984	3,170.3 (CYs)	2008 – 2009	[31][32]
	809	86,818	980	15,778	123,990.1	6,016.5 (TYs)	-	[1]
	5.22	35.44	-	-	-	149 (CYs)	11/1972–10/1979	[2]
	2,992	-	719	-	232,500	7,265 (TYs)	-	[3]
	593	12,651	201	603	32,944	1,005 (TYs)	01/1977–12/1981	[4]
	675	20,426	391	2,099.7	72,363.5	3,002.1 (TYs)	1991 – 2000	[5]
300 – 399 kV	143	804.3	46	330.8	21,423	468.4 (CYs)	2008 – 2009	[21][22]
	115	4,794.4	152	1,324	15,436	717.6 (CYs)	2008 – 2009	[23][24]
	140	6,800.3	127	46,726.5	26,571	933.2 (CYs)	2008 – 2009	[25][26]
	54	7,104.4	56	165.0	6,536	210.0 (CYs)	2008 – 2009	[27][28]
	86	9,497.1	34	172.3	8,853	206.6 (CYs)	2008 – 2009	[29][30]
	41	448.7	35	719.9	8,941	258.9 (CYs)	2008	[33]
	258	856.3	144	2,164.9	20,862	267.8 (CYs)	2008 – 2009	[31][32]
	89	5,547	63	809	28,998.8	1,383.0 (TYs)	-	[1]
	1,752	-	1,430	-	232,900	5,177 (TYs)	-	[3]
	422	13,605	137	1,644	16,880	685 (TYs)	01/1977–12/1981	[4]
	778	48,112	379	1,036.6	56,350.9	2,189.6 (TYs)	1991 – 2000	[5]
400 – 599 kV	11	2.2	6	35.6	2,402	38.0 (CYs)	2008 – 2009	[19][20]
	0	0	0	0	473	2.0 (CYs)	2009	[22]
	0	0	0	0	2,295	32.0 (CYs)	2008	[23]
	5	105.5	24	2,144.7	5,116	139.3 (CYs)	2008 – 2009	[25][26]
	97	10,399.2	46	925.8	17,213	438.6 (CYs)	2008 – 2009	[27][28]
	0	0	0	0	47	1.0 (CYs)	2009	[30]
	228	1,651.3	206	5,496.5	33,605	501.2 (CYs)	2008 – 2009	[31][32]
	119	871	144	1,629	32,035.4	743.5 (TYs)	-	[1]
	7.86	38.8	-	-	-	61 (CYs)	11/1972–10/1979	[2]
	413	-	203	-	78,400	1,221 (TYs)	-	[3]

	7	89.48	30	67.8	4,729.8	40.0 (TYs)	1991 – 2000	[5]
600 – 799 kV	32	282.6	33	112.4	14,502	158.0 (CYs)	2008 – 2009	[23][24]
	18	204.9	13	211.4	4,400	62.0 (CYs)	2008 – 2009	[25][26]
	120	19,062	144	2,336	33,305.8	807.0 (TYs)	-	[1]
	214	-	131	-	24,798.0	456 (TYs)	-	[3]

It should also be pointed out that in some AC circuit data sources, exposure data were given in terms of terminal-years instead of circuit-years (see Column 7 for details). Since the number of terminals was not provided in these sources, an average value of 2.09 terminals per circuit [3] was used in the study. This is an indicator that some changes may be needed in the existing practices of data collection such that the data collected can directly support the intended probabilistic analyses, e.g., the PCA.

Common mode outages (CMOs) are another type of important outage that should be included in contingency analysis because their impacts can be significant and often dominate the result of the contingency analyses [34]. Such data are, however, very rare in the literature, e.g., in [34] and [35], and therefore, not analyzed here. A set of line-related and terminal-caused common mode outage data for transmission lines of different voltage levels is given in [36]. The frequencies and duration times for CMOs of transmission lines can be calculated similar to the calculation for a single transmission line, as shown in Chapter 2. A possible solution to this data scarcity is to develop models to calculate their parameters using, e.g., a multiple Greek Letter model [35] and the parameters of single outages. This, however, is beyond the scope of the study and will not be further discussed.

*Table 3. 2**: Raw Outage Data for Transformers***

Voltage Classes	Outage Number	Duration (Hours)	Transformer-years	References
Up to 109kV	381	103,866	3,038.5	[1]
110–149kV	1,481	253,823	9,114.5	[1]
150–199kV	167	92,824	612.0	[1]
200–299kV	794	107,203	6,044.5	[1]
	4	11	15	[31]
	5	191.3	15	[32]
300-399kV	213	36,038	1,669.5	[1]
	8	156.1	60.6	[21][22]
	9	22,176.2	54	[23][24]
	2	42.6	56	[25][26]
	2	3.8	20	[29][30]
	25	11,996.9	122.3	[31][32]
400-599kV	70	11,280	1,084.0	[1]
	16	226.9	52	[19][20]
	7	21.1	11	[21][22]
	17	6,106.4	84	[23][24]
	14	2,449	126.6	[25][26]
	17	275.4	190.1	[27][28]
	0	0	4.0	[29][30]
	40	5,487.8	354.6	[31][32]
600-799kV	151	23,398	2,364.0	[1]
	25	12,061	131	[23][24]
	8	3,552.4	54	[25][26]

*Table 3. 3**: Raw Outage Data for Conventional Generators***

Fuel Type	MW Rating	Outage Number	Duration (Hours)	Unit-years	NERC Region
Fossil Fuel	**0–399**	11,579	713,558.3	2,838.0	ERCOT
		4,959	238,829.9	782.1	FRCC
		13,092	933,361.6	2,261.2	MRO
		22,614	1,025,640.5	2,826.8	NPCC
		62,474	3,476,200.5	8,220.3	RFC
		43,462	2,349,441.1	8,078.4	SERC
		8,841	590,763.4	2,110.0	SPP
		27,931	1,359,333.4	4,065.6	WECC
	400–799	10,320	406,989.3	1,329.9	ERCOT
		5,459	185,191.8	646.8	FRCC
		7,393	244,865.8	818.7	MRO
		8,157	355,746.6	773.9	NPCC
		32,958	1,465,928.8	2,783.5	RFC
		24,009	1,044,170.5	2,546.0	SERC

		8,882	393,810.2	1,114.4	SPP
		17,183	556,682.5	1,220.3	WECC
	800–999	13,126	511,173.4	1,542.4	All
	1,000+	3,460	233,531.9	397.7	All
Gas/Jet Turbine	**0–99**	6,998	315,135.7	1,448.7	ERCOT
		5,760	355,881.4	2,865.4	FRCC
		5,631	690,826.4	2,491.2	MRO
		23,511	2,206,091.8	5,307.1	NPCC
		57,404	3,964,319.8	12,110.3	RFC
		13,904	1,788,745.6	5,967.1	SERC
		2,741	572,901.2	922.7	SPP
		19,436	793,055.7	3,099.7	WECC
	100–199	419	19,098.7	74.9	ERCOT
		511	20,663.1	106.5	FRCC
		1,061	38,594.1	261.9	MRO
		354	39,274.8	171.0	NPCC
		7,071	324,899.5	1,324.0	RFC
		5,357	265,593.8	1,487.9	SERC
		526	25,103.4	119.3	SPP
		1,170	33,324.0	169.3	WECC
	200+	809	63,047.2	144.3	All
Nuclear	**0–399**	–	–	–	–
	400–799	1,628	354,428.4	621.4	All
	800–999	3,710	693,651.5	1,274.8	All
	1,000+	3,986	997,137.2	1,515.4	All

For generator outages, the data for fossil-fueled plants of capacity greater than 800 MW, all nuclear plants, and gas/jet turbines of more than 200 MW were extracted together. The major reason for this is the relatively small number of such generators in the existing grid. Although the amount of data accumulated over 33 years can be fairly large, very few generating unit outages were from individual NERC regions, and the software (pc-GAR) could not report it.

Also, unplanned or forced outages, i.e., classes 0, 1, 2, and 3 [37], are included. In Table 3.3, unit-years are given in Column 5. It needs to be pointed out that, unlike transmission lines or transformers, a generator may spend a significant amount of time on scheduled maintenance or acting as non-spinning reserve. This is particularly true for gas/jet turbines since most of them only have a few hundreds of service hours annually. Since CA is performed on calendar time basis, using service-hours rather than unit-hours or unit-years to calculate the outage frequencies makes the CA more difficult.

3.1.3 Determination of Data Poolability and Modeling Population Variability

3.1.3.1 A Formal Statistical Test

A Pearson χ^2-test is a type of non-parametric or "distribution free" statistical test, which works on the actual categorical data instead of the assumed distributions of parameters. If a χ^2-test indicates that the failure data difference from different categories is statistically significant enough, then the records should not be pooled together and population variability should be modeled [15][38]. The χ^2 can be calculated using the following equation:

$$\chi^2 = \sum_{i=1}^{N} \frac{(X_i - E_i)^2}{E_i} \tag{3.2}$$

where X_i and E_i, i=1, 2, …, N, are the numbers of observed and expected failures, respectively. The number of degrees-of-freedom for the χ^2-test is $(N\text{-}1)$ and the value of $\chi^2(N\text{-}1)$ will be large if the number of observed failures significantly differs from that of the expected. Use of the χ^2-test may not be appropriate for small sample sizes. A "rule of 5" is widely used, i.e., the number of events in each category should not be smaller than 5 ($N \geq 5$) [38].

To test the poolability of data (including duration hours), the null hypothesis H_0 is that failure parameter is the same for data from different sources [15]. The process then includes the following steps:

1. Aggregate all failure data $\{X_i, O_i\}$ and calculate the mean value using Equation (3.1);
2. Apply the mean value to the data of each category to calculate the expected number of counts $E_i = EO_i$;
3. Calculate χ^2 using Equation (3.2);
4. If the calculated χ^2 is larger than $\chi^2_{Percentile}(N-1)$, and the percentile is large, e.g., 95%, then it is strong evidence that the null hypothesis is false.

A typical value for the 95th percentile can be chosen. If the null hypothesis H_0 is true, the probability of seeing such a large value of calculated χ^2 is less than 0.05 [15].

3.1.3.2 A Lognormal Distribution for PCA Parameters

If the Pearson test indicates that the data should not be pooled, a statistical distribution needs to be developed to capture this statistical difference among disparate data sources. There are no general rules to select a distribution, although different distributions can be compared using sensitivity analysis and additional outage data collection. Lognormal distributions have been commonly used for modeling many reliability parameters in nuclear PRA [15] and are also used herein. A lognormal distribution $LN(\mu, \sigma)$ is characterized by two parameters μ and σ, the location and scale parameters. The mean and variance for a lognormal distribution are $\exp(\mu+\sigma^2/2)$ and $\exp(2\mu+\sigma^2)(\exp\sigma^2-1)$ while the 5^{th} and the 95^{th} percentiles are $\exp(\mu-1.645\sigma)$ and $\exp(\mu+1.645\sigma)$, respectively [38].

An illustration of calculating the parameters for individual lognormal distributions for different outage variables is shown in [8]. The details are not discussed here. By applying the same procedures to all the raw outage data, the statistical results were calculated and are summarized in Table 3.4. The arithmetic means for individual components calculated using equation (3.1) are also included in Table 3.4, which shows that differences between the arithmetic means and means calculated from distributions can be very different. The mean values of outage parameters for components that are either poolable (i.e., outage frequency of transformers between 300 and 399kV) or of a single source (i.e., some generators) are shown in Table 3.5. Note that the units for mean values for outage rates and duration time are number of occurrences per year and hours per outage, respectively. Also, point estimates of parameters for CMOs of transmission lines can be found in [39] and can be directly used to account for transmission line CMOs in probabilistic contingency analysis.

Table 3. 4: ***Parameters and Statistics for Lognormal Distributions of Different Outages***

Components	Parameters	$\chi^2(N\text{-}1)$	μ	σ	Arithmetic Mean	Mean Value	Variance	5th Percentile	95th Percentile	Error Factors
AC Circuit: 200–299kV	fmt	682.8	-4.63	0.37	0.01	0.01	1.62E-5	0.005	0.018	1.84
	dmt	114,420.4	3.27	0.86	41.51	37.88	1,548.35	6.43	107.32	4.09
	ft	810.8	-1.85	0.60	0.20	0.19	0.015	0.059	0.42	2.66
	dt	96,440	3.03	1.17	22.82	41.10	4,995.77	3	142.24	6.89
AC Circuit: 300–399kV	fmt	922.6	-4.74	0.64	0.01	0.01	5.78E-5	0.003	0.025	2.85
	dmt	39,390	3.04	1.12	45.90	39.06	3,804.74	3.32	131.56	6.30
	ft	839.3	-1.45	0.55	0.34	0.27	0.03	0.095	0.58	2.46
	dt	312,978	3.46	1.49	46.97	96.25	76,004.2	2.74	367.93	11.60
AC Circuit: 400–599kV	fmt	77.1	-5.96	0.59	0.005	0.003	3.89E-6	0.00098	0.0068	2.64
	dmt	27,341.6	1.53	1.91	28.09	28.7	30,818.8	0.2	107.21	23.15
	ft	207.4	-0.90	0.82	0.31	0.57	0.31	0.11	1.57	3.87
	dt	6,336.2	2.65	1.12	22.59	26.54	1,752.6	2.26	89.36	6.29
Transformer: 300–399kV	d	174,195	4.23	2.18	271.87	734.3	6.16E+7	1.9	2,464.02	36.01
Transformer: 400–599kV	f	81.5	-1.60	0.70	0.09	0.26	0.04	0.065	0.64	3.14
	d	10,569.3	3.49	1.45	142.80	94.56	64,916.8	3.01	359.2	10.92
Fossil Fuel Generator: 0–399MW	f	8,655.9	1.74	0.21	6.25	5.83	1.46	4.08	8.00	1.40
	d	159,300	4.04	0.14	54.82	57.40	62.87	45.35	71.29	1.25
Fossil Fuel Generator: 400–799MW	f	4,326.8	2.35	0.18	10.18	10.63	3.77	7.76	14.08	1.35
	d	66,860	3.64	0.096	40.69	38.14	13.56	32.40	44.48	1.17
Gas/Jet Turbine: 0–99MW	f	11,892.9	1.27	0.35	4.14	3.77	1.80	2.01	6.27	1.77
	d	1,778,084.2	4.53	0.50	78.94	104.48	3053.19	40.81	209.1	2.26
Gas/Jet Turbine: 100–199MW	f	961.3	1.33	0.37	4.42	4.04	2.35	2.07	6.91	1.83
	d	43,616.0	4.03	0.41	46.56	61.23	698.09	28.49	110.95	1.97

Table 3. 5: ***Mean Values for Poolable Outage Data***

Components	Parameters	Mean Value
Transformer: 300–399kV	f	0.13
Fossil Fuel Generator: 800–999 MW	f	8.51
	d	17.8
Fossil Fuel Generator: 1,000+ MW	f	8.70
	d	67.5
Nuclear Generator: 400–799 MW	f	2.62
	d	217.7
Nuclear Generator: 800–999 MW	f	2.91
	d	187.0
Nuclear Generator: 1000+ MW	f	2.63
	d	250.2
Gas/Jet Turbine: 200+ MW	f	5.61
	d	77.9

Note that the mean frequencies for most of the generator outages are very high compared to transmission lines and transformer outages. This is mainly because the outages include all types of forced outage events, i.e., Classes 0, 1, 2, and 3 for startup, immediate, delayed, and postponed failure outages [37]. CA is based on a steady-state condition for the system and, therefore, the start-up generator contingencies are not of concern. For delayed and postponed generator outages, the system operators may have sufficient time to perform a unit-commitment, economic dispatch, or scheduled import of power. Therefore, Classes 2 and 3 are also not relevant to the CA of interest. This is again an indicator that the current data collection or presentation practices may need to be improved to better support the needed probabilistic analysis. If outages of a particular class cannot be extracted, a simple approximation can be used to determine the percentage of outages for different classes.

On the other hand, variances for generator outage parameters are small, which indicates the distributions are much less diffuse and have smaller data variability across different regions compared to those for transmission circuits and transformers. A very likely explanation is that generators are located indoors and, therefore, are less susceptible to degradation by environmental conditions, unlike transmission lines. Also, variances for duration times shown in Table 5 are generally large indicating a spread of data points and the significant difference of utilities' maintenance practices. This is also an interesting observation from the data collected in this study.

It should be pointed out that distributions other than the lognormal distribution can be used to capture differences among various data sources. In addition, the population variability can be better accounted for by using a two-stage Bayesian method, as shown in [40]. The hierarchical Bayesian method (HBM) captures the variability in failure rates due to different factors, such as environmental impacts. The generated component failure rates can be further updated if component specific failure data is available. This HBM is, thus, also applicable to the derivation of grid component failure rates.

3.2 An Investigation of Potential Intermittency Induced Outage Modes for Wind Generation

To capture the uncertainties associated with renewables, especially the uncertainties due to the uncontrollable variability of the generation. The probabilistic models and the parameterization of the models for such contingencies need to be developed to provide input data for frequencies and duration times of different outage modes in the contingency analysis. The focus here is on under- or over-generation outages, as well as common mode outages. Generation from renewables is generally dependent on environmental and weather conditions. The variation patterns of these conditions can be very different across regions[10]. Therefore, parameters for renewable outage modes need to be developed for specific areas based on locational data.

[10] This is again a data poolability issue, which is not investigated here since we do not have renewable generation data from other regions.

Studies were performed to extract rampings or ramping events for renewable generation and analyze the statistics of these rampings, e.g., in [9][41][42][43]. These studies focused on a relatively longer-term trend of un-dispatchable renewables and are more suitable for applications such as flexibility studies, unit commitment, or economic dispatch. Therefore, to model renewables in a PCA, the focus of this study is to investigate fast, intermittency induced rampings or ramping events that can be potentially considered as generator outages, as well as the modeling and parameterization of such outage modes that have not been studied. Historical time-series wind generation data are used in this study to extract different types of rampings and perform statistical analyses to obtain input data such that the intermittency induced outages (IIOs) can be included in a PCA.

3.2.1 Outage Modes of Renewable Generation

Modeling of renewable generation in transmission contingency analysis needs to be studied first. For dispersed renewable generation resources embedded in distribution systems, especially roof-top solar panels, although they can be aggregated and lumped, they are usually not modeled in transmission models or contingency analyses since a sudden loss of all these sources is extremely unlikely. Therefore, the focus of the modeling here is on centralized wind farms and/or solar power plants. A single wind farm or solar farm may have sufficient capacity to affect grid operation and can be modeled as a single generator in planning models. A cluster of such renewable generation sources that are geologically close to each other may have even more significant impacts and can be modeled explicitly.

Outage modes need to be carefully identified, as indicated in Chapter 2. For the IIOs of renewable generation, their durations are much shorter than those of conventional generators. In addition to a complete loss of generation, which is always modeled for a conventional generator, under- or over-generation of renewable generators also must be explicitly modeled. This is because the impacts and causes of these outage modes are very different. An over-generation event may only cause some voltage issues, which may be tolerated better in transmission networks, while an under-generation event may introduce frequency issues and a loss of load, which is more severe in terms of stability and economic impact. Explicitly modeling different outage modes will require the collection and analysis of different outage data.

More importantly, for each of these outage modes, there exist associated common mode outages (CMOs) of multiple renewable generators. Renewables largely rely on environmental conditions, such as wind speed and direction, or solar irradiance. In a relatively limited area, such conditions can be highly co-dependent. Therefore, a sudden change in weather conditions can cause similar and concurrent changes in wind/solar farms that are collocated or close to each other. This type of behavior is similar to CMOs of two or more collocated conventional generators, i.e., multiple wind/solar farms can experience an outage in a similar manner concurrently. Therefore, CMOs may also include a loss of generation, or under- or over-generation events.

3.2.2 Probabilistic Outage Models and Parameterization of Renewable Resources

3.2.2.1 Pearson' Correlation for Probabilistic Outage Models

Similar to the practice of obtaining probabilistic parameters for conventional generator outages, historical data for wind and solar power generation and/or weather condition data are needed to analyze the statistics of these outage modes, e.g., how often the reduction in wind generation at a certain level occurs and how long this reduction lasts. This method was used to calculate the frequencies and probabilities of occurrences of different fast solar rampings using solar irradiance data and electrical models of solar plants, as shown in [12]. In [12], fast solar rampings were captured by analyzing high-resolution irradiance measurement data and characterizing the irradiance changes within a certain time period, which was used for calculating the statistics of solar generation reduction. In this study, a collection of wind generation data measured at a utility with a high penetration level of wind is used to demonstrate the process of parameterizing the outage modes.

Before developing the probabilistic models and parameters for IIOs of renewables, one important aspect that needs to be investigated is whether the outage parameters can be characterized by independent random variables. Pearson's correlation is a standard quantitative method used to measure the linear correlation between random variables (see e.g., [44]) and will be used in this study.

The cross-correlation coefficient for realization of two random variables X, i.e., a set of n samples of $X=\{X_i\}$ and Y and $Y=\{Y_i\}$, $i=1,2,\ldots, n$. is given as:

$$R_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} \tag{3.3}$$

where $S_{XX} = \sum(X_i - \overline{X})^2$,$S_{YY} = \sum(Y_i - \overline{Y})^2$, and $S_{XY} = \sum(X_i - \overline{X})(Y_i - \overline{Y})$ with $\overline{X}$ and $\overline{Y}$ the mean values of the samples. Cross-correlation between wind farms has been used in a number of studies mainly for long-term forecasting of wind speed or generation, e.g., [45].

3.2.2.2 Historical Wind Generation Data Cleaning

One and a half years of historical wind generation data (including wind speed data) were obtained from ERCOT for this study. The generation data, which were collected from more than 100 generation sites, have a five-minute time interval. Some of these sites are different phases of one big wind project and obviously can be combined based on site locations, i.e., sites that are co-located in a wind farm. The wind generation data were normalized first with the rated capacities of different wind plants.

The original data consist of both wind speed and power data that were collected at a five-minute interval from a total of 110 sites. The data quality is generally good. However, the continuity of the generation across different generation sites are very different. The major reason is that some wind farms were not built yet when the data collection started, and some wind farms may have been down for maintenance or other reasons at the time. This issue of data discontinuity was addressed by using several indices, i.e., number of missing data, number of continuously unchanged zero data, number of continuously unchanged non-zero data, and the correlation between wind speed and power.

The data can remain constant or unchanged for different time lengths, as shown in Figure 3.1, where a block of unchanged data is considered bad data or anomalies if the block length exceeds a pre-specified value while a block of unchanged data is considered normal if the length of the unchanged data block is smaller than the threshold. For example, Figure 3.1 has two blocks of zeros with lengths of 8 and 4, i.e., 40 and 20 minutes, respectively. If assuming a threshold of 6 (or 30 minutes), only the first block of zeros is considered abnormal while the second block of zeros is good. In our data cleaning, the threshold is pre-specified as 100. In addition, if the number of unchanged zero (or non-zero) data points from a site exceeds a pre-specified value (set to 10,000 in this study), wind speed and wind power data from this site will not be included in the study. Similarly, if the total number of missing wind speed/power data of a site exceeds a pre-specified value (again, set to 10,000 in this study), this wind generation site will be excluded.

The missing data (no data entries) and long sequences of continuously unchanged zeros or non-zeros are considered bad data. Bad data need to be excluded when calculating the correlation between wind speed and power. Then, the sites with a low correlation between wind speed and power can be removed. In this study, a site with wind speed-power correlation values lower than 0.4 is deleted. To see the impact of the correlation value on the data quality, scatter plots of the wind power vs. wind speed are provided for three sites, i.e., site 10 (correlation is 0.93), site 25 (correlation is 0.58), and site 71 (correlation is 0.16.) As it is well-known that the power curve is used to describe the relationship between wind power and wind speed for a single wind turbine. The collections of the scatter plots in Figure 3.2 looks roughly like a typical power curve. The scatter plot in Figure 3.4 looks like a rectangle and is quite different from a typical power curve. Figures 3.2-3.4 show that the higher the correlation value, the closer the scatter shape is to a typical power curve.

In fact, when plotting time series wind speed and wind power data in the same figure for those low-correlation sites, very frequently high wind speeds correspond to low wind power outputs and low wind speeds correspond to high wind power outputs. That is, wind speed data and power output are inconsistent and one or both are probably erroneous. Hence, we only keep the sites having a relatively high correlation

(>0.4) between wind speed and power. After the data cleaning, data from 57 sites were kept and these sites were divided into 28 groups based on the geolocations. The rated capacities of these groups are very different with the maximum and minimum of 820 MW and 40 MW, respectively.

index	data
1	1
2	2.1
3	2.2
4	3
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	2
14	0
15	0
16	0
17	0
18	6.1
19	9
20	9
21	9
22	9
23	9

Figure 3. 1: ***Illustration of continuously unchanged zero and non-zero data***

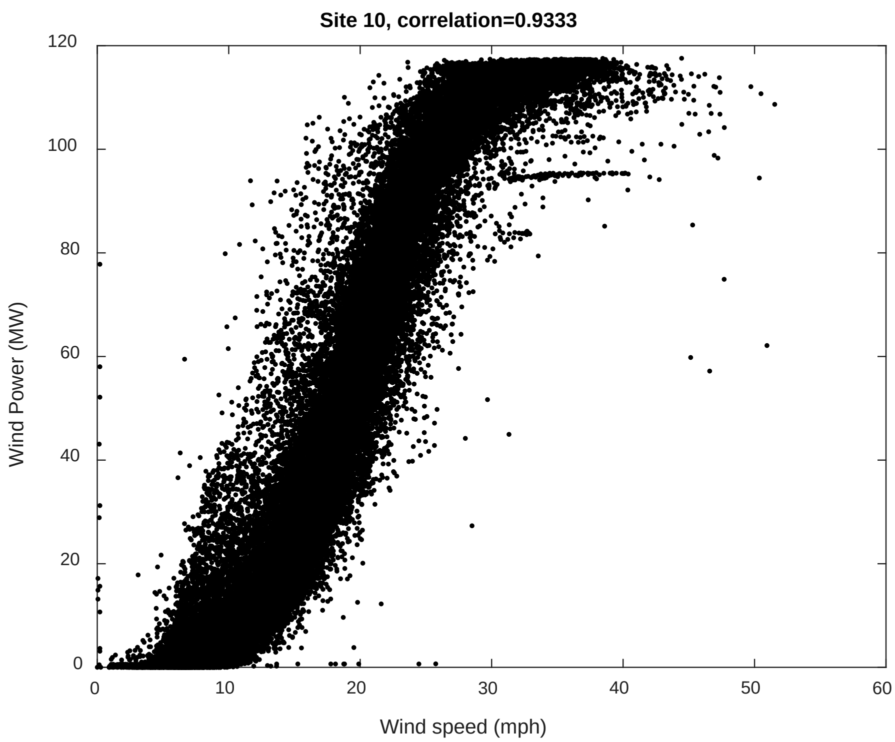

Figure 3. 2: ***Scatter plot of wind power vs. wind speed for site 10.***

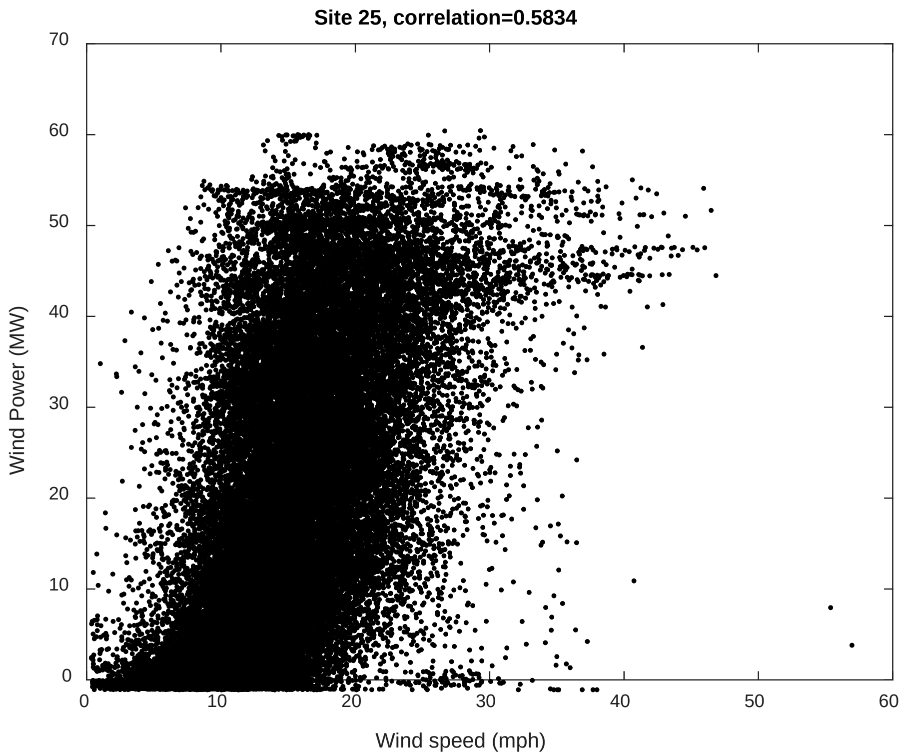

Figure 3. 3: ***Scatter plot of wind power vs. wind speed for site 25.***

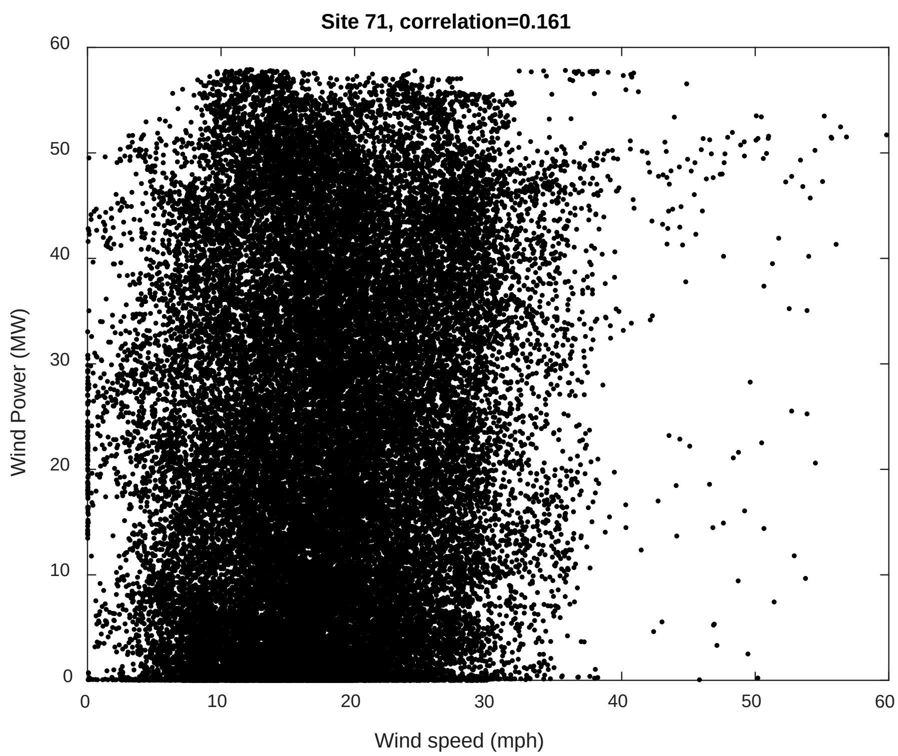

Figure 3. 4: ***Scatter plot of wind power vs. wind speed for site 71.***

3.2.2.3 *Extraction of Ramp-up and -down Ramping Events from Time Series Wind Generation Data*

For frequencies of wind generator outages or fast rampings, it is natural to count the number of outages at a specific wind farm and calculate the annual frequencies and durations of single outages. Therefore, the major task here is to extract the fast ramp-up and ramp-down events in wind farms caused by, e.g., the sudden changes in gust speed/direction of the wind.

Many different methods have been developed to extract ramping events in renewable generation (e.g., see [4–7]). In general, all these methods intended to capture relatively long-term trends in wind generation variation instead of fast ramp-ups and -downs in terms of minutes or tens of minutes that can be considered outages. In some of the methods, a smooth-out of the time-series is needed to ignore those fast rampings. In those studies, different definitions for ramps were used. Usually, these definitions allow some power fluctuation between the start and the end of a ramping. Since the target application of this study is PCA, only fast rampings can qualify as generator outages. Therefore, a different definition for a ramping is provided here considering the interval of the wind generation and speed data, i.e., 5-minute.

For a set of given time series generation data represented by pairs of time and power $\boldsymbol{X}=\{(t_1,p_1),(t_2,p_2),\ldots,(t_N,p_N)\}$, a fast ramping event $T(i,j)=\{(t_i,p_i),\ldots,(t_j,p_j)\}$, for i, j in $[1, N]$ is defined as a subsequence of $\boldsymbol{X}$ that monotonically and continuously increases (for an up ramp) or decreases (for a down ramp) in the generation, i.e., $p_l > p_{l-1}$ (for an up ramp) or $p_l < p_{l-1}$ (for a down ramp) for l in $[i, j]$. The requirement for monotonic change is because the interval for the wind generation data is five minutes. Since the generation at the end of the ramping may become flat for some time before it changes to another level, this time period is considered the duration of the ramping, similar to the outage duration of a conventional generator. Therefore, an IIO consists of a segment representing the fast ramping $T(i, j)$ and another segment indicated by $D(j+1, k)=\{(t_{j+1}, p_{j+1}),\ldots,(t_k, p_k)\}$, where $p_l \approx p_j$ for l in $[j+1, k]$, and this outage is denoted here by $O(i, j, k)$ for $i < j < k$.

Based on the definition of fast ramping events, since the time interval for the available wind generation data is 5-minute, the fast ramping events here usually only take tens of minutes to finish. The following criteria were developed and used to extract outages $O(i, j, k)$:

1) A monotonic decrease or increase in the wind generation for ramp-downs and –ups, i.e., $p_l<p_{l-1}$ for l in $[i, j]$;
2) A minimum variation of δ in the wind generation when the rampings are completed, i.e., $dev=|p_j - p_i|\geq \delta$; and
3) After the wind generation reaches the highest point (for a ramp-up) or the lowest point (for a ramp-down), the duration of such a ramping is defined as the period in which the generation level is maintained within its $\pm\varepsilon$ range of the highest or lowest generation level, i.e., $p_l \leq p_j(1\pm\varepsilon)$ for l in $[j+1, k]$.

A scoring scheme was developed to extract fast rampings more efficiently. A subsequence must score 10 points to be considered a fast ramping. The scores are assigned based on step-wise changes in generation, e.g., 10 points for a deviation of 20% from the previous generation level in two consecutive generation entries and only 1 point for a deviation of 2% in two steps. In addition, any ramping, if it does not earn 10 points within 30 minutes will be discarded since the ramping is considered too slow. Also, a ramping process that lasts for more than 30 minutes will be discarded.

It should be pointed out that, although an outage is defined to consist of a ramping segment and the duration segment, rampings and outages are used interchangeably below for simplicity. These criteria can be used to study the rampings that have relatively significant impacts on contingency analyses. Such criteria have been implemented in Matlab scripts. The time sequence wind generation data from each wind farm will be examined against these criteria. Rampings that satisfy these criteria will be extracted and analyzed later.

Each of the wind farms experienced a different number of such rampings, e.g., 417 such rampings were identified in this process for one of the wind farms. Figure 3.5 shows two example rampings that lead to significant under- and over-generation within 20 minutes at this specific wind farm.

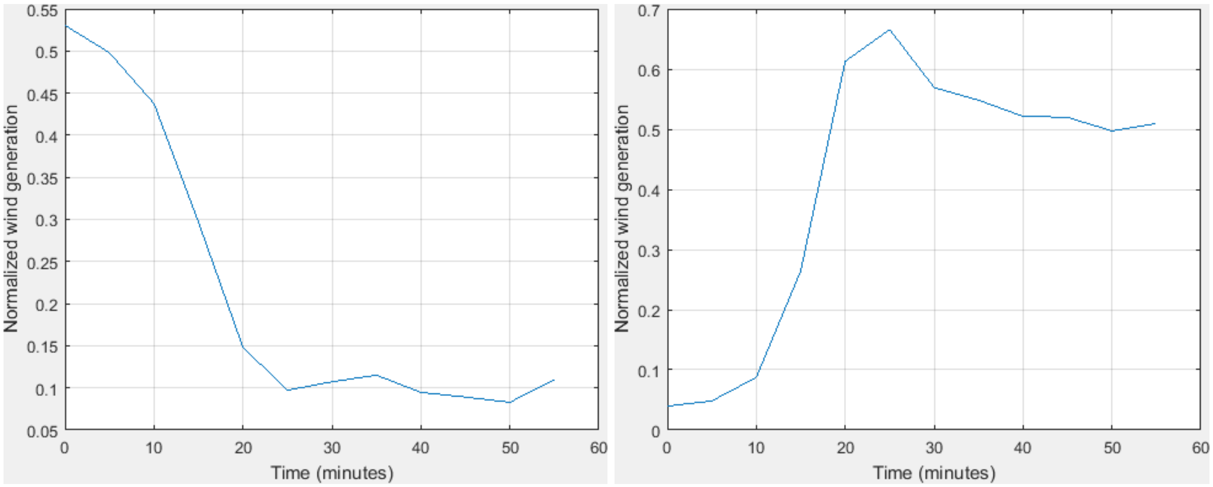

Figure 3. 5: **Example under- and over generation modes of a single generation site.**

Inspection of the rampings in Figure 3.5 shows that a fast ramping or IIO can be characterized by several quantities, such as the initial generation level (Gen_{init}) prior to the ramping, the deviation of generation (dev) from the generation at the beginning of the ramping, and the duration of the ramping $(t_{duration})$. Note that the duration does not represent the time it takes the ramping to reach the lowest or highest generation point. It simply means how long the generation stays at the post ramping level. For an extracted outage $O(i, j, k)$, we can easily calculate these quantities:

$$Gen_{init} = p_i$$

$$dev = |p_j - p_i|$$

$$t_{duration} = k - j$$

There are a large number of rampings and all of them will impact the grid operation to some extent. However, not every ramping event qualifies as an outage if the generation changes are too small and/or too slow. By changing parameters δ and ε, rampings with different potential impacts can be extracted. The extracted outages can always be further screened by, e.g., limiting the values of $(j - i)$ to select those rampings, i.e., taking less time to complete. In addition, the criteria can also be modified to accommodate new rules, e.g., power fluctuation during the ramping, by changing the first criterion above.

3.2.3 Parameters for Outage Models of Wind Generation Sites

The probabilistic models and the parameterization of the models for IIOs were developed to provide input data to fit the PCA framework. In a PCA, the input data needed for individual generator outages are the frequencies of the outage occurrences and durations of the outages. Since renewables are still modeled as generators, the parameters for renewable outages should also include frequencies, i.e., ramp-up and -down outages or rampings, and durations, i.e., the length of time the wind plant stays at the post-outage generation level.

As shown in Figure 3.5, the variation (dev) of the generation level during a ramping is not fixed and these variations are expected to differ from each other. This is a unique and yet important parameter for a renewable outage that measures the severity of a ramping. Generally speaking, one of the failure modes of a conventional generator can be de-rated generation, i.e., the generation level can only be sustained at a lower level due to a partial failure of the generator. However, this de-rated generation mode is rarely modeled, partially due to a lack of data. This variation, however, is very common for renewables and therefore, important for modeling the IIOs of wind plants, especially for the common mode outages of multiple wind pants. Another important parameter is the generation level prior to the rampings or the initial power generation level (Gen_{init}), which is needed for the initial power flow calculation in contingency analysis.

Derivation of these parameters will be based on extracted ramping events from historical data, as discussed in Chapter 2.

To model the single contingencies or outages of a single wind farm, the following parameters will be needed for the PCA:

(1) the frequency (f) of different types of outages, i.e., fast ramp-ups and –downs;
(2) the duration time ($t_{duration}$) of such outages, i.e., the time the wind plant stays at the post-outage generation level; and
(3) the variation of the generation (dev) from a relatively steady generation level of a wind farm due to the intermittency.

The first two parameters are similar to the parameters for conventional generators while the third one can be used to model de-rated generation for a conventional generator but is rarely used. This de-rated generation or, more precisely, variation of generation level, is much more common and therefore, more important for the intermittency induced outages of a wind plant.

These parameters enable the IIOs of renewable generation to fall naturally into the existing PCA framework and the tool and need to be carefully derived for the PCA from those fast ramping events retrieved from historical renewable generation data, as discussed below.

It should be pointed out that, because renewable generation is generally dependent on environmental and weather conditions, the variation patterns of these conditions can be very different across regions. Therefore, parameters for renewable outage modes for a specific region may not be applied to other regions without further analysis, which is an important factor to consider in PCA applications.

3.2.3.1 Analyses of All Individual Rampings in Wind Generation

After applying the previous criteria, i.e., δ=30% and ε=30%, a total of 7,750 rampings (both ramp-up and down) were identified for all 53 wind plants. The numbers of ramp-ups and downs are very close to each other, i.e., 3,954 vs. 3,796. In addition, the numbers for individual wind plants, the ramp-ups and –downs are also close to each other.

A preliminary analysis of frequencies for all of the extracted outages is presented here. The average frequencies for individual wind plants can be calculated easily. However, it is more useful to characterize the variability of rampings at different generation sites such that the probabilistic model can be used for a variety of wind plants. This is very important in transmission planning with wind generation that is being planned for installation across the system. This can be done by deriving the distribution (probability density function or PDF) of ramping frequencies across different wind farms.

Figure 3.6 shows an empirical distribution that was fitted using kernel density estimation based on the histogram of ramping frequencies across different generation sites. The mean value and the standard deviation are also shown in Figure 3.2.

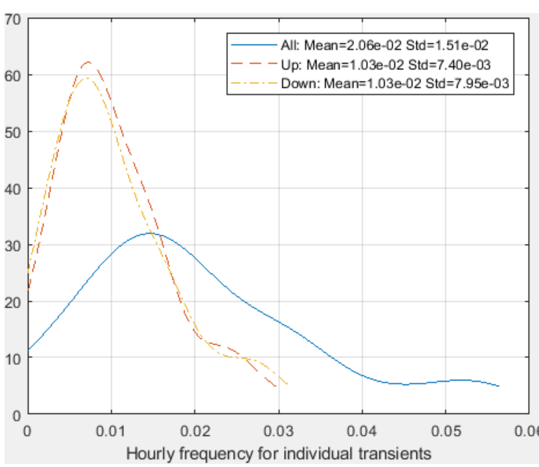

*Figure 3. 6: **Frequency distributions for individual rampings.***

Figure 3.6 shows the empirical hourly frequency distributions for ramp-up, –down, and all individual rampings. The frequency distribution of total rampings or outages is indicated by the solid line while the distributions for up ramps and down ramps are plotted in dashed and dashed-dotted lines in Figure 3.6. The mean value of the total rampings is the sum of the mean values of up ramps and down ramps. In addition, the shapes of the distribution and the mean values and standard deviations are almost identical in Figure 3.6, which indicates that distributions for both up and down rampings can be considered the same and modeled using the same distribution. The frequency distributions are plotted across different wind generation sites for this data source. If frequency information from other utilities can be obtained, we can develop a frequency distribution across different regions.

3.2.3.2 Analyses of Single Rampings in Wind Generation

The frequency distributions shown in Figure 3.6 are for all of the individual outages or rampings. However, some of these outages occur concurrently, e.g., rampings from two, three, or more wind farms may occur at about the same time due to a strong correlation between them, as discussed in Section II. Such concurrent rampings can be considered CMOs and need to be studied separately to model them. Theoretically, for M wind farms, there can be concurrent rampings from all these M wind farms, then concurrent rampings from every $(M-1)$ wind farms, and all the way to every three and two wind farms (denoted as triple and double rampings in this study). After excluding all these concurrent rampings from the individual rampings, the remaining rampings are the single ones.

However, almost no concurrent rampings at four wind farms have been observed from the historical data. Therefore, the single rampings are obtained by removing triple rampings first, followed by double rampings. The total number of single rampings is 2,995. The empirical distributions of outage frequencies across different wind generation sites are shown in Figure 3.7. Less than half of the total rampings are single rampings, as also seen from the mean values in Figure 3.7. The shapes, mean values, and the distributions for ramp-ups and –downs are slightly different from each other, but still very similar to each other.

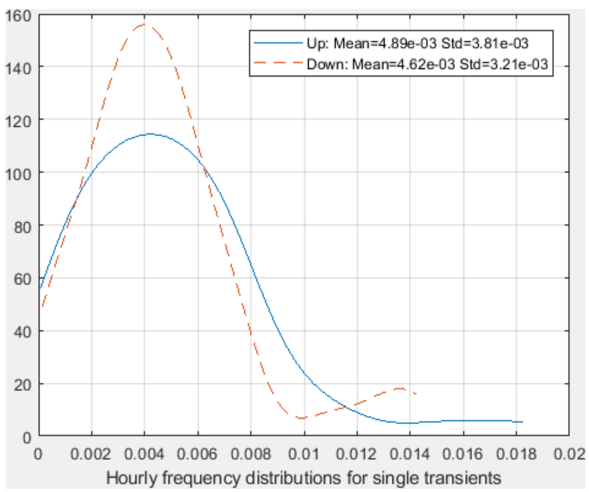

*Figure 3. 7: **Frequency distributions for single rampings.***

As indicated above, a wind farm outage cannot be solely described by the outage frequency or rate and the duration. The changes of the generation level in different rampings or outages are different and the initial generation level prior to the outage occurrence also needs to be known to understand the impacts of such single rampings. Therefore, three other quantities that characterize the under- or over-generation curves, i.e., the initial generation levels, the percentage variations of generation, and the durations of the rampings, need to be captured in the outage model.

Unlike frequencies that are plant specific, duration $t_{duration}$, generation level prior to the ramping Gen_{init}, and the generation variation *dev* are all ramping specific. While information for these three quantities is contained in the extracted outages, it is important to know, before developing the probabilistic distributions for these parameters, whether these three parameters are correlated, i.e., whether a joint distribution is needed to model the properties of these three quantities. The parameterization for the joint distributions will become more complicated if the three quantities are strongly correlated.

A scatter diagram of the generation levels Gen_{init} prior to the rampings and the absolute values of maximum deviations *dev* of the generation in the rampings, together with their histograms, can be used to provide a rough idea of their correlations, as shown in Figure 3.8. The scatter diagram shows that, within the "M" shaped area, the data pairs for Gen_{init} and dev (the absolute value of the differences between the initial generation level and final level) are scattered fairly evenly. Note that no data appear outside of the "M"-shaped area; this will be explained below.

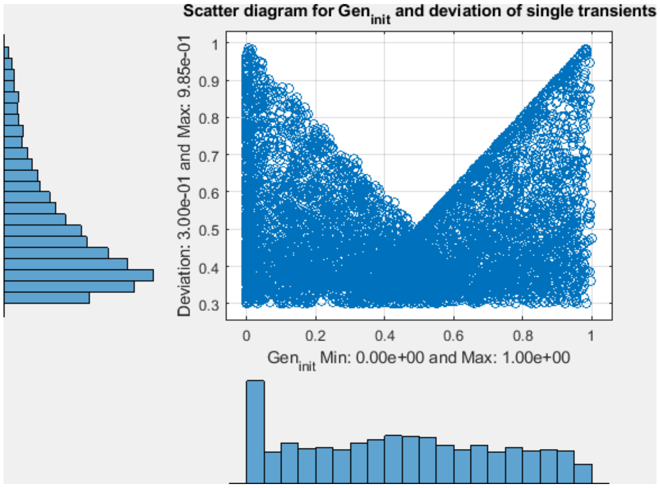

*Figure 3. 8: **Scatter diagram for initial generation level and deviation.***

The formulation of the "M"-shaped area is due to the criterion introduced to extract the rampings, i.e., the minimum variation of 30%, i.e., $dev \geq 0.3$ for both ramp-ups and -downs. Therefore, for the ramp-ups, the maximum Gen_{init} is 70% and for the ramp-downs, the minimum Gen_{init} is 30%, which is clear if we plot such scatter diagrams for up-ramps and down-ramps, respectively, as shown in Figures 3.9 and 3.10.

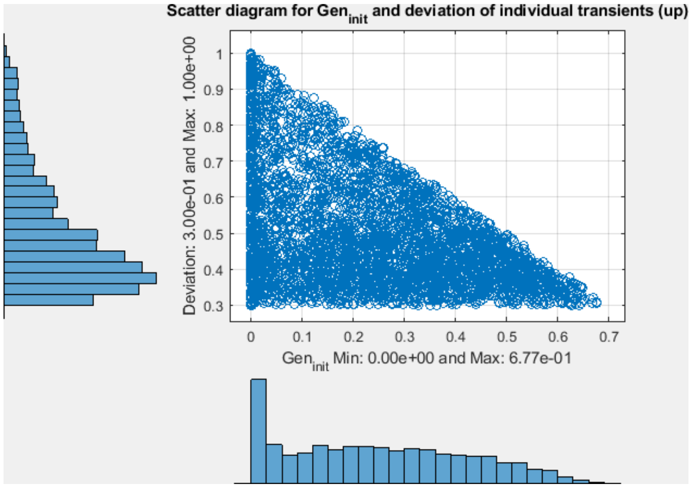

*Figure 3. 9: **Scatter diagram for initial generation level and deviation of single ramp-ups.***

Based on the criteria developed in Section 3.2.2.3, the constraints that formulated the "M"-shaped area in Figure 3.8 are given as:

$$Gen_{init} + dev \leq 1.0$$
$$Gen_{init} - dev \geq 0.0 \tag{3.4}$$

for a ramp-up and a ramp-down with $dev \leq 0.3$, respectively. Thus, the fact that the data pairs fall into the "M"-shaped area does not necessarily indicate a correlation between the two quantities. The scatter diagrams with durations are not shown here but do not indicate a strong correlation either.

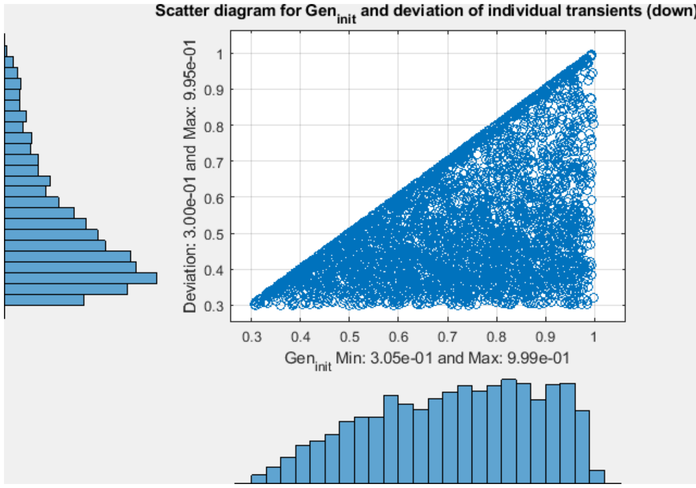

Figure 3. 10: ***Scatter diagram for initial generation level and deviation of single ramp-downs.***

This can be verified by calculating the correlation of these three quantities for all rampings, which are shown in Table 3.10. The small values of the off-diagonal elements in Table I indicate the very weak correlations between these quantities. Therefore, $t_{duration}$, Gen_{init}, and dev can be modeled separately, i.e., using independent distributions.

Table 3. 6: ***Correlation of Gen_{init}, dev, and $t_{duration}$.***

	Initial Generation Level (Gen_{init})	Variation of Generation (dev)	Ramping Times ($t_{duration}$)
Initial Generation Level (Gen_{init})	1.00e+00	-2.28e-02	-9.84e-03
Variation of Generation (dev)	-2.28e-02	1.00e+00	1.32e-01
Duration ($t_{duration}$)	-9.84e-03	1.32e-01	1.00e+00

The distributions of initial generation levels for both ramp-ups and -downs across the 53 generation sites are shown in Figure 3.11. The Gen_{init}, of all outages at each generation site were averaged for all up ramps and down ramps separately. The ranges of the initial generation levels for up and down rampings are different, as expected. Therefore, it is clear that the ramp-ups and -downs need to be modeled separately, e.g., by using the fitted distributions in Figure 3.11.

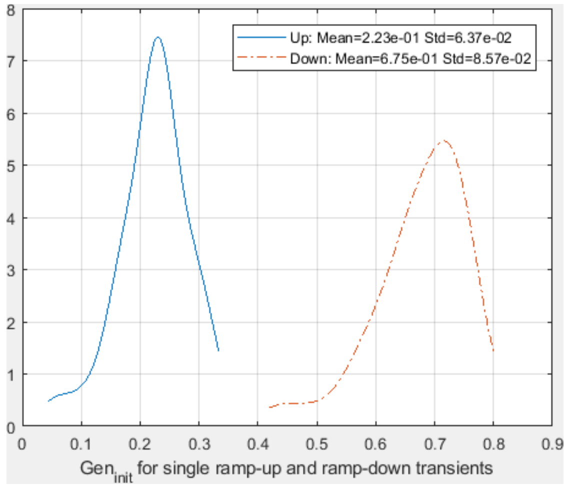

Figure 3. 11: ***Distributions of initial generation levels Gen_{init} for single rampings.***

The fitted distributions for deviations of ramp-ups and –downs across generation sites are shown in Figure 3.12. In Figure 3.12, the distributions of deviations for ramp-ups and -downs are similar, and mean values and standard deviations are both very close to each other; therefore, they can be modeled using the same distribution.

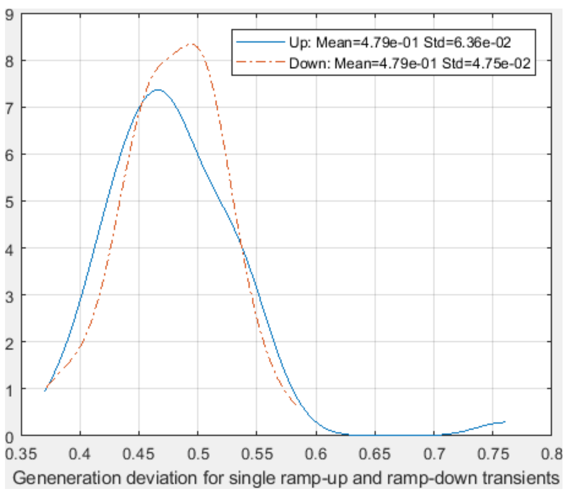

Figure 3. 12: ***Distributions of deviation dev for single rampings.***

The distributions of duration hours for all single ramp-up and -down outages across generation sites are shown in Figure 3.13. The distributions were fitted for average duration hours at individual generation sites. Figure 3.13 shows that the duration distributions for ramp-ups and -downs still appear similar to each other. In addition, the durations for most of the rampings are fairly short, i.e., around one hour. However, this is mainly because of the average effect of duration hours at each generation site. For some of the outages, the duration hours can be very long, e.g., even 10s of hours.

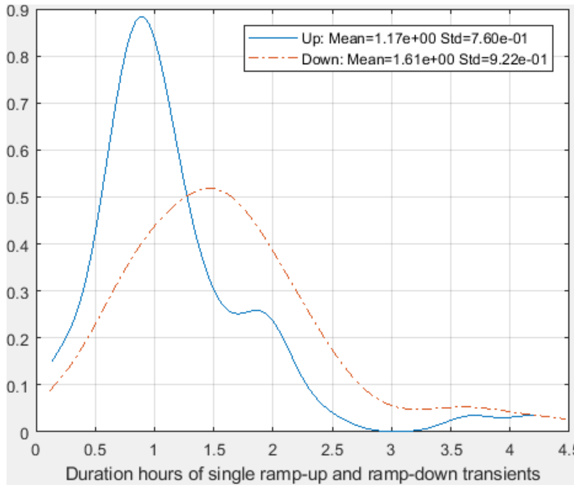

Figure 3. 13: ***Distributions of duration $t_{duration}$ for single rampings.***

Different thresholds can be selected during the extraction of fast rampings. The extraction process can be repeated to confirm the above statements by selecting δ=50% and ε remains 30%. This implies that only rampings with larger generation deviations will be extracted and the total number of rampings should be smaller. A total of 1,338 rampings (both ramp-up and down) were identified for all 28 wind plants. The numbers of ramp-ups and downs are 739 and 599, respectively. Again, the numbers for individual wind plant ramp-ups and –downs are close to each other. The frequency distributions across the wind generation sites for the up- and down-ramps are shown in Figure 3.14. Similarly, the scatter diagrams for initial generation levels against deviations are shown in Figures 3.15 – 3.17, where the difference due to the selection of δ=50% is obvious compared to Figures 3.8 – 3.10. The Pearson correlation coefficients for the initial generation levels, deviations, and durations are presented in Table 3.7. Again, the weak correlations among them are confirmed.

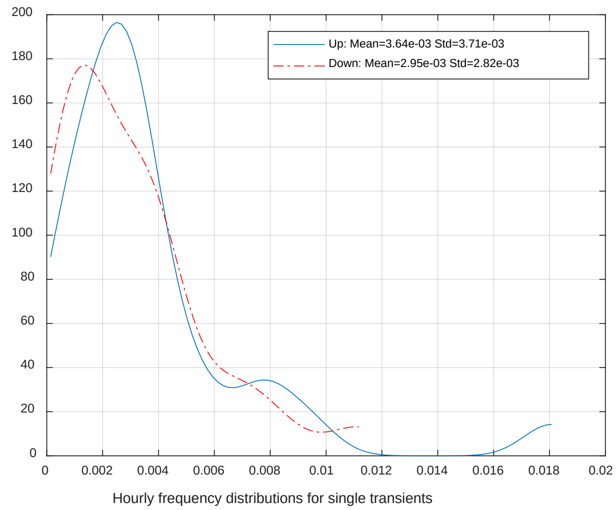

Figure 3. 14: ***Frequency distributions for single rampings.***

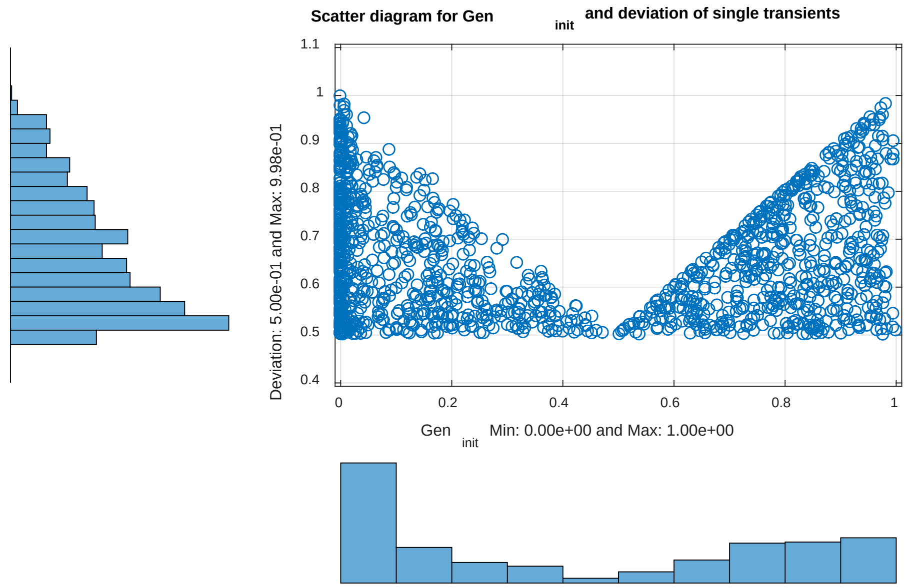

Figure 3. 15: ***Scatter diagram for initial generation level and deviation.***

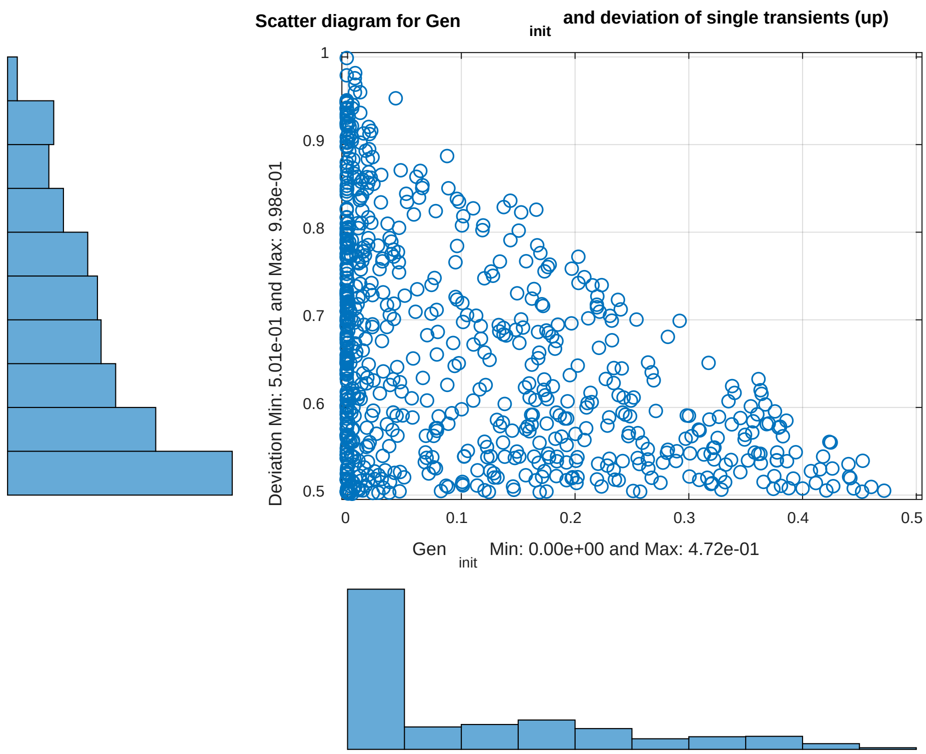

Figure 3. 16: ***Scatter diagram for initial generation level and deviation of single ramp-ups.***

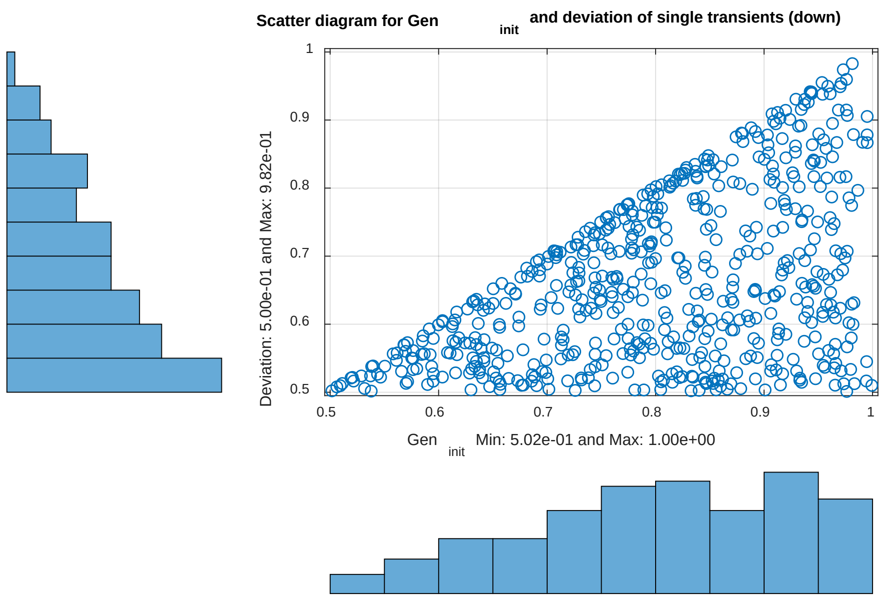

Figure 3. 17: ***Scatter diagram for initial generation level and deviation of single ramp-downs.***

Table 3. 7: ***Correlation of Gen_{init}, dev, and $t_{duration}$.***

	Initial Generation Level (Gen_{init})	Variation of Generation (dev)	Ramping Times ($t_{duration}$)
Initial Generation Level (Gen_{init})	1.00e+00	-1.78e-02	-1.38e-01
Variation of Generation (dev)	-1.78e-02	1.00e+00	1.17e-01
Duration ($t_{duration}$)	-1.38e-01	1.17e-01	1.00e+00

The distributions of initial generation levels for both ramp-ups and -downs across the 28 generation sites are shown in Figure 3.18.

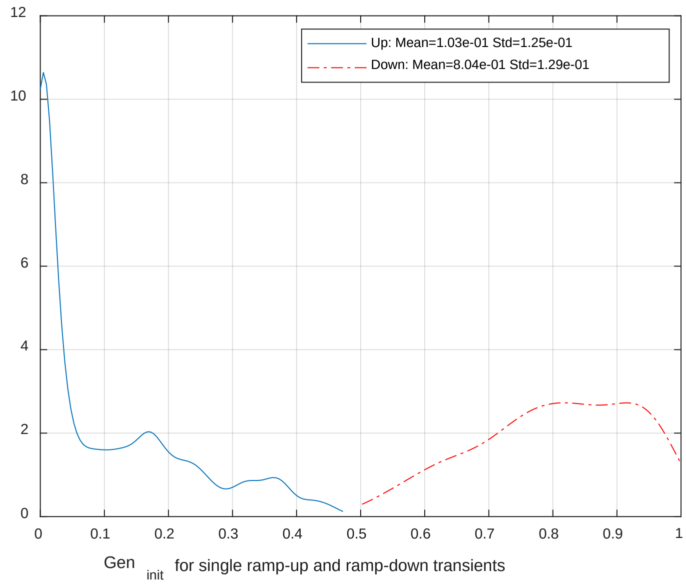

Figure 3. 18: ***Distributions of initial generation levels Gen_{init} for single rampings.***

The fitted distributions for deviations of ramp-ups and –downs across generation sites are shown in Figure 3.19. As shown in Figure 3.19, the distributions of deviations for ramp-ups and -downs are similar, and mean values and standard deviations are both very close to each other; therefore, they can be modeled using the same distribution.

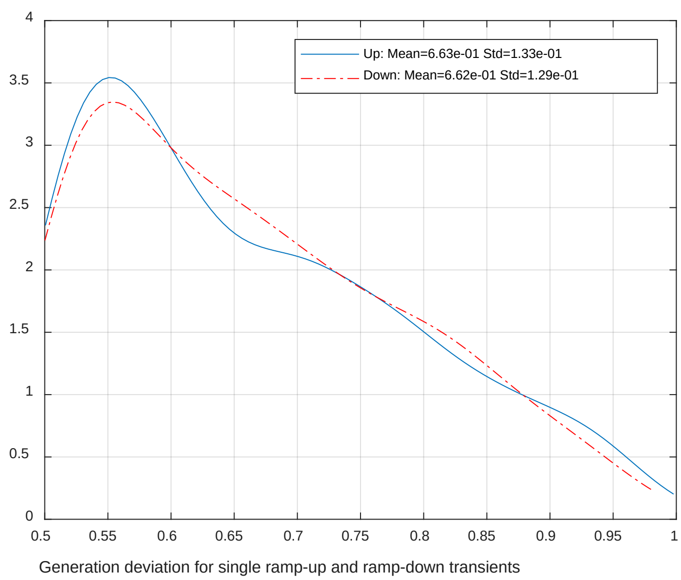

Figure 3. 19: ***Distributions of deviation dev for single rampings.***

The distributions of duration hours for all single ramp-up and -down outages across generation sites are shown in Figure 3.20. The distributions were fitted for average duration hours at individual generation sites. Figure 3.20 shows that the duration distributions for ramp-ups and -downs still appear similar to each other. In addition, the durations for most of the rampings are very short, i.e., less than one hour.

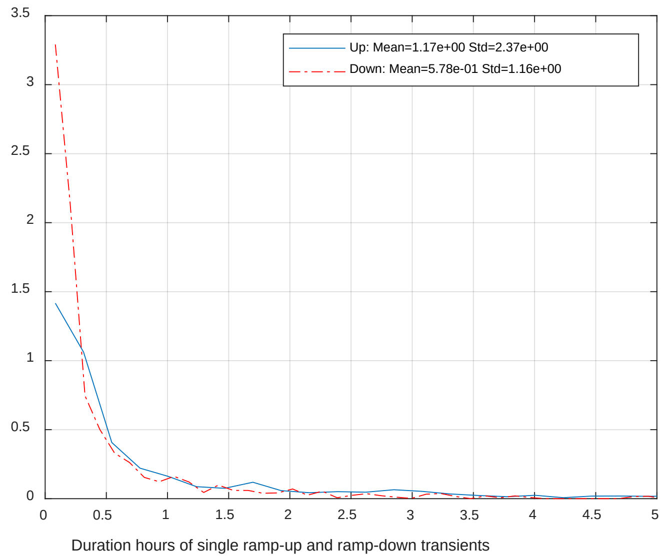

Figure 3. 20: ***Distributions of duration $t_{duration}$ for single rampings.***

Results in Figures 3-12 to 3.20 are based on ramping events extracted using different thresholds. These results confirm the major conclusions drawn from results of original thresholds, i.e., (1) the frequencies for up- and down-ramps are similar to each other and can be modeled using the same distribution; (2) the initial generation, deviation, and duration are very weakly correlated and, therefore, can be modeled using independent random variables.

3.2.3.3 Parameters for Common Mode Outage Models of Multiple Wind Generation Sites

A single ramping or IIO can have limited negative impacts on the grid since the generation deviation is limited by the capacity of the wind farm, even for a threshold of 50% deviation in extracting the ramping events. However, the concurrent rampings or IIOs at multiple wind farms may cause much larger generation variation and have much more severe impacts, although such concurrent IIOs are usually less frequent than the single rampings. However, it is reasonable to postulate that the likelihood of having such concurrent IIOs can be large if these multiple wind farms are highly correlated because correlation can measure the tendency to change in the same direction. Extra attention must be paid to these scenarios.

The output of two wind generation sites may be highly correlated because they are very close to each other or located along the same wind path. In these situations, the wind speed and direction are very uniform across the two generation sites. Also, the output of the other site tends to change in the same direction (for a positive correlation) or the reverse direction (for a negative correlation that is possible but very unlikely for wind farms) from the first output changes. Therefore, a strong correlation between two wind farms may lead to ramping-induced common mode outages or CMOs, i.e., the generation at two or more wind sites or plants increase or decrease at approximately the same time. CMOs are usually less frequent than single rampings but the combinational effect of a simultaneous increase or decrease in generation of multiple wind farms can be very severe. CMOs usually dominate the results in probabilistic risk analyses and, therefore, should be carefully captured, as well.

Two examples of generation reductions at two and three highly correlated sites (i.e., correlation efficient larger or equal to 0.8 in this study), respectively, are shown in Figure 3.21. The wind generation at different wind farms drops similarly in terms of both times and magnitudes. Note that it is (and should be) very rare that the beginning and end times of these concurrent rampings are exactly the same, even for highly correlated sites (unless the correlation coefficient is 1.0, as calculated using Equation 3.3, which is impossible in reality), because of the different physical locations of wind farms.

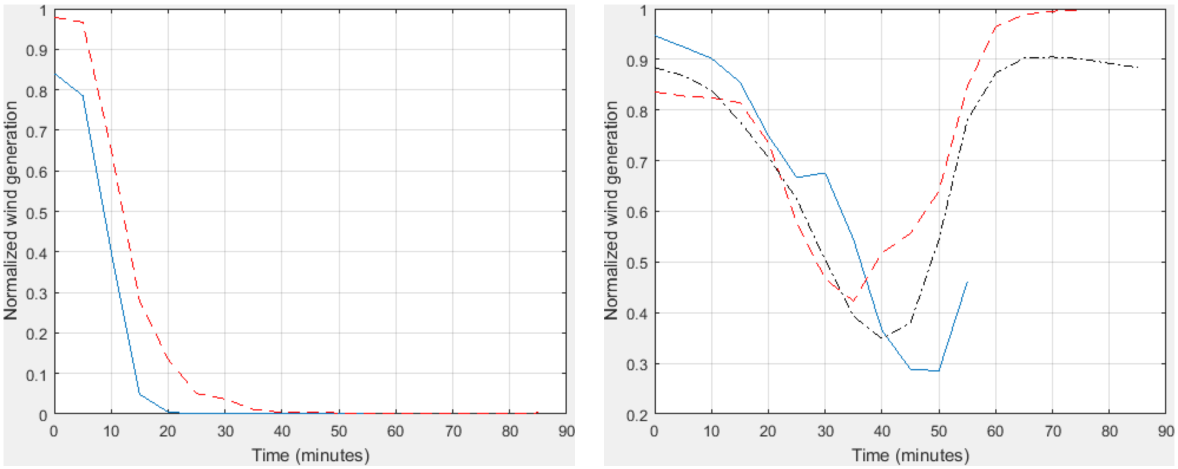

Figure 3. 21: **Concurrent under-generation modes among two and three generation sites.**

A procedure was developed to extract high order rampings. The most important criterion is that the starting time and ending time for the same type of rampings (up or down) should not differ too much from each other. In this study, it was decided that a 20-minute difference or less among rampings from two or three wind farms in the beginning and end times is acceptable. If the time difference is too large, the two outages should not be considered a concurrent mode. If two outages are correlated, the error introduced by combining them is not very severe and, therefore, can be ignored. For the same reason, unless the correlation coefficient is 1.0, the output variations from two generation sites cannot be exactly the same. It is anticipated, however, that the stronger the correlation is, the more frequent the concurrent changes at two wind farms can occur.

To model the CMOs, the same set of parameters is needed, i.e., the frequencies, duration times, and variation during the rampings. It is, therefore, reasonable to extract concurrent over- or under-generation modes from two sites with different levels of correlation from historical data and estimate the frequencies and other parameters of such concurrent modes. Such a process was repeated exhaustively for every pair of wind farms. This result can be used to provide parameters for all CMOs between every pair of wind farms explicitly in the contingency list, similar to the modeling of single outages at individual wind farms. In order to extract parameters for single rampings and CMOs, the single rampings and CMOs need to be exclusive. The issue of possibly double-counting the same rampings is addressed by extracting CMOs or high-order rampings first, i.e., the rampings involved in a triple concurrent ramping will be removed from candidate rampings for extracting double rampings and so on. Since quadruple concurrent rampings have not been observed from the historical wind generation data, the highest order CMOs considered in this study are triple concurrent rampings. By doing so, the concurrent ramping extraction process will also become tractable. In the following discussion, we present results for double rampings first.

Parameters for CMOs: Double Concurrent Rampings

A total of 293 double rampings were extracted with 166 ramp-ups and 127 ramp-downs (δ=50%). According to the definition of a correlation, it is reasonable to postulate that a larger number of concurrent rampings or CMOs between two wind farms should be observed for a stronger correlation. A comparison is performed here by extracting concurrent rampings from historical wind generation data and categorizing them according to different correlation values. The numbers of double concurrent rampings for different correlation values are shown in Tables 3.8 and 3.9 (for δ=30% and δ=50%), respectively. This clearly indicates that the occurrences of concurrent modes involving two wind farms are generally (except for the numbers of rampings for correlation values in [0.8,1.0] and [0.6,0.8]) proportional to the correlation values, and the relationship between the percentage of common modes and the correlation values is consistent. It should also be noted that, even for very weak correlation, i.e., [0.0, 0.2], still a small number of concurrent modes can be observed. It is believed that this is very likely due to coincidence. Table 3.8 also shows the average point-wise distance between each pair of the double rampings, which is used to measure how close the two rampings are. Again, the general trend is that a small distance corresponds to a higher correlation value (except for the last row in the table, which might not be meaningful because of the small number of rampings in this category). Similar observations can be made from Table 3.9.

Table 3. 8: ***Correlation vs. concurrent modes between two wind generation sites (*** *δ=30%)*

Pearson's Correlation between wind farms (*i*, *j*)	**Numbers of simultaneous under- or over-generation between *i* and *j***	**Average point-wise distance between two rampings**
0.8 – 1.0	487	0.25
0.6 – 0.8	516	0.31
0.4 – 0.6	459	0.31
0.2 – 0.4	236	0.37
0.0 – 0.2	53	0.27

Table 3. 9: ***Correlation vs. concurrent modes between two wind generation sites (*** *δ=50%)*

Pearson's Correlation between wind farms (*i*, *j*)	**Numbers of simultaneous under- or over-generation between *i* and *j***	**Average point-wise distance between two rampings**
0.8 – 1.0	102	0.14
0.6 – 0.8	149	0.15
0.4 – 0.6	24	0.19
0.2 – 0.4	14	0.17
0.0 – 0.2	4	0.14

Distributions of frequencies across the wind generation sites, generation levels prior to the rampings, deviations, and durations for rampings are shown Figures 3.22 – 3.25 for any two wind sites with a correlation between 0.8 and 1.0. For correlations between 0.6 and 0.8, the same type of plots are shown in Figures 3.26 – 3.29.

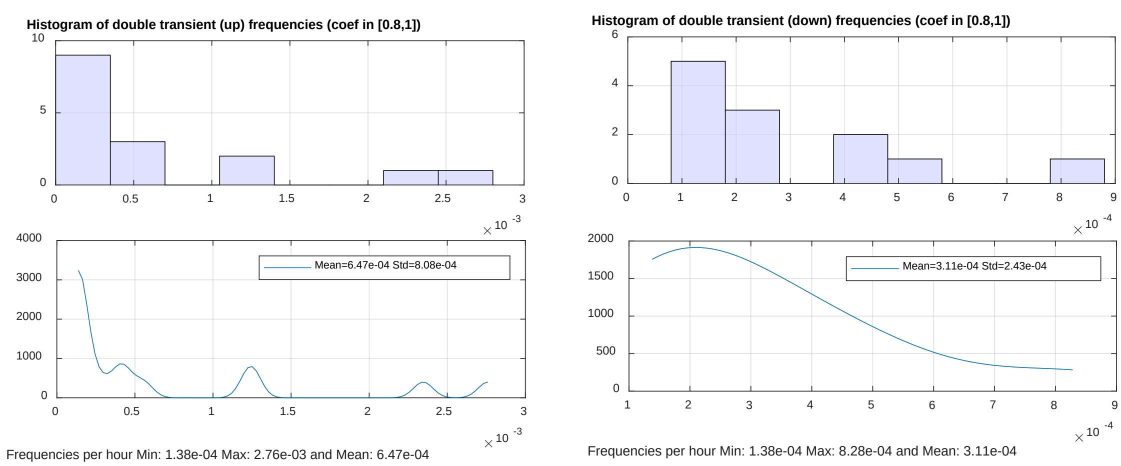

Figure 3. 22: ***Frequency distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0]).***

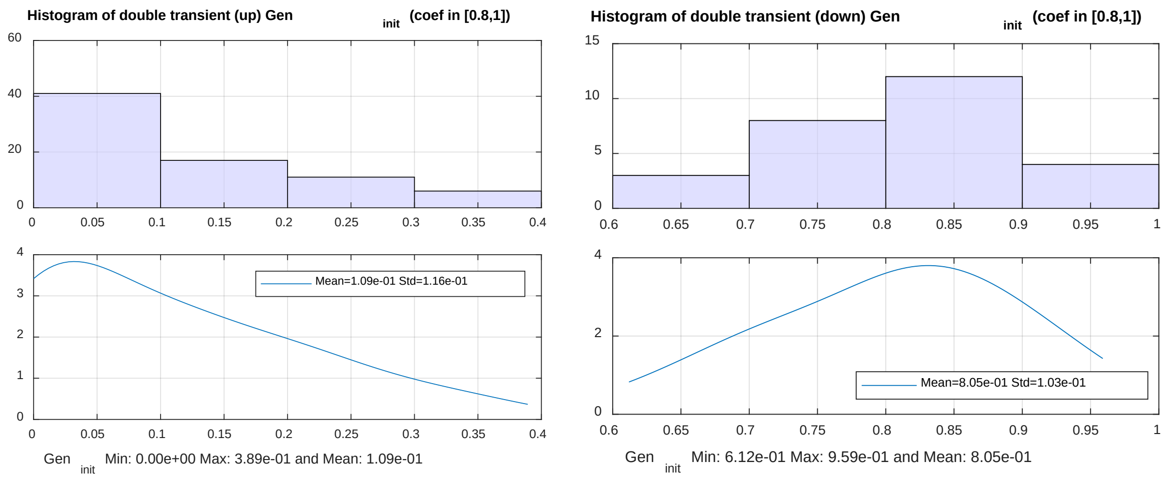

Figure 3. 23: ***Initial Generation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0]).***

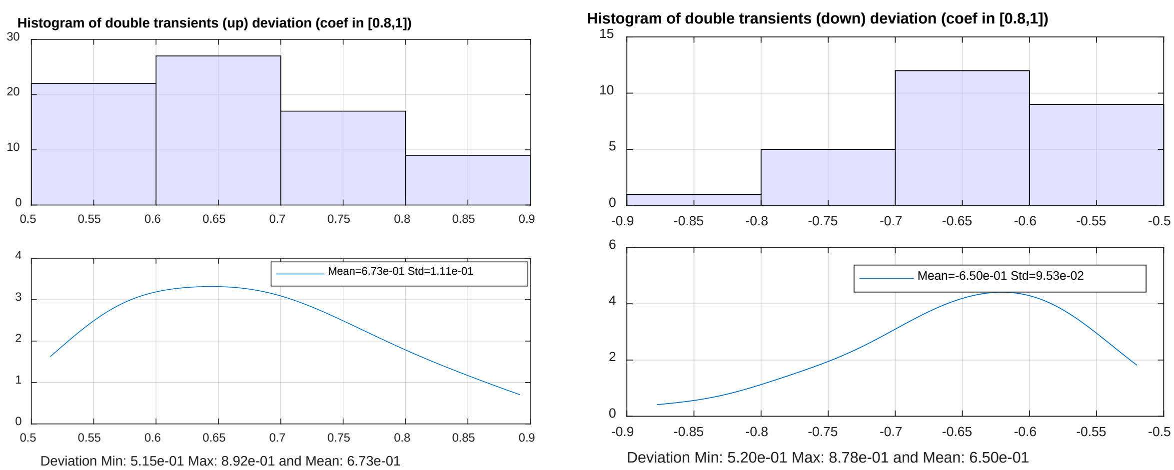

Figure 3. 24: ***Generation deviation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0]).***

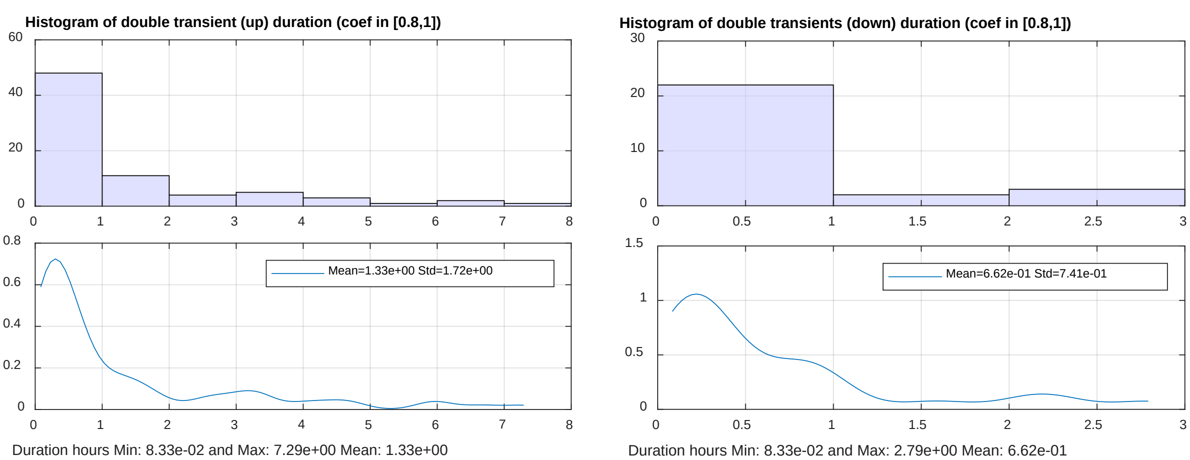

Figure 3. 25: ***Duration distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.8,1.0]).***

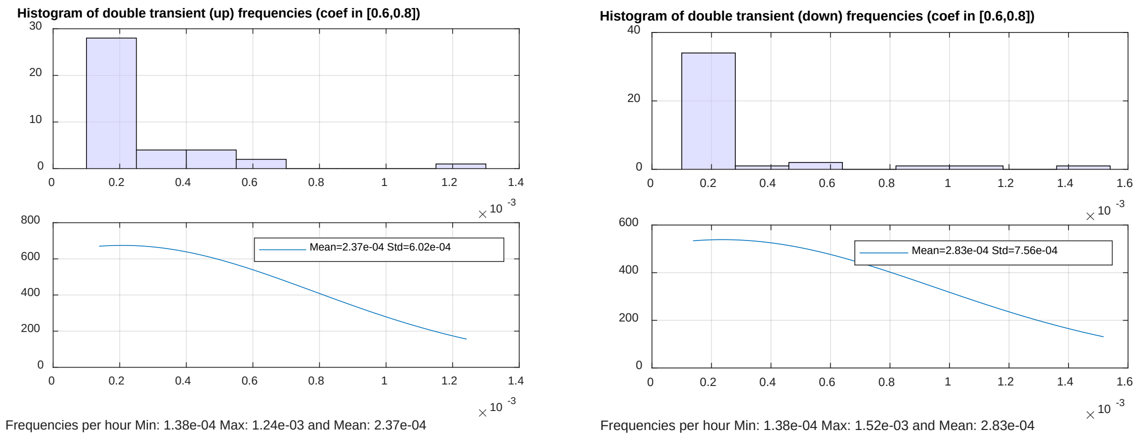

*Figure 3. 26****: Frequency distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8]).***

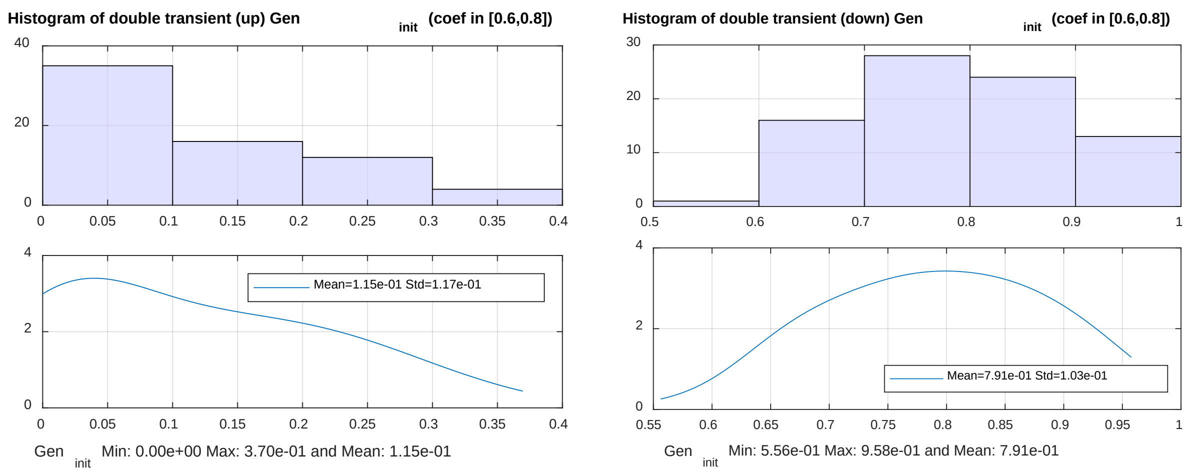

*Figure 3. 27****: Initial generation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8]).***

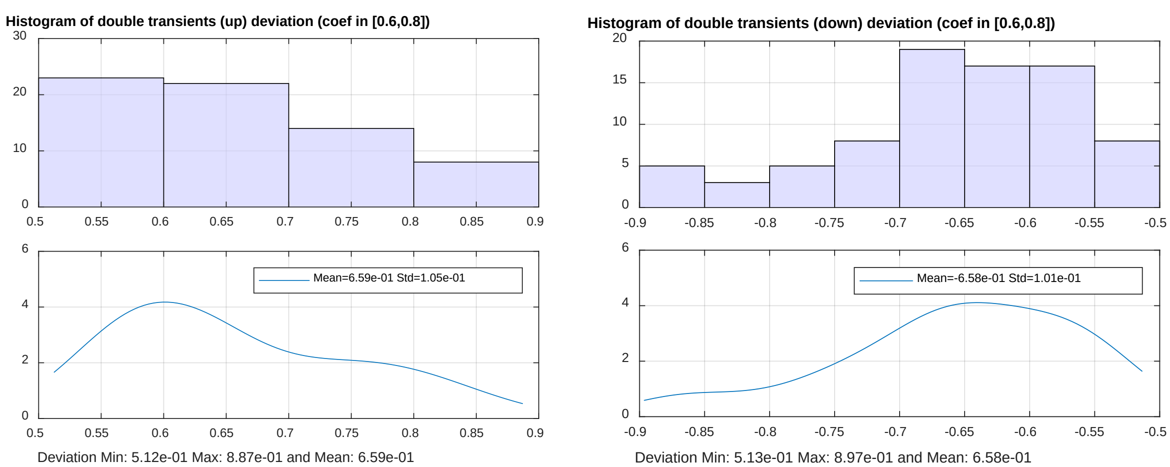

*Figure 3. 28****: Generation deviation distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8]).***

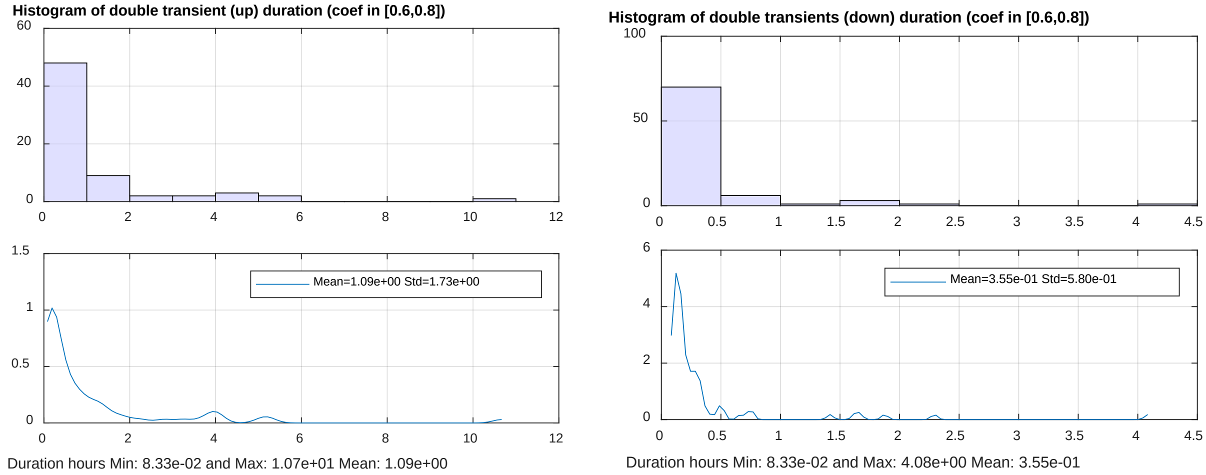

Figure 3. 29: **Duration distributions of single ramp-ups (left) and -downs (right) (Corr. in [0.6,0.8]).**

Parameters for CMOs: Triple Concurrent Rampings

Triple concurrent rampings can be extracted similarly. To simplify the problem, it is assumed that two of the wind farms have a strong correlation, i.e., between [0.8, 1.0]. The numbers of rampings for different correlation values with the third wind farm are shown in Table 3.10 when δ=30%. The number of triple concurrent rampings is obviously much smaller than that of the double ones. This is consistent with the fact that higher-order failures are less frequent.

Table 3. 10: **Correlation vs. concurrent modes between three wind generation sites (*δ=30%*)**

Pearson's Correlation between wind farms (*i*, *j*)	**Pearson's Correlation between wind farms (*i*,*k*) and (*j*,*k*)**	**Numbers of concurrent under- or over-generation of wind farms *i*, *j*, and *k***	**Average point-wise distance between two rampings**
0.8 – 1.0	0.8 – 1.0	217	0.08
	0.6 – 0.8	68	0.11
	0.4 – 0.6	40	0.13
	0.2 – 0.4	21	0.11
	0.0 – 0.2	7	0.12

To model the frequencies for double rampings, the rampings for the same pair of wind farms were aggregated. The distributions of the frequencies for the pairs of wind farms with correlation value between 0.8 and 1.0 are developed based on the histograms for ramp-ups and –downs, respectively, as indicated in Figs. 3.30 and 3.34. Again, the distributions for both ramp-ups and ramp-downs are very similar to each other.

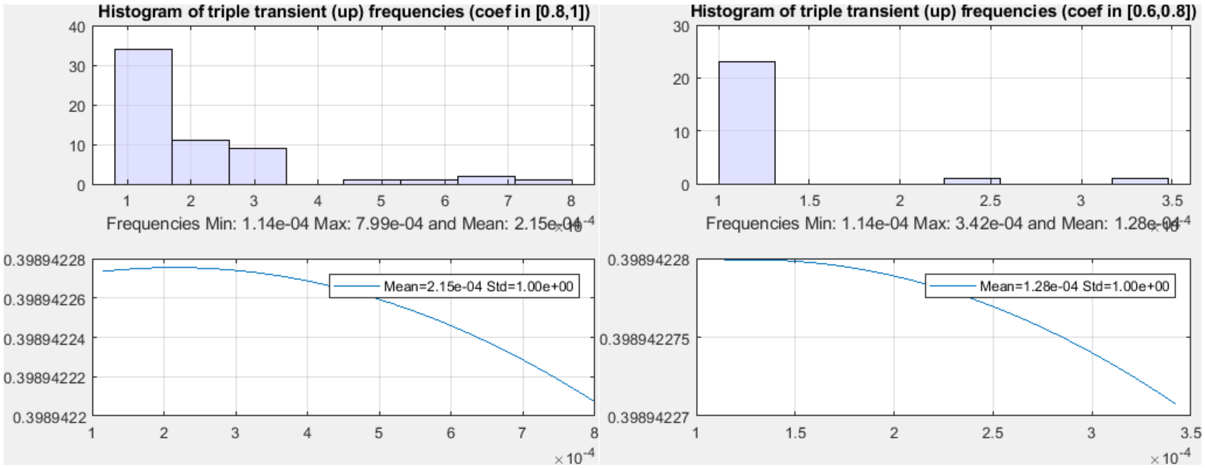

Figure 3. 30: ***Frequency distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

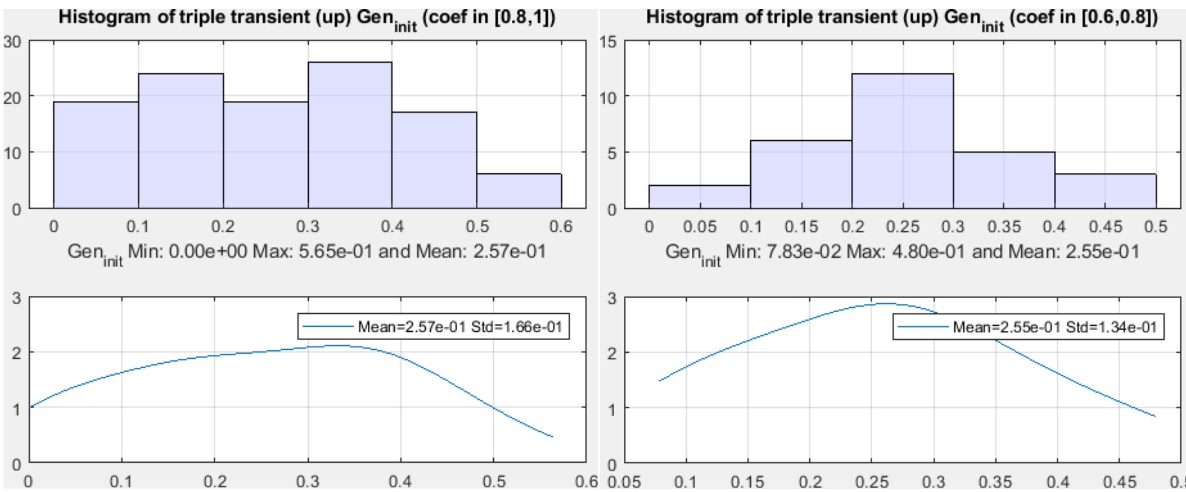

Figure 3. 31: ***Initial Generation distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

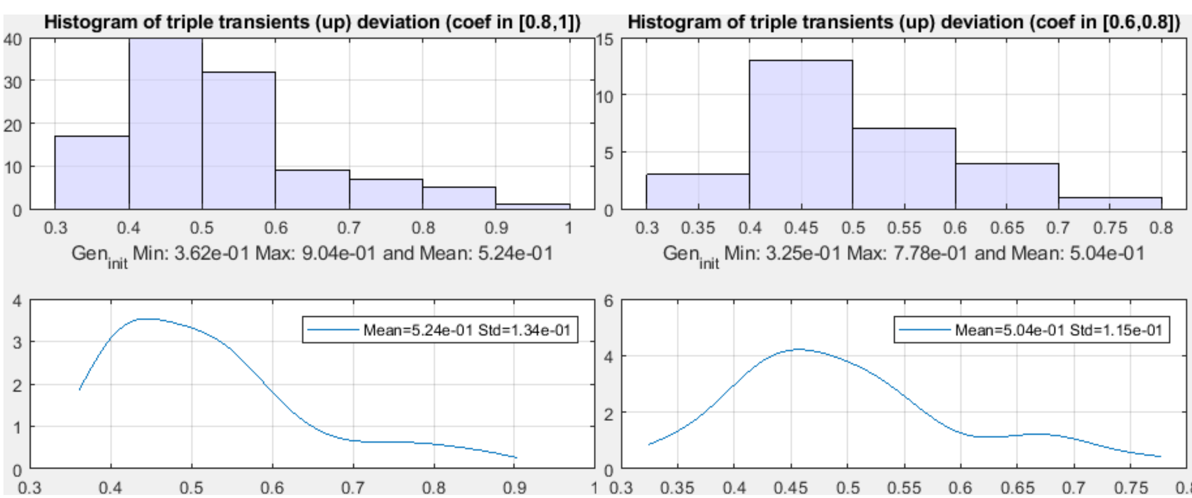

Figure 3. 32: ***Deviation distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

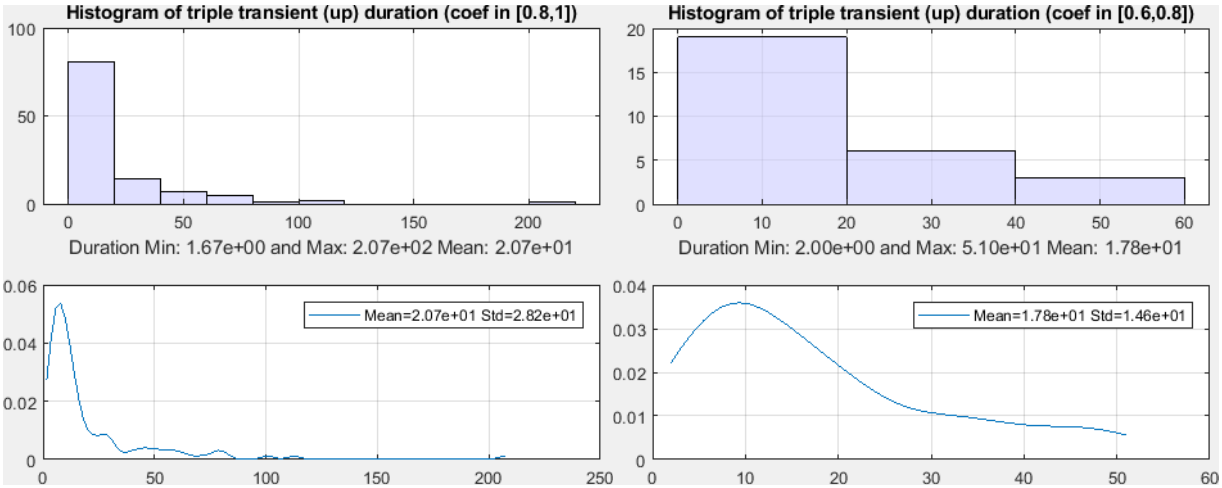

Figure 3. 33: ***Duration distributions of triple ramp-ups for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

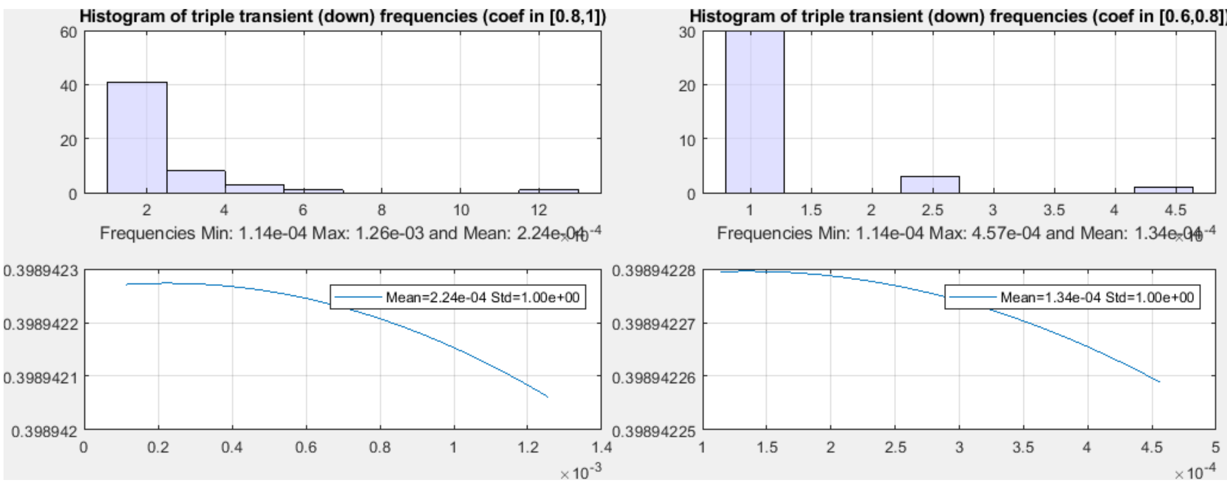

Figure 3. 34: ***Frequency distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

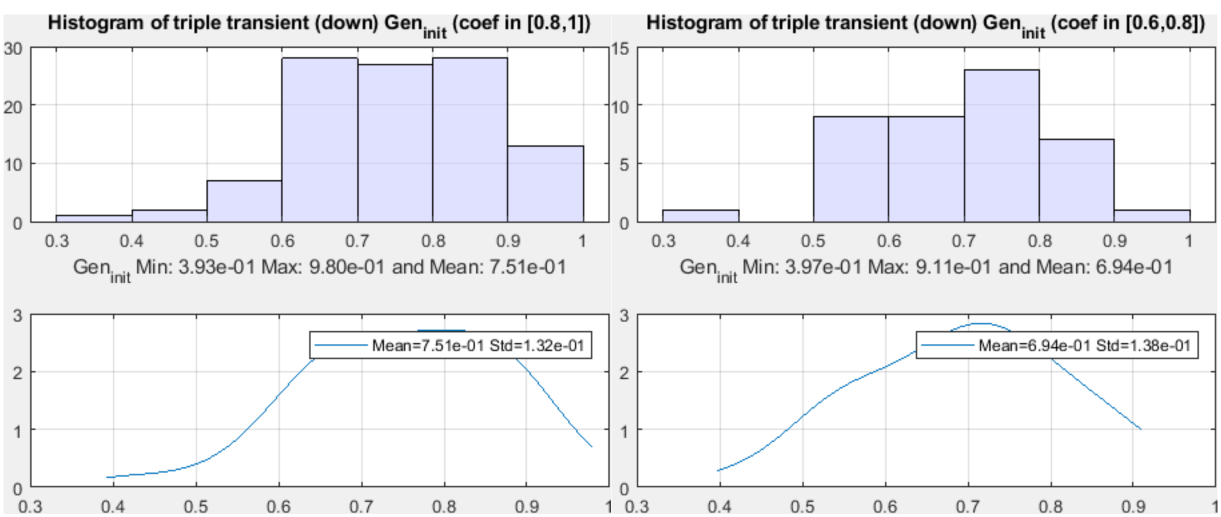

Figure 3. 35: ***Initial generation distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

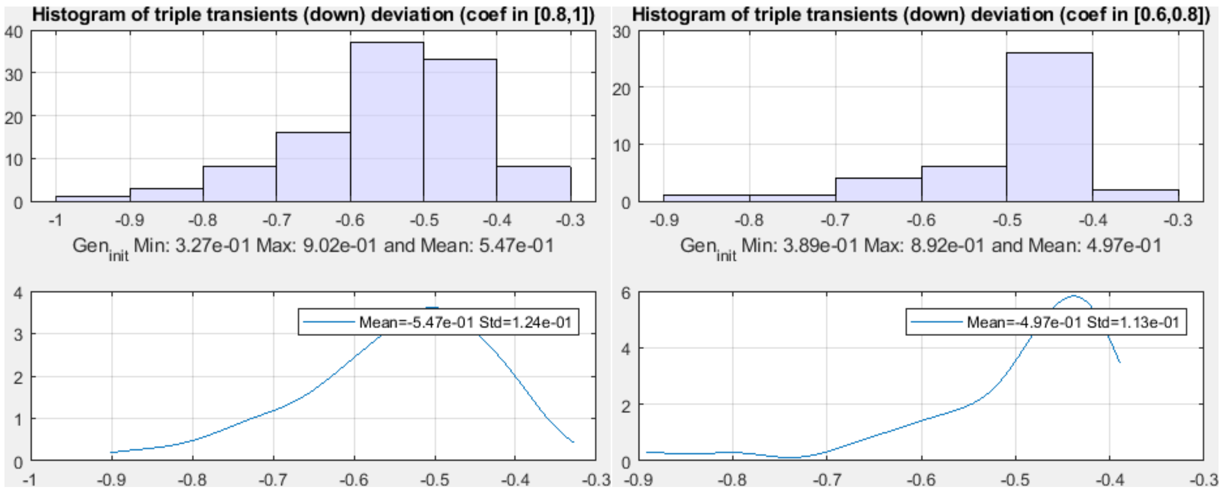

Figure 3. 36: ***Deviation distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

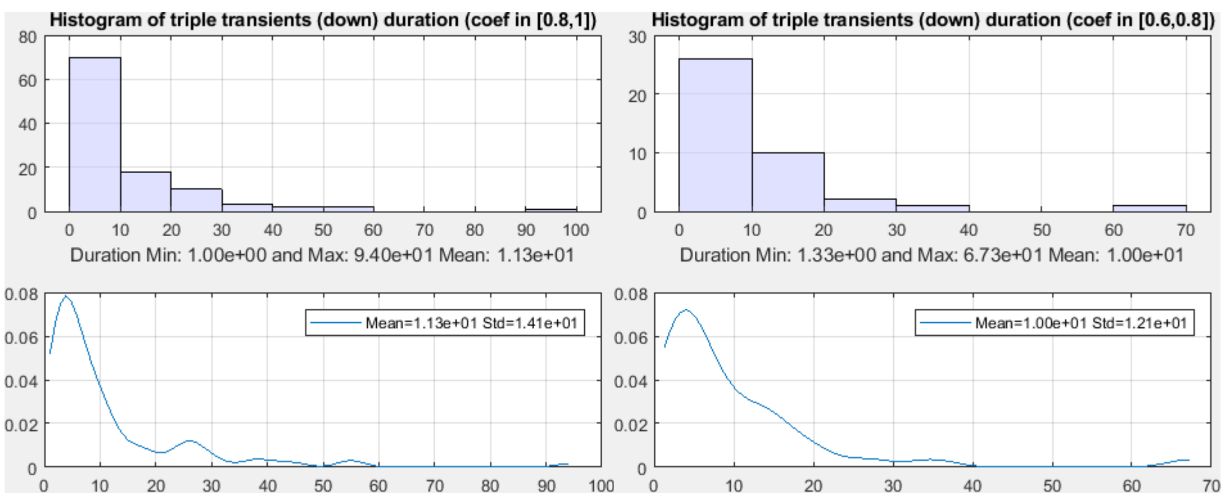

Figure 3. 37: ***Deviation distributions of triple ramp-downs for Coef in [0.8,1.0] (left) and [0.6,0.8] (right).***

It is suggested that common modes should only be modeled for highly-correlated wind farms to reduce complexity. This is considered reasonable since a weak correlation usually means that the two wind farms are distant from each other. In addition, the common modes are much more frequent for wind farms of stronger correlations. Such probabilistic models, once developed for wind or solar generators, are readily used as input and implemented in the developed tool for the enhanced PCA analysis.

When δ=0.5, the number of triple rampings is small because the numbers of wind farms and years of our data are relatively small. Therefore, the histogram and the fitted distribution function are not meaningful and not shown. Instead, the values are tabulated below as an example. Only the correlation coefficients between 0.6-1 are given. The total number of triple rampings for correlation coefficients between 0-0.6 is one and, therefore, is not listed.

Table 3. 11: ***Mean Values of Frequencies, initial generation, deviation, and durations (δ=0.5)***

Triple Rampings	Correlation	Frequency	Initial Generation	Deviation	Duration (Hours)
Up	0.8 – 0.1	0.0001	0.20	0.64	3.04
Down		N/A	0.86	0.67	0.57
Up	0.6 – 0.8	0.0002	0.12	0.75	1.78
Down		0.0001	0.88	0.75	0.26

The parameters for various outages developed in this chapter can be readily used for contingency analysis. This will be further illustrated in Chapter 5.

4. Facilitation of Decision-making Process in Transmission Planning Using Probabilistic Reliability Metrics

Deterministic transmission planning has been routinely performed and successfully applied by utilities for transmission expansion based on compliance with the NERC transmission planning standards. The most famous deterministic criterion is the (N-1) contingency, i.e., an outage of one single important component that does not cause any system instability, thermal overloading, loss of load, or cascading outages. The major issue with the deterministic criterion is that it does not account for the likelihoods of different disturbance occurrences (i.e., it assumes that all likelihoods for different disturbances are equal). Also, no matter how severe an impact a disturbance may have, if it is very unlikely to occur, then the disturbance may be ignored. These criteria are very conservative and may not be realistic, particular for the modern grid.

Deterministic planning is usually based on the worst-case scenario and often leads to overinvestment for utilities. However, the current generation power grid is experiencing transformational changes, such as the introduction of more and more renewable generation, and worst-case scenarios used in the past may no longer be the worst case. One such example is that the summer peak condition has long been considered the most challenging case to address in transmission planning. However, when wind power penetration level is high during low load conditions, many conventional generators may be offline, which leads to a significant reduction of the total inertia in the system. This subsequently may cause deterioration of frequency performance during the occurrence of ramping events associated with wind power. Therefore, it is impossible to identify the worst-case scenarios without rigorously treating uncertainties.

The increasing penetration level of renewables poses a greater challenge to deterministic transmission planning. Traditionally, renewable generation is modeled similar to conventional generators in contingency analysis. The uncertainties associated with the intermittency and variability of renewables are literally neglected, or more precisely, cannot be accounted for or fully captured by the deterministic planning criteria. These are the major reasons probabilistic planning has become more popular, i.e., the increasing uncertainties in the grid across the nation. Deterministic planning is no longer effective in terms of handling such uncertainties and is expected to be even less capable in the future for higher penetrations of renewables. As an alternative, probabilistic planning can rigorously model and, thus, capture the uncertainties inherent to renewable generation. However, the major issue with probabilistic planning gaining wide-spread use is that it is difficult for a utility to use probabilistic results in the decision-making process. Unlike the deterministic process that focuses on compliance with well-established transmission planning standards, there is a lack of probabilistic planning criteria for the transmission planner to adopt.

In this study we will review the probabilistic planning criteria and illustrate how to use them together with the deterministic criteria to facilitate and help the utilities in making a planning decision. The focus of this study is on the discussion of probabilistic reliability metrics calculated from contingency analyses (PCAs), a major technique used in transmission planning. Due to the limited resources available, case studies are not presented but can be readily performed using an enhanced PCA tool.

4.1 Probabilistic Planning Criteria

In transmission planning practices, there are only a limited number of options for utilities to upgrade their systems. Usually, the number of alternatives that are available to utilities is very small. Therefore, the utility is able to evaluate alternative expansion plans based on different assessments, such as environmental, societal and political considerations, and economic and reliability analyses. The evaluations of environmental, societal, and political assessments that will screen out some alternatives are beyond the scope of this study. Only probabilistic planning criteria for reliability and economic analyses are discussed here.

There exist three types of probabilistic reliability criteria, namely, reliability target, probabilistic cost, and incremental reliability index. One can choose different criteria for transmission and distribution systems, but the focus of this study is on transmission planning.

4.1.1 Reliability Target

Probabilistic risk assessment (PRA) has been practiced in the nuclear industry for decades. A PRA is mandated in individual nuclear power plants (NPPs) and probabilistic criteria have been defined in terms of the reliability or risk goal [47]. For example, one criterion is that the core damage frequency (CDF) in any nuclear power plant shall not exceed 10^{-5} per year. For any changes to be made to an NPP system in the utility, the PRA has to be redone and the same risk goal needs to be met to show the change does not increase risk and is acceptable. Since NPP operation is very stringently regulated, failure data collection and analysis are common practices, and reporting events of relative severity is mandated by the regulators. The availability of failure data and the tools for risk assessment make it feasible for NPP operators to perform PRAs and determine compliance with the risk goal(s). Therefore, given the availability of the data and the tools, the most straightforward criterion is to designate a reliability goal that must be satisfied.

However, when it comes to transmission or distribution planning, there is usually no such criterion, or it is very difficult to define a goal or goals for reliability. Unlike the nuclear industry, the electric power industry is not as rigorously regulated with regard to risk assessments. At the transmission level, the grid is regulated by FERC and NERC, and at the distribution level, the systems are mainly (loosely) regulated by states. At present, a well-accepted probabilistic criterion is the one defined for the loss of load expectation (LOLE) in generation adequacy assessment, e.g., a LOLE of one day per ten years has been adopted by many utilities. Even this criterion is somewhat arbitrary, and it is hard to justify why this specific number is used.

For distribution systems, there are many indices that can be used to characterize the reliability, e.g., System Average Interruption Duration Index (SAIDI) etc. However, there are no reliability goals for these lagging metrics. Instead, these may be reported by utilities to the state commissions as their performance evaluation.

Theoretically, these indices can also be calculated (and thus become leading reliability metrics) using the component failure frequencies and repair times. However, it is difficult to do so. Distribution systems are particularly vulnerable to weather events, which can be very variable. As such, the frequencies, durations, and strength of the different weather events are extremely difficult to predict.

4.1.2 Probabilistic Cost

The probabilistic cost criteria are the most popular ones, and also very straightforward to calculate since reliability and economics are very tightly coupled. These criteria can be based on the PCA and economic analysis. Both the total cost method and benefit/cost ratio method [48] can be used as probabilistic cost criteria. As an example of the total cost criterion, we can increase the reliability of a system by using more reliable (and likely more expensive) components or building more redundancy into the system (which usually means more investment is needed) [11]. As the reliability increases, continuously increasing amounts of investment will be needed to increase the reliability further. Once a certain reliability level is reached, huge amounts of investment will be needed for even a slight increase in reliability. Ultimately, a point will be reached at which the next incremental investment will not be justified by the small increase in reliability realized. At this point the utility will decide not to make any further investments. Just like any other engineering system, the reliability of a power grid cannot be 100%. Therefore, utilities always need to make decisions based on the trade-off between reliability and cost.

4.1.2.1 Total Cost Method

In the total cost method [48][49][50], the total cost (C_{Total}) is divided into investment cost (C_{Invest}), operational cost $C_{Operation}$, and unreliability cost ($C_{Unreliability}$).

[11] This is based on an implied assumption that is commonly adopted in reliability theory, i.e., a (properly designed) physical system is coherent because the system reliability will increase if the reliability performance of one or some of its components is improved [49].

$$C_{Total} = C_{Invest} + C_{Operation} + C_{Unreliability}$$

The investment cost, operational cost, and unreliability cost are closely related via the concept of system reliability. Investment cost and operational cost are expected to increase to enhance system reliability. Improved reliability will reduce the unreliability cost. Unreliability cost calculation is based on the assessment of customer cost due to power supply interruptions caused by various system problems such as overloading conditions and voltage violations, as calculated by the PCA. This unreliability or customer cost will decrease as the reliability becomes lower. Therefore, there always exists an optimum point of total cost if we evaluate the curve of the total cost vs. reliability.

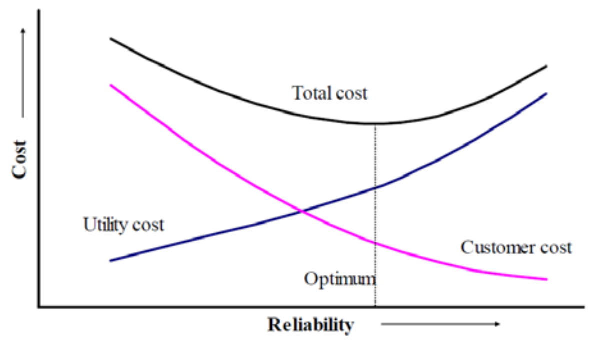

*Figure 4. 1: **A Total cost method.***

The calculation of investment cost is routinely performed in transmission planning for individual transmission expansion options. The operational cost includes operation, maintenance, and administration (OMA) expenditures, network losses, financial charges, and other ongoing costs, which can be estimated by utilities. Unreliability cost can also be assessed on a monetary basis, i.e., the assessment of customer cost due to power supply interruptions caused by various system problems such as overloading conditions and voltage violations. System problems can be identified from the contingency analysis, and the frequencies and durations of system problems can be calculated from statistics of those contingencies that cause these problems. The unreliability cost is obtained using the expected energy not supplied (EENS) index (in kWh/year) times the unit interruption cost (UIC) ($/kWh), i.e., the unit cost of interrupting customer load.

$$EENS = \sum_{i \in \Omega_{Loss}} P^i * L^i * T^i_{duration} * UIC$$

where P^i is the probability of contingency i that causes a loss of load L^i kW with a duration time $T^i_{duration}$ hours, Ω_{Loss} is the set of all contingencies that cause loss of load, N_{loss} is the number of contingencies that causes losses of load with different amounts. Here UIC is assumed to be constant, but it is possible to differentiate the costs of different types of load (e.g., critical load vs. non-critical load). Probabilities and durations of contingencies can be directly obtained from historical outage data [8] while loss of load scenarios can be evaluated by using contingency analysis. The enhanced PCA tool developed in [44] can be used for this.

The unreliability cost[12] will decrease as the reliability becomes higher and the losses of load will decrease, which means more investment and operational costs are needed. The UIC can be estimated using various methods, such as a customer survey, as discussed in [50]. If a set of pre-defined alternative options for the planning is made available, such an evaluation process is straightforward by plugging these options into a tool with enhanced PCA capability to compute the reliability using the system problem indices.

It should be pointed out that UIC is not easy to fully develop. In general, the impacts of power outages can be grouped into direct and indirect categories. The direct impacts cover loss of manufacturing and production, interruption of services such as transportation, telecommunication and so on, loss of sales, damage to equipment, spoiled goods, damage to electronic data, accidents and injuries. Compared to the evaluation of indirect cost, the direct cost assessment is easier. The indirect cost is mainly related to the indirect and usually longer-term impacts of the power disruptions. The longer-term consequence of power outages may include public disorder and crimes after the blackouts, overtime payments to personnel, cancellation of social activities, property losses, increasing insurance rates, and so on. After the reliability becomes high and people are used to it, the expectation from the public about the system is higher and it is more intolerant for people to accept power outages.

Either direct or indirect costs are related to different groups of customers with different characteristics and, therefore, it is necessary to analyze such costs for different customer groups, i.e., industry customers, service customers, and residential customers. Some have suggested that all economic cost and worth considerations will be determined by the “market” and “market forces.” This makes the estimate of the reliability and unreliability costs more dynamic and volatile.

4.1.2.2 Benefit/Cost Ratio Method

Alternatively, probabilistic cost criterion can be evaluated using the benefit/cost ratio method, where the cost includes the capital and operational costs to enhance reliability, and the benefit is the reduction in the unreliability cost. The cost and benefit can be calculated in the same manner and amortized based on the planning horizon. Again, utilities can rank the benefit/cost ratio by evaluating a set of candidate expansion plans.

4.1.2.3 Incremental Reliability Index

If it is difficult or inappropriate to use the unreliability cost in some cases, an incremental reliability index (IRI) can be applied [49]. The IRI is defined as the reliability improvement per million dollars of investment, which can be expressed as follows:

$$IRI = \frac{RI_B - RI_A}{Cost}$$

where RI can be one of the quantities calculated or used in the probabilistic contingency analysis (probabilities, durations, or EENS). Again, cost indicates the cost needed to bring system reliability from RI_A to RI_B.

It is worth noting that WECC developed its own probabilistic criterion based on different categories of disturbances. Disturbances are categorized based on the number of components that are in the outage, and each category is characterized by an associated frequency range. Then, minimum requirements for transient voltage dips, post-transient voltage deviations, and frequency nadirs are designated for different

[12] Unreliability cost can be used alone in probabilistic planning. It can be considered an application of PRA. The consequence is expressed in terms of cost and the risk is the product of the probabilities of undesired events (the contingencies that cause a loss of load) and the unreliability cost. Similarly, the system’s reliability metrics, such as probabilities/durations of system problems and EENS can also be used alone by utilities to select expansion plans from a set of alternatives. The difficulty of using them alone is that system problems also include voltage violations, which have no direct impact on customers and are difficult to associate cost with.

categories of disturbances. However, the only probabilistic part of the criterion is the frequencies of disturbances for the different categories. The rest of the criterion is all deterministic.

4.2 Combining Probabilistic and Deterministic Planning Criteria: The Well-being Approach

While the probabilistic criteria in the previous section can be used alone by the transmission planner to make a decision in selecting one of the expansion plans, e.g., the probabilities and durations for the system problems, it sometimes is still considered insufficient. Many utilities need to know not only whether the system is in trouble but also how far the system is from having issues. This is traditionally described using deterministic criteria.

The grid is designed based on the (N-1) contingency criterion and the system will be weakened after a single contingency, although the system can be operating without any violations. In this case, if an additional contingency or contingencies occur, then the system will likely be experiencing some violations. For utilities, it is desirable to know whether the system is operating at a completely normal state or under weakened conditions. This can be illustrated by using the system states [52] defined for grid operation, as shown in Figure. 4.2.

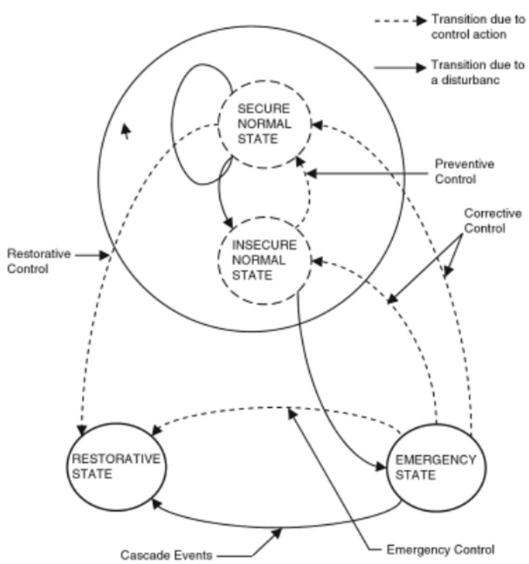

Figure 4. 2: System States for Grid Operation.

(1) A normal secure state is a state in which the following conditions are met:
  i. No equipment is disconnected;
  ii. All system loads are satisfied at the specified voltage levels and rated frequency;
  iii. The operating point exists and is stable
  iv. No device is operating outside thermal or voltage limits;
  v. The transient trajectory is asymptotically converging to an equilibrium.

  and conditions (vi – x) are that (i – v) are still satisfied after (a – e) occur

  a. Plausible load pattern and level change that can occur in 15 minutes;
  b. Plausible transfer pattern and level change that can occur in 15 minutes;
  c. Any single line outage;
  d. Any generator outage;
  e. Any fault and fault clearing time.

(2) A normal insecure state is a state in which above conditions (i) – (v) are satisfied, but one or more of (vi) – (x) are not satisfied after one or more of (a – e) occur (preventive control may be applied).
(3) An emergency state is one in which, if some operating limits are violated, e.g., overloaded transmission lines, condition (i) is satisfied; one or more of conditions (ii – v) are not satisfied because one or more of (a – e) have occurred (corrective control may be applied); or conditions (vi – x) are satisfied, or one or more may not be satisfied if one or more of (a – e) occur (preventive control may be applied).

Note that the restorative state is not defined as it is not relevant to the study.

A normal secure state indicates that the system is very healthy. The normal insecure state indicates that the system is still fine, but the operators should be alert for possible issues. An emergency state means that the system is in trouble. By using these definitions, operators can tell how close the system is to a trouble state. This is why there is a growing interest in the so-called "well-being" [51] of the system and to evaluate the likelihood, not only of the system's entering a complete failure state, but also the likelihood of being close to trouble. The underlying concept is that reliability indices obtained using the PCA can be associated with these different system states.

Three states defined in the "well-being" approach are:

(1) **Healthy State:** i.e., all equipment operates within its constraints and generation is adequate to satisfy all load demands. In addition, there is sufficient margin such that the loss of any major system component, such as generating units and transmission lines, specified by the deterministic criterion, will not result in an operating limit being violated or load curtailed. The criterion of being healthy will depend on the philosophy of individual utilities[13].
(2) **Marginal State**: i.e., the system is operating within its limits, but no longer has sufficient margin to satisfy the specified deterministic criterion. This means that the loss of some major components will result in the criterion being violated, although the system is currently still within limits and no load is actually shed.
(3) **At-risk State**: i.e., equipment and/or system constraints are violated and/or load is shed. Such states correspond to the inadequacy states enumerated by conventional composite reliability evaluation algorithms. The indexes defining the likelihood, frequency, duration, and load/energy curtailed for each type of state can be evaluated using conventional probabilistic techniques.

These three states clearly correspond to the normal secure state, normal insecure state, and emergency state defined above. Thus, the system well-being analysis presents a combined deterministic and probabilistic framework that offers a quantitative interpretation of the degree of system security (healthy state) and insecurity (marginal state) in a bulk electric power system in addition to the traditional risk measures.

The well-being approach can be easily implemented in the PCA. For the given contingency list,

(1) A contingency that causes voltage violation, overloading, or loss of load issues is categorized as a contingency that brings the system to an at-risk state, and its probability contributes to the probability of the system being in an at-risk state.
(2) A contingency that does not cause any system problems needs to be further studied by adding another contingency for the PCA.
    a. If the system still does not have any issues with the additional contingency, then the system is in a healthy state for the given contingency, and it contributes to the probability of the system being in a healthy state.
    b. Otherwise, the given contingency will contribute to the system being in an alert state. By doing so, one can calculate the probabilities of the system being in a healthy state, an alert state, and an at-risk state.
(3) The probabilities, frequencies, and durations of the system being in a healthy[14], alert, and at-risk states can be summed.

[13] A system can still be healthy with outages as long as the system satisfies the conditions of being healthy.

[14] Note that a system operating in a state without any outage is also healthy according to the N-1 design criterion. The total probability of a system being without any outages can be approximated by the product of the probabilities of being healthy for all individual components, i.e., $P_{no\_outage} = \prod_{i \in \Omega_F}(1 - P^i)$, where Ω_F is the set of all single outages of individual components.

(4) The process continues until all contingencies in the list have been evaluated.

Then, the decision maker may compare the probabilities of being in a healthy, alert, and at-risk states to the predefined thresholds to determine the best alternative among the expansion plans.

In summary, this study reviews two major approaches that can be used to assist utility planners in selecting the best option in transmission planning. The first approach is based on probabilistic planning criteria, including reliability goals, probabilistic cost, and incremental reliability indices, while the second one is based on a "well-being" analysis of the system. Theoretically, any probabilistic indices including system problem indices can be used as criteria in transmission planning. The major difficulty is the non-existence of reliability goals or targets. The probabilistic cost approach is based on the total-cost concept that links reliability and cost, which can be an efficient way of helping utilities make expansion decisions. Alternatively, a "well-being" approach combines both deterministic (whether the system is one contingency away from an at-risk state) and probabilistic (the probabilities of system states) criteria. It not only considers the likelihood of the system being in an at-risk state, but also how far the system is from the at-risk state.

5. The Enhanced PSS/E Driven by Python

To implement the probabilistic contingency analysis in existing planning platforms, such as PSS/E, scripts were developed using Python. This section describes the scripts and their architecture.

5.1 Python Code Architecture

Scripts external to the PSS/E software platform were developed using Python 2.7 and are used to drive the enhanced probabilistic contingency analysis using PSS/E 33.4. The major functional modules consist of contingency screening, an IIO/CMO processor, a sample generator, and an output module. All these modules are independent of PSS/E and interface with PSS/E to provide the required input data and analytical results of the enhanced PCA. The flowchart of the Python tool is given in Figure 5.1. The details of the flowchart are given in Section 5.6.

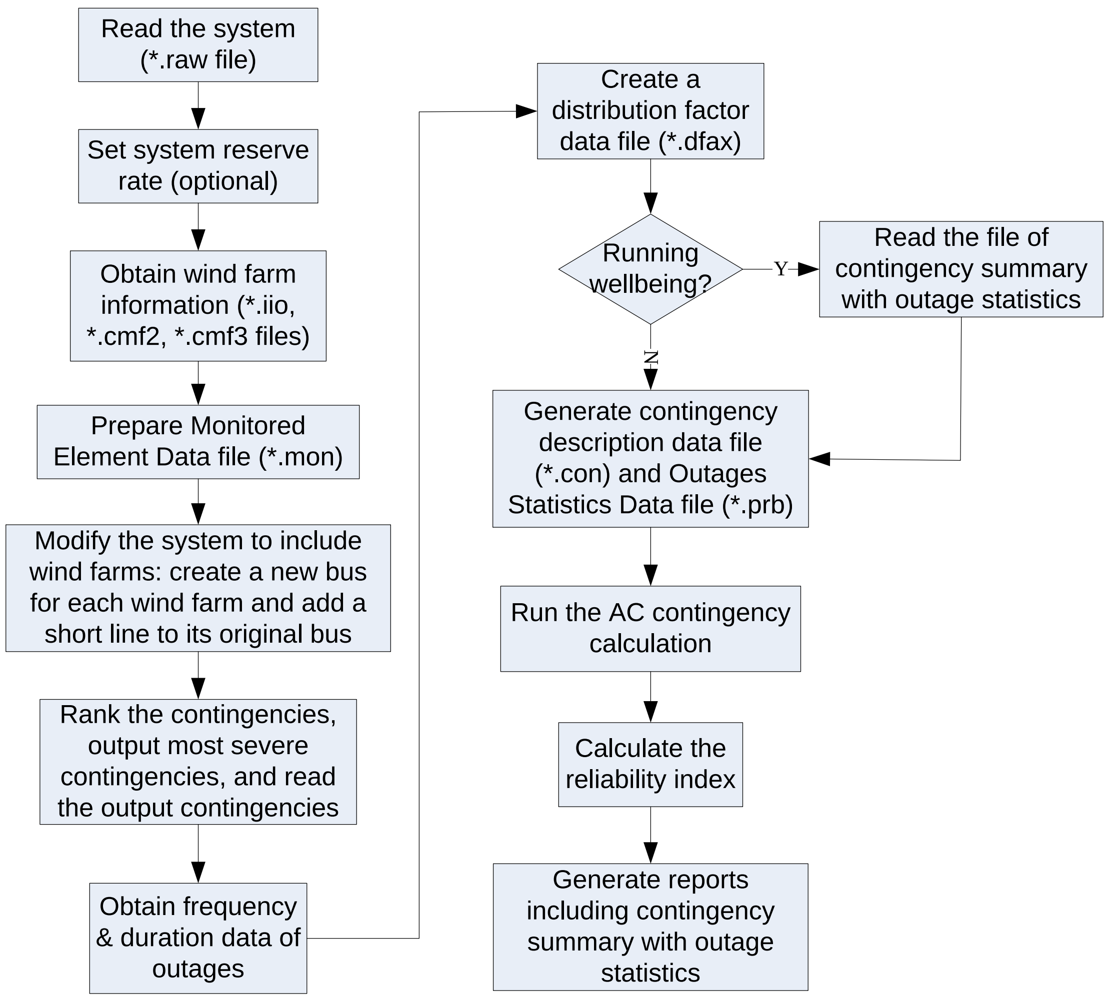

*Figure 5. 1: **Flowchart of the Python tool***

PSS/E provides Application Program Interface (API) routines for Python to run PSS/E functions. The documentation for PSS/E API details the usage of each API routine. More information about the API routines and related file inputs are available in the PSS/E Program Operation Manual. To enable the PSS/E interface, ‘import psspy’ needs to be added to the python code. Python 2 needs to be used and for this study, PSS/E 33.4 and Python 2.7 were used.

5.2 Contingency Screening

For this study, a large system was considered, e.g., with 3,000 machines and 15,000 branches. The number of single contingencies for a system of this size is 18,000. The number of double contingencies is 3.24×10^8, which is a huge number. Adding another level of contingencies will cause an exponential increase in the number of total contingencies, so the number of triple contingencies will be extremely large. Theoretically, every single, double, and triple contingency needs to be evaluated if the requirement is to perform (N-3) studies. This undoubtedly will be practically infeasible considering the time required to perform such simulations. In fact, it is also not necessary to simulate every contingency since it is possible to estimate the contingency impact without simulating it, as is done by many software tools.

It is straightforward to estimate the severity ranking of contingencies, using built-in routines, such as 'RANK_BRN_AND_MAC' in PSS/E. This API estimates the severity of designated single element outage contingencies and outputs a preset number of the most severe contingencies. N-1 contingencies of all online elements are used as input. Three preset numbers need to be specified, i.e., 1) the number of branch contingencies from overload ranking to include, 2) the number of voltage depression contingencies to include, and 3) the number of machine contingencies from overload ranking to include.

These output contingencies from the ranking process are included in the N-1 contingencies. For the N-2 contingencies, not all branches are included for the combinational contingencies, but only the outputted branch contingencies from the 'RANK_BRN_AND_MAC' are used. In addition, for the N-2 contingencies involving the generators and wind farms, a simple threshold-based method is used to reduce the number of contingencies, i.e., a contingency that has two generator outages with a total generation variation being smaller than a pre-specified threshold value is excluded from the contingency list. The details of this method are as follows. If only one generator or wind farm outage is involved, a contingency with the total variation in generation being smaller than a threshold value, denoted as Thred1, will be ignored. Note that a CMO is considered a single outage or contingency, and the same method applies to CMOs of conventional generators or wind farms. That is, the total variation of the power outputs of all the wind farms in the CMO is compared to Thred1 for screening. Similarly, if only two (or three) generators and/or wind farm outages are involved in a double (or triple) contingency, a contingency that causes the total variation of generation to be smaller than a pre-selected threshold value, denoted as Thred2 (or Thred3), will be ignored.

Consider the N-2 contingency given in Figure 5.2 as an example. It consists of two simultaneous outages, i.e., machine 1 from bus 14932 failing and a CMO of two wind farms ramping down (wind farm at bus ID 800060 ramps down by 116 MW and wind farm at bus 800061 ramps down by 150 MW). Then, the total variation of generation needs to be compared to Thred2 as two generators and wind farm outages are involved. Assuming the output of machine 1 from bus 14932 is 1,379 MW, this contingency's total variation of power output is equal to 1,379+116+150=1,645 MW. Hence, this contingency will be included in the contingency list if Thred2 is set to be smaller than 1,645 MW; otherwise, it will be ignored.

```
CONTINGENCY DBL_135_4710
REMOVE MACHINE 1  FROM BUS 14932
DECREASE BUS 800060 GENERATION BY 116.0000 MW
DECREASE BUS 800061 GENERATION BY 150.0000 MW
END
```

Figure 5. 2: ***Contingency DBL_135_4710***

5.3 IIO/CMO Processor

It is always possible that two or more wind farms are connected to the same bus. In order to specify contingencies of two wind farms that are connected to the same bus, each wind farm is connected to a new bus in the system model. For a specified bus with multiple wind farms, add a new bus for each of the wind farms and a very short line with very small impedance value (i.e., R=0, B=0, X=1e-4) between this new bus and the specified bus. By doing so, different outage modes can be defined for these wind farms.

A wind farm can have multiple outage modes, i.e., a change in the generation for different levels, which can be a continuous variable, as shown in Section 3.2. Note it is assumed that the wind farm can only fail to one of the outage modes. To simplify the problem, the generation deviation level is discretized, e.g., a reduction of 30%, 50%, 70%, and 90% in generation. Also, a CMO of multiple wind farms is treated as a separate or independent outage with parameters obtained from the distributions indicated in Section 2.2. More details are shown in Section 5.4.

Note that the maximum number of buses allowed in PSS/E is 999,997. It is better, but not mandatory, to reserve a continuous block of numbers only for new buses to be used for wind farms, e.g., 800,000-900,000 are selected for the WECC system in our study.

5.4 Sample Generator

The sample generator module is used for generating a pre-selected number of samples for the Monte Carlo simulation. It uses a list of outage distributions (annual frequencies and duration hours) for a variety of outages associated with grid components, including intermittency induced outages (IIOs). Lognormal distributions are developed for transmission circuits, transformers, and conventional generators (see Tables 3.1 to 3.5). For each contingency in the list, the corresponding component (e.g., a fossil fuel plant) and its category (between 400 and 799 MW) will be identified and the distribution will be used to generate samples.

Sampling from a given parameterized probability density function can be performed using an inverse transform sampling method, as described below:

1. Generate a random number u from the standard uniform distribution in the interval [0,1], e.g. from U~Unif[0,1]
2. Find the inverse of the desired cumulative distribution function (CDF), e.g. $F_X^{-1}(x)$.
3. Compute $X = F_X^{-1}(u)$. The computed random variable X has distribution $F_X(x)$.

The PDF of lines, generators, and wind farm IIOs/CMOs have been obtained from historical data, as described in Section 3, and used in this tool. For lines and generators, the frequency and duration values are directly obtained from the distributions. For an IIO or CMO of wind farms, the sample generation is performed in multiple steps because frequency and duration are not sufficient to describe these events. For an IIO of a wind farm, the frequency of the IIO is generated from the distribution of the frequencies across different wind farms. For this specific outage, initial generation level, deviation, and duration need to be determined separately using three independent distributions, as shown in Section 3.2.

Initial generation levels of wind farms are given in the system data. The duration can be directly sampled. To reduce the complexity, several discrete levels of deviation are specified for each IIO/CMO by three parameters, i.e., starting value, number of steps, and width of each step. The deviation sample generation is illustrated as below. The PDF of the deviation is divided into a number of equal-width bins. The center of a bin represents the deviation level and the area of the bin represents the probability of this deviation level. The discrete levels can be represented as $\{s, s-w, \cdots, s-(n-1)\cdot w\}$, where s, n, w represent starting value, number of steps, and width of each step, respectively. The probability of a deviation level, $s - i \cdot w$, is calculated as CDF$(s-(i-0.5)w)$ −CDF$(s-(i+0.5)w)$, where CDF is the corresponding cumulative distribution function of the distribution function for a deviation level. As an illustration, for a bin between the two vertical lines shown in Figure 5.3, the deviation level is 70%, the width is 10%, and the probability of this deviation level is equal to CDF(75%)-CDF(65%). Note that although a normal distributed is plotted in Figure 5.3, the description given above also applies to any other distribution, including an empirical distribution.

The parameters for a CMO can be generated in exactly the same manner.

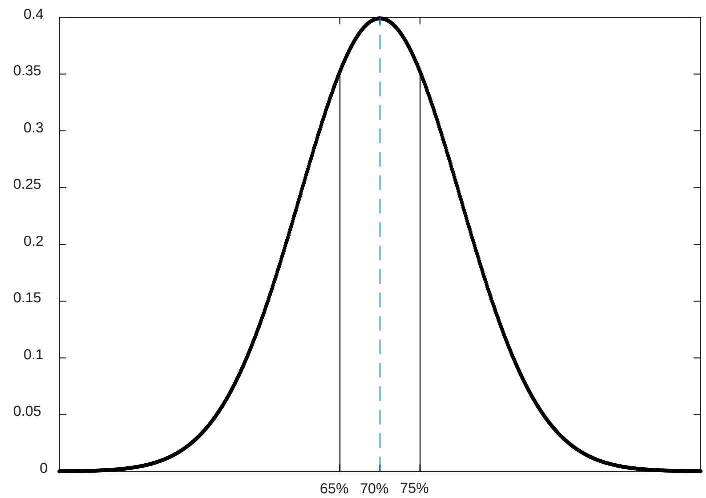

Figure 5. 3: ***Illustration of a bin for the probability of a deviation level.***

5.5 Output Module

The tool can generate 12 types of reports via the PSS/E API routine 'relind_2', i.e., system problem summary report, system loss of load report, bus loss of load report, branch flow overloading report, bus voltage violation report, contingency summary, system problem indices, system loss of load indices, bus loss of load indices, branch overload problem indices, bus voltage problem indices, and contingency summary with outage statistics.

The results in the report of system problems and bus voltage problem indices can be fully retrieved, i.e., the frequency, duration, and probability of the entire system problems and bus voltage violation (out of the range of bus voltages) are extracted and stored in a text file. Also, the frequency, duration, and probability of system overload information are extracted and stored in a text file.

The histograms and kernel density estimates of the frequency/duration of the system problem, overload, overvoltage, and undervoltage are plotted where data are read from these files. Note that the kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. The histograms and kernel density estimates can visually show that the frequency/duration of an outage is probabilistic, but not deterministic.

5.6 Flowchart of Python Code

The flowchart of the Python code architecture has been provided in Figure 5.1. The details of each step of the flowchart are given below.

1. Read the system
2. Set system reserve rate (e.g., reserveRate=0.15) and then modify the machine active power upper limit (PMax) such that the spinning reserve does not exceed the set value. The detailed modification is given in Section on the WECC system. This step is not required but can be used to adjust the system reserve margin, as needed.
3. Obtain wind farm information[15].

[15] Such information is included in *.iio, *.cmf2, *.cmf3 files representing IIO, and CMOs of double and triple rampings. In these files, each line has the machine id, bus number, PMax, and the machine active power output (PGen). Detailed information for each column is given in the Appendix. In the *.cmf2 and *.cmf3 files, the first value in each row is the correlation between two/three machines.

4. Prepare a Monitored Element Data file[16].
5. Modify the system to include wind farms. Move the wind farms to a new bus to ease the wind ramping related contingency generation for the PCA.
   a) Select a bus with a wind farm;
   b) Create a new bus and add a wind plant to it;
   c) Disconnect a wind farm from the original bus and move it to the new bus, i.e., modify the wind farm unit by replacing the original bus number with the new bus number[17];
   d) Add a very short branch between the original and new buses (R=0, B=0, X=1e-4)
6. Rank the contingencies using PSS/E's built-in API and store the output contingencies. This function can have input parameters, i.e., the number of branch contingencies from overload rankings to include, the number of voltage depression contingencies to include, and the number of machine contingencies from overload rankings to include.
7. Obtain frequency & duration data of outages for lines with/without transformers, generators, wind ramp up/down events.
   a) The wind farm output ramp-up and ramp-down events are not deterministic, but probabilistic. Based on data analysis, the probability distribution function (i.e., probability of different per-unit power deviations) uses an empirical distribution, as shown in Fig. 3.12 of Section 3.2.3.2. Samples are taken from the distribution to represent the probability and power deviation of ramp up and down events. As a small deviation has an insignificant effect on the PCA of the WECC system, only large and discrete deviations are sampled. In this study, 100%, 90%, 80%, 70% deviations of IIO and CMO are adopted in this study (see Section 5.4).
8. If it is running the wellbeing mode, read the contingency result generated in the basic mode and store the contingency label
9. Create a distribution factor data file
10. Generate a Contingency Description Data (CDD) file (*.con) and Outage Statistics Data (OSD) file (*.prb)
    a) Name the contingency label 'SGL_#' for an N-1 contingency and 'DBL_#_#' for an N-2 contingency where '#' is the index in the contingency ID list
    b) Generate the N-1 and N-2 contingencies in CDD files and OSD files[18].
11. Run the AC contingency calculation
12. Calculate the reliability indices
13. Generate reports, including a system problem summary report, system loss of load report, bus loss of load report, branch flow overloading report, bus voltage violation report, contingency summary, system problem indices, system loss of load indices, bus loss of load indices, branch overload problem indices, bus voltage problem indices, and a contingency summary with outage statistics.

---

[16] File *.mon via function 'generate_mon_file'. The file includes the voltage deviation and/or voltage range of all/specific buses . The *.mon file used for the WECC system is given in the Appendix.
[17] The maximum number allowed in PSS/E is 999997. It is better, but not mandatory, to reserve a continuous block of numbers only for new buses for wind farms, e.g., 800000-900000 are selected for the WECC system.

[18] Pseudo codes for generating these files are shown in the appendix.

6. Case Study

Multiple case studies are evaluated in this chapter using different systems, including a 23-bus example system and a WECC system. The major focus of the case study is to show how the data poolability issue is addressed, and to compare the results using the built-in PCA capability and the enhanced PCA capability. Another major focus is on the demonstration of the impacts of the intermittency induced outages (IIOs) due to renewable generation. In addition, the well-being approach is performed to show how this can help utility planners in their decision-making process for transmission expansion.

6.1 A Case Study Using a 23-bus System

The example system used in this demonstration study is a three-area (Areas 1, 2, and 3) 23-bus system [7]. The study will be based on a pre-defined contingency list, a monitoring file, and the subsystem definition, as can be found in [7][13].

The system problems include overloading, and over-voltage and under-voltage violations in Area 2 (this depends on the area[s] of interest defined in the sub system file). The indices include the frequencies of such violations and the duration hours of a single violation. The subsystem problem total indices are derived using the indices of individual violations.

For a comparison of the existing and enhanced PCAs, two cases were performed and are presented. For Case 1 of the existing PCA, the arithmetic mean values of raw data were used. For Case 2, 10,000 samples were generated following the distributions of poolable grid components in Table 3.4, as well as the point estimates for other (i.e., non-poolable) components in Table 3.5 and CMOs in [39]. These were used as input for the enhanced PCA. The distributions for the system problem indices were developed using the histogram and a kernel for the corresponding outputs of the samples. The mean and standard deviation are also calculated using the fitted probability density functions (PDFs), as shown in Figures 6.1 and 6.2.

Table 6. 1: Mean Values of System Problem Frequency and Duration

	Case 1: Built-in PCA	Case 2: Enhanced PCA
System Problem Frequency (per Year)	9.27	17.64
System Problem Duration (Hours)	47.1	34.15

The mean values for Cases 1 and 2 are shown in Table 6.1. For Case 1, the mean values of the frequencies and duration hours, i.e., 9.27 occurrences per year of the system problem and 47.1 hours duration per system problem. Compared to the mean values calculated in Case 2, i.e., 17.64 per year and 34.15 hours per system problem, the differences between the two cases are significant, especially for the frequency. The means in Case 2 are calculated using a Monte Carlo simulation, and, therefore, are the true mean values of the PCA outcomes, as long as the number of samples is large enough. Very often, the mean values alone are used for making a decision. Therefore, an accurate calculation of mean values is vital.

Compared to the results in [17], the mean value of system problem frequency is much higher mainly due to the high generator outage frequencies. It should be noted that the contribution to the high frequencies includes startup (Class 0 or U0), delayed (Class 2 or U2), and postponed (Class 2 or U3), and forced outages that are not very relevant to the CA. One solution to this can be an investigation of the percentages for different types of forced outages.

Additional information can also be extracted from the PCA outputs in Case 2. For example, the standard deviations calculated in Case 2 indicate how the estimates of system problem frequencies and duration hours deviate from the mean. Standard deviations for both system problem frequency and duration are shown in Figures 6.1 and 6.2. The difference between the standard deviations for frequencies and durations is relatively large. This is also consistent with the large differences in variances and error factors for frequency and duration distributions indicated in Table 3.4. It suggests that there are some contingencies that may cause extremely long outages. These outcomes, although they may be rare, may pose a much bigger threat to the system.

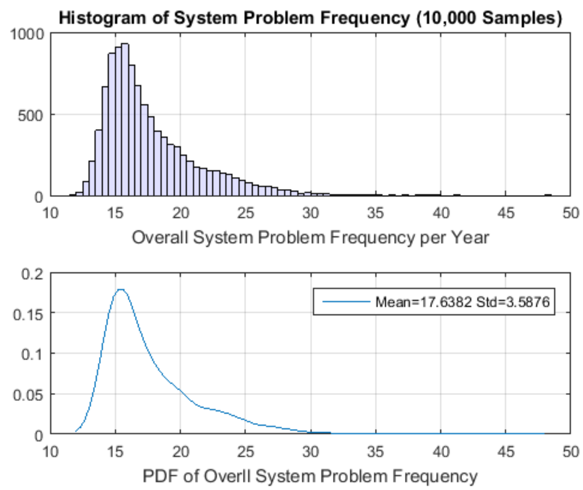

Figure 6. 1: ***Histogram and pdf of system problem frequencies.***

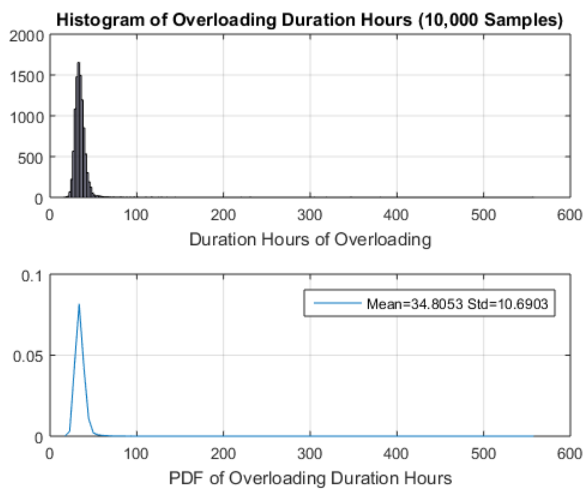

Figure 6. 2: ***Histogram and pdf of system problem durations.***

Thirdly, while the mean values of frequency and duration provide much information about the system problems, it is also interesting to see the shapes of the PDFs for system problem frequencies and durations. In this study, all outage parameter distributions were assumed to be lognormal, which takes only positive real values and skews to the right. The distribution for the system problem frequencies is fairly different from a lognormal distribution. This can be observed in Figure 6.1 for the system problem frequencies. However, the distribution for the system overload frequencies shown in Figure 6.3 is more obvious. Figure 6.3 clearly shows that the histogram for the overload frequencies has two modes, and a bi-modal distribution was obtained using a kernel-based fit.

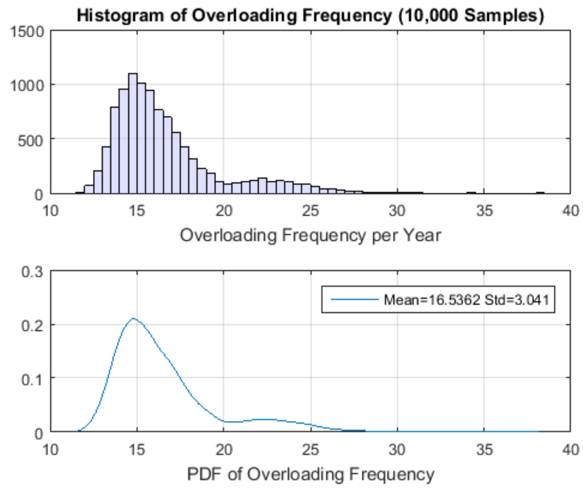

Figure 6. 3: ***Histogram and pdf of system overloading frequencies.***

In Figure 6.3, while most of the overloading frequencies are between 13.0 and 20.0 per year, the frequencies for a significant number of system problems are between 20.0 and 25.0 per year. This is apparently due to the combinational effects of different distributions. This suggests that, if one intends to develop mitigations and fix the system problems, attention should be paid to contingencies causing system problems within both frequency ranges. Different system problems can be studied separately, e.g., over-voltage or overload issues, to better understand the issue and provide mitigations for it [13]. Note that this bi-modal distribution for overload frequencies is not as obvious as in [17] because of less diffuse distributions (or smaller variances in Table 3.4) of the generator outages at large.

6.2 A Case Study Using the WECC System

6.2.1 WECC System

For demonstration purposes in this study a section of the WECC system was selected that is comprised of 21,549 buses, including 3,113 online and 1,093 offline generators, and 16,056 in-service and 1,332 out-service branches. The total current output (PGen) of the online machines is 178,747.8 MW. The total capacity (PMax) of the online machines is 222,952.6 MW. The spinning reserve is 24.73% and the total spinning reserve capacity is 44,204.8 MW, which is very high. The capacity of the largest single machine in the WECC system is 1,379 MW, which is 1/32 of the total spinning reserve capacity. A system with a large amount of spinning reserve generally has less contingency-caused problems compared to a system with less spinning reserve. Therefore, for purposes of this study the system operating parameters were modified to reduce the spinning reserve to no more than 15%, i.e., the active power upper limit of each machine is set to the smaller of its actual value and the machine active power output multiplied by 1.15, i.e., PMax = min(PMax, PGen*1.15).

Based on knowledge of the renewable generation available in the WECC system, there are a number of wind farms and solar plants in the system. However, all generation sources are modeled exactly like conventional generators and it is unclear which generators are renewables. Therefore, in this study we refer to the WECC system with unspecified renewable sources as the **base case**. To consider IIOs and CMOs for wind generation, additional wind farms were added and used to replace the original generators in the base case to represent different penetration levels of wind generation. The PMax & PGen for these generators were assigned to the wind farms. The original generators were randomly selected to be replaced by wind farms. Therefore, the locations of these new wind farms are not related to the real locations of the wind farms in the WECC system. Although this is not realistic, it is an acceptable assumption that allows us to exercise the enhanced PCA developed in this study to capture both random failures and IIOs/CMOs inherent to renewable generation.

6.2.2 Parameter Settings for Enhanced PCA Studies

The parameter settings used in this study are listed in Table 6.2. For purposes of this study, the variation of each event (a failure of a machine, IIO or CMO of wind farms) needs to be larger than 50 MW; otherwise, this event was excluded from participating as a contingency.

*Table 6. 2. **Parameter settings for Enhanced PCA***

Parameters	Functionalities	Settings
iterlib	Determine whether a well-being analysis will be performed.	['basic'] if running only the basic mode; ['basic', 'wellbeing'] if running both the basic and wellbeing modes
PRCMode	Whether parallel AC contingency analysis will be performed.	2 if using parallel ACCC; 3 if using sequential ACCC
para.NumSamples	Number of samples in PCA	500 (number of samples)
number_of_threads	Number of threads for parallel implementation of ACCC	10 (number of threads for parallel ACCC)
para.compare_mode	Used to determine whether to use PCA and wind farm IIO/CMO	1, 2, 3, or 4. more information is available at Section 6.2.5
penetration	Penetration level of wind	0.1, 0.3, or 0.5 (penetration of wind power)
reserveRate	System reserve margin	0.15 (reserve rate)
iter_max_FNSL	Maximum iteration of fully-coupled Newton-Raphson power flow	1,000
para.num_rank_branch	Number of branch contingencies from overload ranking to be included in the contingency list	30
para.num_rank_voltage_depression	Number of voltage depression contingencies to be included in the contingency list	30
para.num_rank_machine	Number of machine contingencies from overload ranking to be included in the contingency list	30
para.threVal_level1	Thred1, threshold for one generator or wind farm outage as explained in Section 4.2	1,200 MW
para.threVal_level2	Thred2, threshold for two generators or wind farms outage as explained in Section 4.2	1,600 MW
para.threVal_level3	Thred3, threshold for three generators or wind farms outage as explained in Section 4.2	3,000 MW
para.number_of_wind_ramp_steps	Number of deviation steps for wind IIO or CMO	6
para.wind_ramp_step	Interval between consequent deviation steps of wind IIO or CMO, 10 represents 10%	10

To improve convergence, the fully-coupled Newton-Raphson power flow was adopted. However, setting it to a small value (e.g., 20 as default or 50) for the WECC system used in this study resulted in many cases stopping after reaching the maximum iteration limit with a large mismatch, i.e., the power flow does not converge. Thus, to resolve this issue the maximum iteration limit was set to 1,000 in this study**.**

For the example case of the WECC system used in this study, the voltages of some buses, which are denoted as set O (given in the appendix), were always out of their specified range, i.e., [0.9, 1.1], even in the base case without any component failures. To focus on issues due to contingencies, a subsystem named VOLHL was defined to include all buses except those in set O. In the Monitored Element Data file, subsystem VOLHL was monitored instead of monitoring the whole system. Note that the normal range of voltage levels was set to [0.9, 1.1].

6.2.3 Comparison between built-in PCA and enhanced PCA with and without IIOs for Wind Generation

As another example to demonstrate the use of the enhanced PCA, two more case studies, i.e., Cases 3 and 4, were performed similar to Cases 1 and 2 but using the WECC system described above. To investigate the effect of wind ramping on the PCA, cases 5 and 6, were performed based on the WECC system with and without wind ramping events; the results were then compared. Note that in Cases 3 and 4, all generators were treated as conventional generators since the wind farms were not specified in the raw system data file. The only information provided was whether a generator was in on or off status. In Cases 5 and 6, additional wind generation was added to specific buses, as indicated in the IIO data file given in the Appendix. IIOs and CMOs for wind farms were considered in Case 5, but not in Case 6. The various cases studied are summarized as follows:

- Case 3: used the PSS/E built-in PCA for the WECC system base case (i.e., with wind farms modeled as conventional generators but IIOs not considered)
- Case 4: used the enhanced PCA for the WECC system base case (i.e., with wind farms modeled as conventional generators but IIOs not considered)
- Case 5: used the enhanced PCA for the WECC system with 10% more wind generation than the base case considering their IIOs and CMOs
- Case 6: used the enhanced PCA for the WECC system with 10% more wind generation than the base case but not considering their IIOs and CMOs

Since the number of components in the WECC system is huge, it is impossible to evaluate all the contingencies, especially the higher order contingencies. The number of single and double contingencies considered after screening in each case are tabulated in Table 6.3. The double contingencies were generated using a combination of the selected single contingencies. In Cases 5 and 6, the penetration level of wind generation was 10%. Note that the numbers of both single and double contingencies are relatively small because high threshold values[19] and small values of rank parameters[20] were used in the ranking such that many contingencies were discarded.

To perform the enhanced PCA, a total of 100 samples was taken for calculating system problem indices for Cases 4 - 6. The mean and standard deviation of frequencies and durations of the overvoltage, undervoltage, and system problems for Cases 3 - 6 are tabulated in Table 6.4. Note that in Case 3, arithmetic average values for frequencies and durations for each category of grid components are given in Table 3.4 and were directly used as input to the PCA. The EENS for each of Cases 3 - 6 is given in Table 6.5.

Table 6.4 shows that the system problem indices for Case 3 are not very different from Case 4, unlike the large differences for Cases 1 and 2 in Table 6.1. This is because the arithmetic means of many components are close to the means of the distributions, as shown in Table 3.4. However, Table 6.5 shows the large differences in EENS between Cases 3 and 4. The reason is that the mean values of frequency and duration in Case 3 are similar to that in Case 4; however, the frequency/duration of some contingencies that cause load shedding in Case 3 are quite different from Case 4. This leads to the large differences in EENS between Cases 3 and 4. Also, using the enhanced PCA can provide more information (e.g., the distribution) about the EENS of the system than the built-in PCA. This indicates the necessity of using the enhanced PCA instead of the built-in PCA.

The results for Cases 5 and 6 show that the frequency of overvoltage problems and system problems (referring to any system problem, including overvoltage, undervoltage, loss of load, voltage collapse, non-convergence, or bus voltage change being greater than a specified value) in Case 5 is larger than Case 6 while the duration in Case 5 is smaller than Case 6. The reasons are as follows. Wind ramping contingencies, including the IIOs and CMOs, are considered in Case 5 but not Case 6. Hence, contingency frequency in Case 5 is larger than Case 6. The duration of a wind ramping event is usually within several hours or less while conventional contingencies, such as generation and branch failures, can last for weeks. This explains why the duration in Case 5 is smaller than that in Case 6. Table 6.4 indicates that consideration of IIOs/CMOs will significantly increase system problem frequencies but reduce system problem durations. This is a very interesting observation and yet easy to understand considering the nature of renewable outages. This trend is expected to be more pronounced when wind penetration level becomes higher in the system.

[19] See parameters para.threVal_level1, para.threVal_level2, and para.threVal_level3 given in Table 6.2.
[20] See parameters para.num_rank_branch, para.num_rank_voltage_depression, and para.num_rank_machine in Table 6.2.

*Table 6. 3: **Number of single and double contingencies for Cases 3, 4, 5, and 6.***

Cases	# of single contingencies	# of double contingencies
3	63	2,405
4	63	2,405
5	63	3,267
6	63	2,466

*Table 6. 4: **Frequency, duration, and probability of overvoltage, undervoltage, and system problems for Cases 3, 4, 5, and 6.***

Problems	Cases	Frequency		Duration		Probability (hours)
		Mean	Standard deviation	Mean	Standard deviation	
Overvoltage	3	642.6	N/A	4.1	N/A	2,634.7
	4	655.3	227.8	4.4	2.6	2,712.9
	5	993.1	389.7	2.8	1.8	2,572
	6	680.6	273.5	4.7	4.4	2,889.8
Undervoltage	3	70.4	N/A	3.5	N/A	246.4
	4	72.3	27.0	4.2	4.8	277.4
	5	71.0	30.1	2.6	3.4	170.9
	6	75.8	33.2	3.5	6.1	227.9
Loss of Load	3	0.14	N/A	257.1	N/A	36.0
	4	0.14	0.01	71.7	94.6	10.0
	5	0.01	0.02	2.7e-3	2.2e-3	2.7e-5
	6	0.01	0.01	3.7e-3	4.1e-3	3.7e-5
System	3	649.8	N/A	4.2	N/A	2,729.2
	4	663.0	230.9	4.4	2.7	2,741.5
	5	994.7	392.0	2.8	1.8	2,582.1
	6	681.1	274.2	4.7	4.4	2,901.2

*Table 6. 5: **EENS for Cases 3, 4, 5, and 6.***

Cases	EENS (MW)	
	Mean	Standard deviation
3	81.66	0
4	22.69	29.44
5	8.8e-4	3.9e-3
6	2.14e-3	8.68e-3

Table 6.4 also shows that voltage problems are an issue for the example case study. One of the reasons is that many buses had voltage issues in the original base case, as discussed above. Another major reason is the rare event approximation used in the calculation, as discussed in future work.

To further illustrate the additional information that can be provided by the enhanced PCA, histograms and distributions for system problem indices from Case 4 are shown in Figures 6.4 – 6.6. It can be seen from Figure 6.4 that the mean overvoltage frequency is 655.3, as opposed to 642 for Case 3, as shown in Table 6.3. Figure 6.4 also shows that the standard deviation is large, i.e., 227.8. A similar analysis can be applied to Cases 5 and 6, as demonstrated in Figures 6.7 – 6.12. These results again indicate that using the enhanced PCA is beneficial because the distributions of frequencies and durations of system problems can provide more information, i.e., mean values as well as standard deviations.

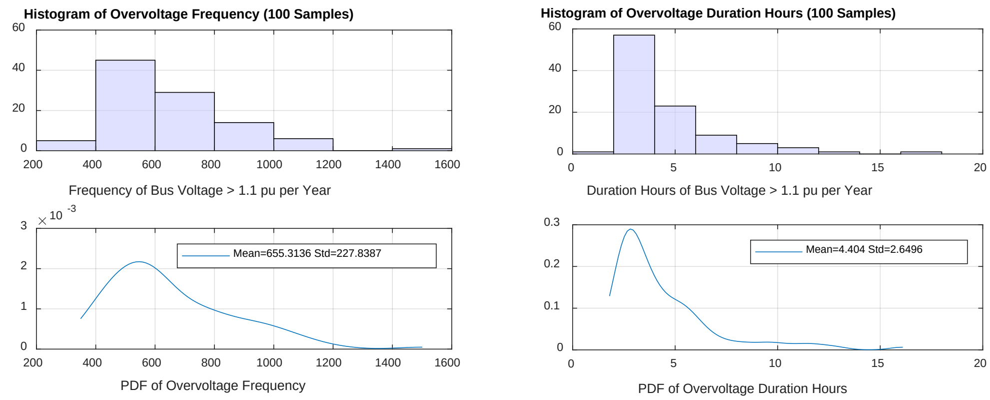

Figure 6. 4: ***Histogram/distribution of overvoltage contingencies in Case 4: (a) frequency and (b) duration.***

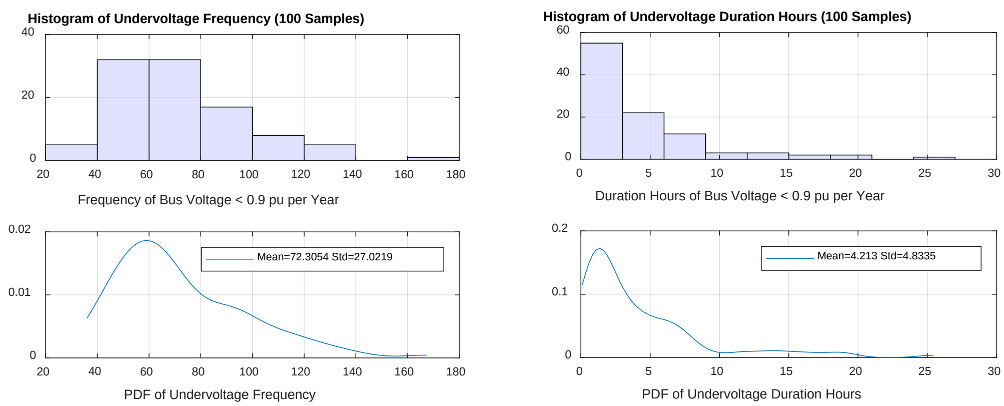

Figure 6. 5: ***Histogram/distribution of undervoltage contingencies in Case 4: (a) frequency and (b) duration.***

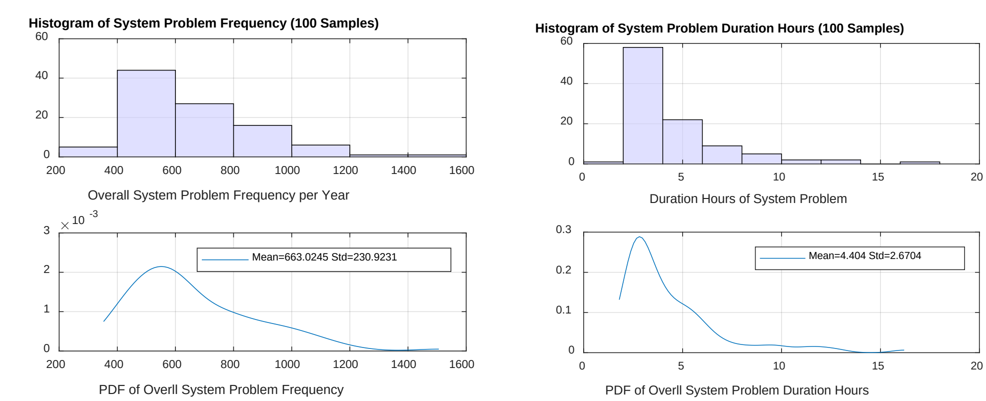

Figure 6. 6: ***Histogram/distribution of system problem contingencies in Case 4: (a) frequency and (b) duration.***

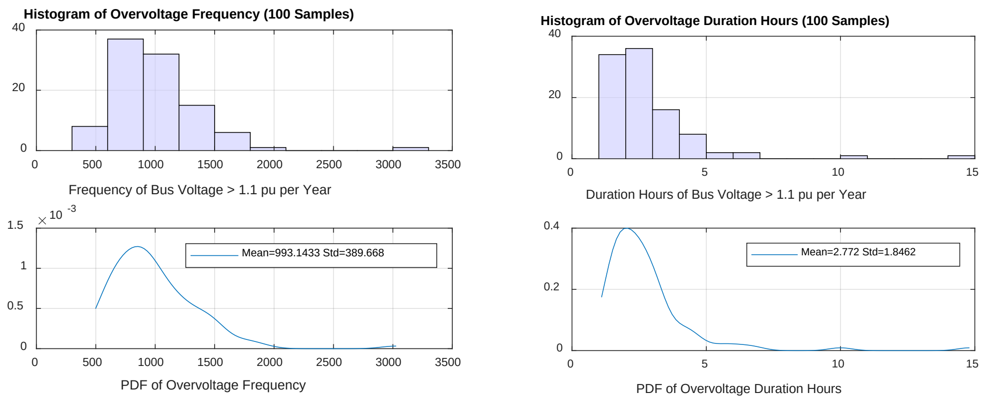

Figure 6. 7: ***Histogram/distribution of overvoltage contingencies in Case 5: (a) frequency and (b) duration.***

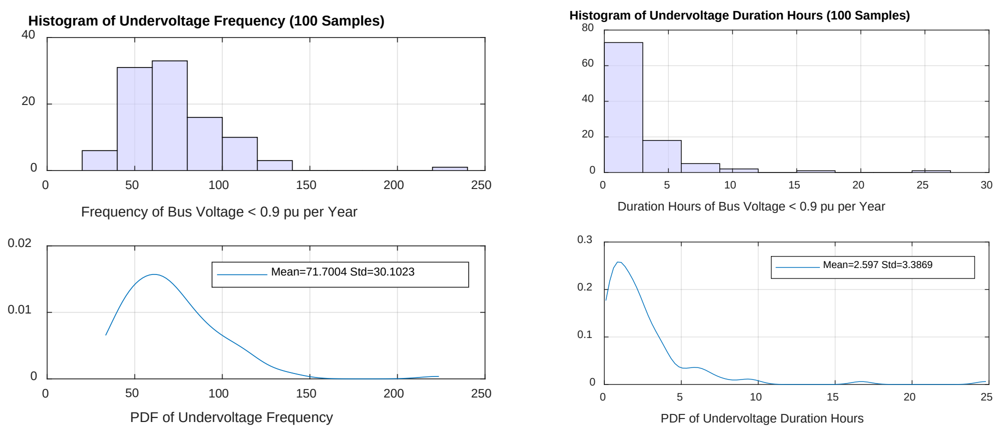

Figure 6. 8: ***Histogram/distribution of undervoltage contingencies in Case 5: (a) frequency and (b) duration.***

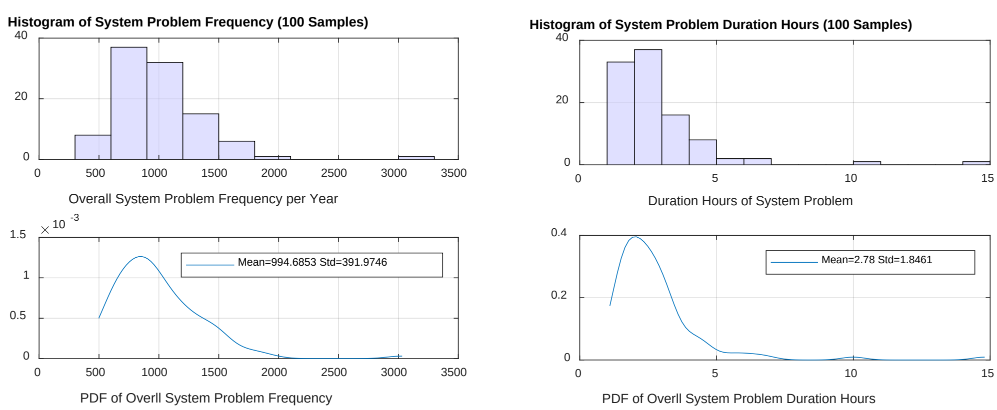

Figure 6. 9: ***Histogram/distribution of system problem contingencies in Case 5: (a) frequency and (b) duration.***

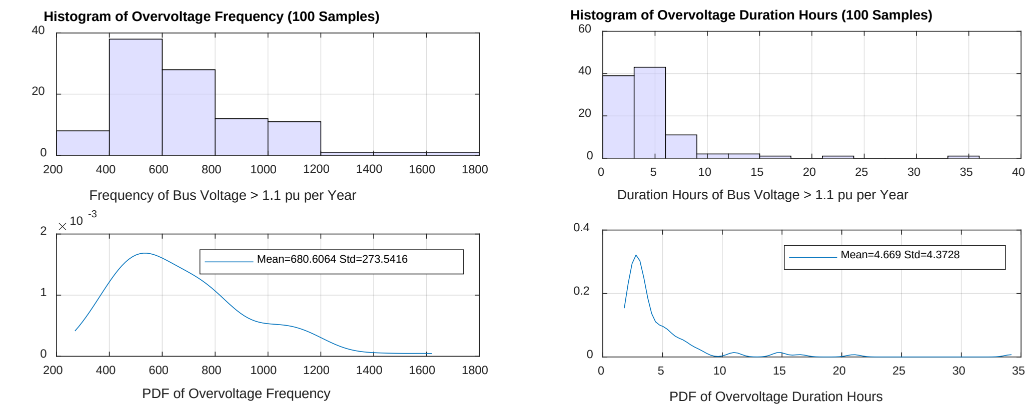

Figure 6. 10: ***Histogram/distribution of overvoltage contingencies in Case 6: (a) frequency and (b) duration.***

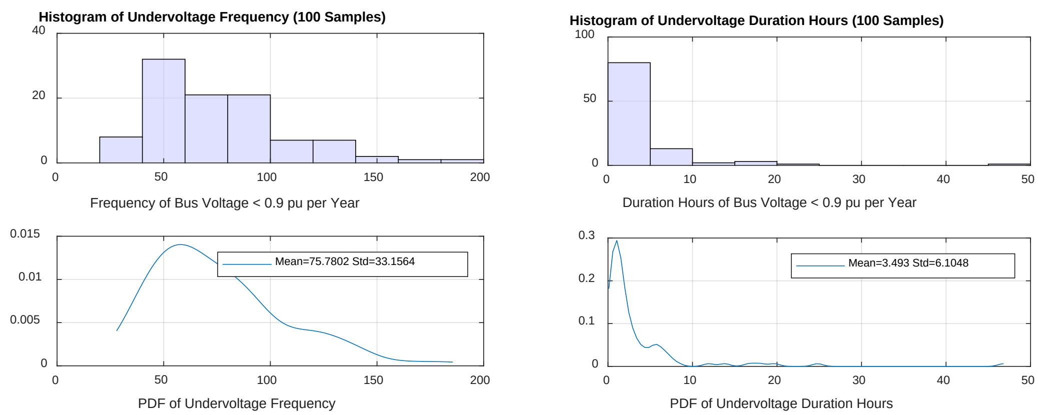

Figure 6. 11: ***Histogram/distribution of undervoltage contingencies in Case 6: (a) frequency and (b) duration.***

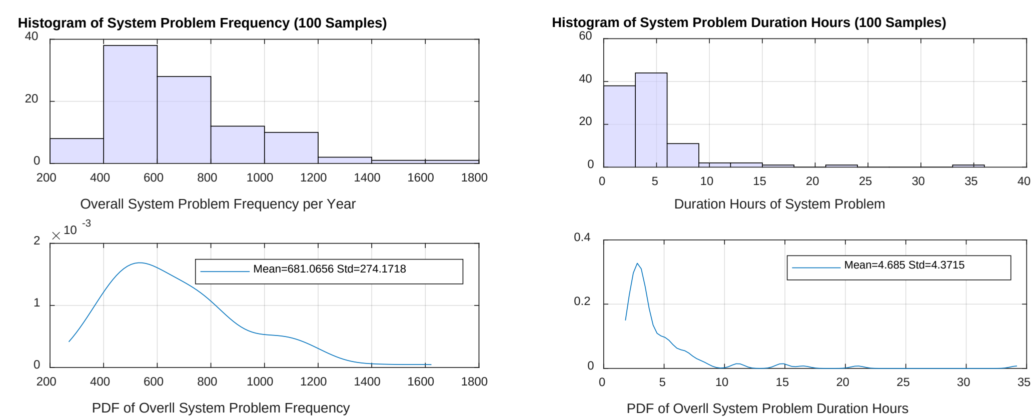

Figure 6. 12: ***Histogram/distribution of system problem contingencies in Case 6: (a) frequency and (b) duration.***

Table 6.5 shows that the values of EENS for Cases 3 and 4 are much larger than Cases 5 and 6. To determine the reason, Cases 3 - 6 were evaluated again with different numbers of contingencies, which are referred to as Cases 3v2, 4v2, 5v2, and 6v2. In these cases, the rank parameters (i.e., para.num_rank_branch, para.num_rank_voltage_depression, and para.num_rank_machine) were set to 150, 200, and 100, respectively. The number of contingencies in Cases 3v2, 4v2, 5v2, and 6v2 are tabulated in Table 6.6. Tables 6.3 and 6.6 show that the number of contingencies increases significantly by increasing the rank parameter.

EENS values were computed again and are shown in Table 6.7. Comparing Tables 6.5 and 6.7 shows that the EENSs of Cases 3 and 3v2 are similar and the EENSs of Cases 4 and 4v2 are also similar. However, the EENS of Case 5v2 is much larger than for Case 5, and the EENS of Case 6v2 is also much larger than for Case 6. Examining the outage statistics in the contingency summary report, it is found that a credible contingency that significantly affects the EENS, i.e., the product of probability and load shedding value for this contingency is large, exists in Cases 3, 4, and 3v2-6v2, but is missing in Cases 5 and 6. This caused the EENS values for Cases 3 and 4 to be much larger than for Cases 5 and 6. Thus, it is necessary to set the rank parameters to relatively large values to avoid missing credible contingencies.

Table 6. 6: ***Number of single and double contingencies for Cases 3v2 - 6v2.***

Cases	# of single contingencies	# of double contingencies
3v2	353	50,345
4v2	353	50,345
5v2	353	60,716
6v2	353	59,915

Table 6. 7: ***EENS for Cases 3v2 – 6v2.***

Cases	EENS (MW)	
	Mean	Standard deviation
3v2	81.7	0
4v2	23.3	29.1
5v2	20.6	26.8
6v2	21.2	27.5

6.2.4 Impacts on System Reliability of Different Wind Penetrations Considering IIOs/CMOs

To investigate the impacts of wind penetration on system reliability when IIOs and CMOs are considered, three different wind penetrations were studied, i.e., 10%, 30%, and 50% represented by Cases 5, 7, and 8, respectively. Under these three settings, 113, 266, and 507 wind farms were virtually deployed in the WECC system, respectively. The total number of contingencies in these three settings were 3,330, 5,913, and 85,480, respectively. The number of single and double contingencies were 63 and 5,850 in Case 7 (79 and 85,401 in Case 8), respectively, after screening. Corresponding time consumptions for the ACCC are 106.2 min, 179.9 min, and 63 hours for Cases 5, 7, and 8, respectively.

For the sake of comparison, the mean and standard deviation for frequency and duration of the overvoltage, undervoltage, and system problems are tabulated in Table 6.8. This table shows that as the wind penetration increases, the mean and standard deviation of the frequency increase while the mean and standard deviation of the duration decrease. The reason is that when the wind penetration increases, there are more wind farms which have more IIOs, and CMOs, and, therefore, the frequency of problems increases. Also, the wind farm IIOs and CMOs have short durations; thus, the duration of the problems decreases as the wind power penetration increases. On the other hand, fewer conventional generators remain in the system as the wind penetration level increases, and outages associated with conventional generators are less. This causes increasing frequencies or system problems, but decreasing durations of system problems, which can be seen clearly in Table 6.4. Still, the probability of problems has an increasing trend.

It can be concluded that, for increasing penetration levels of wind generation, system problem frequencies will increase, and durations will decrease if IIOs are considered in contingency analyses.

*Table 6. 8. **Frequencies, durations, and probabilities of the overvoltage, undervoltage, and system problems for with different wind penetrations.***

Problems	Case	Penetration	Frequency		Duration		Probability
			Mean	Standard deviation	Mean	Standard deviation	
Overvoltage	5	10%	993.1	371.2	2.8	1.8	2,572.0
	7	30%	1911.1	629.8	1.3	0.8	2,558.9
	8	50%	5,993.6	3,520.4	0.6	0.5	3,368.4
Undervoltage	5	10%	71.7	28.2	2.6	3.3	170.9
	7	30%	283.0	93.1	0.8	1.2	230.4
	8	50%	1,410.1	1,039.7	0.5	0.8	738.9
System	5	10%	994.7	372.9	2.8	1.8	2,582.1
	7	30%	1,937.7	627.2	1.6	1.2	3,195.3
	8	50%	6,479.9	3,518.2	0.5	0.4	3,291.8

6.2.5 Implementation of Well-being Approach and Case Study

As discussed in Chapter 4, the well-being approach divides all system states into three categories, i.e., healthy states, marginal states, and at-risk states. The system problem probabilities shown in Tables 6.4 and 6.8 all correspond to the probabilities for the system to be in an at-risk state for the list of given single and double contingencies.

The well-being approach was implemented by following the procedures in Section 4.2 for the given contingencies, which may consist of single, double, triple contingencies etc. If a given contingency causes system problems, it contributes to the probability of the system being in a risk state. Among the remaining contingencies, if adding another single contingency causes system problems, the contingency contributes to the system being in a marginal state. Otherwise, the contingency contributes to a healthy state.

Double and triple contingencies are generated and added to the single contingencies to formulate the contingency list. As the number of N-3 contingencies is very large for the WECC system, only the results of Case 3 are presented here. The probability of the selected N-3 contingencies is very small as the number is small. The cutoff limit for the probability of a contingency is set to 1e-9. That is, a contingency will be ignored if its probability is smaller than 1e-9, and only 17 N-3 contingencies remain. None of them cause a system problem. The frequency, duration, and probability of overvoltage and undervoltage problems are listed in Table 6.9, which also represent the frequency, duration, and probability of the system being in a marginal state. The frequency, duration, and probability of the system being in an alert state have already been given in Table 6.4.

*Table 6. 9: **Frequency, duration, and probability of the system being in a marginal state with overvoltage, undervoltage, and system problems for Case 3.***

Problems	Case	Frequency	Duration	Probability
Overvoltage	3	8.2	3e-8	2.6e-7
Undervoltage	3	2.4	3e-8	7.2e-8
System	3	0	0	0

Figure 6.13 provides a schematic of N-1, N-2, and N-3 contingencies and is used to explain the calculation of probabilities of different system states in a well-being evaluation. S in each circle is a set representing system states. p in each circle is a set of probabilities for each corresponding state in S. sum(p) represent the sum of all elements in p.

In this project, N-1 and N-2 contingencies are considered. The total probability of at-risk states is the sum of sum(p_{11}) and sum(p_{21}). The total probabilities of healthy and marginal states are sum(p_{30}) and sum(p_{31}), respectively. Tables 6.3 and 6.5 provide the total probability of at-risk states. The total probability of marginal states is listed in Table 6.6. The total probability of healthy states is approximately equal to 8,760 - sum(p_{11}) - sum(p_{21}) - sum(p_{30}), ignoring higher order contingencies.

The full contingency list in Figure 6.13 refers to a list of all possible single-component contingencies in the system. Using the full contingency list will result in a huge number of N-2 and N-3 contingencies. Therefore, the contingency screening method described in Section 5.2 is used such that a resonable number of contingencies remains. The contingencies excluded by the contingency screening method are assumed to represent the system in a healthy state. This assumption will result in an overestimation of the probability for the system being in a healthy state.

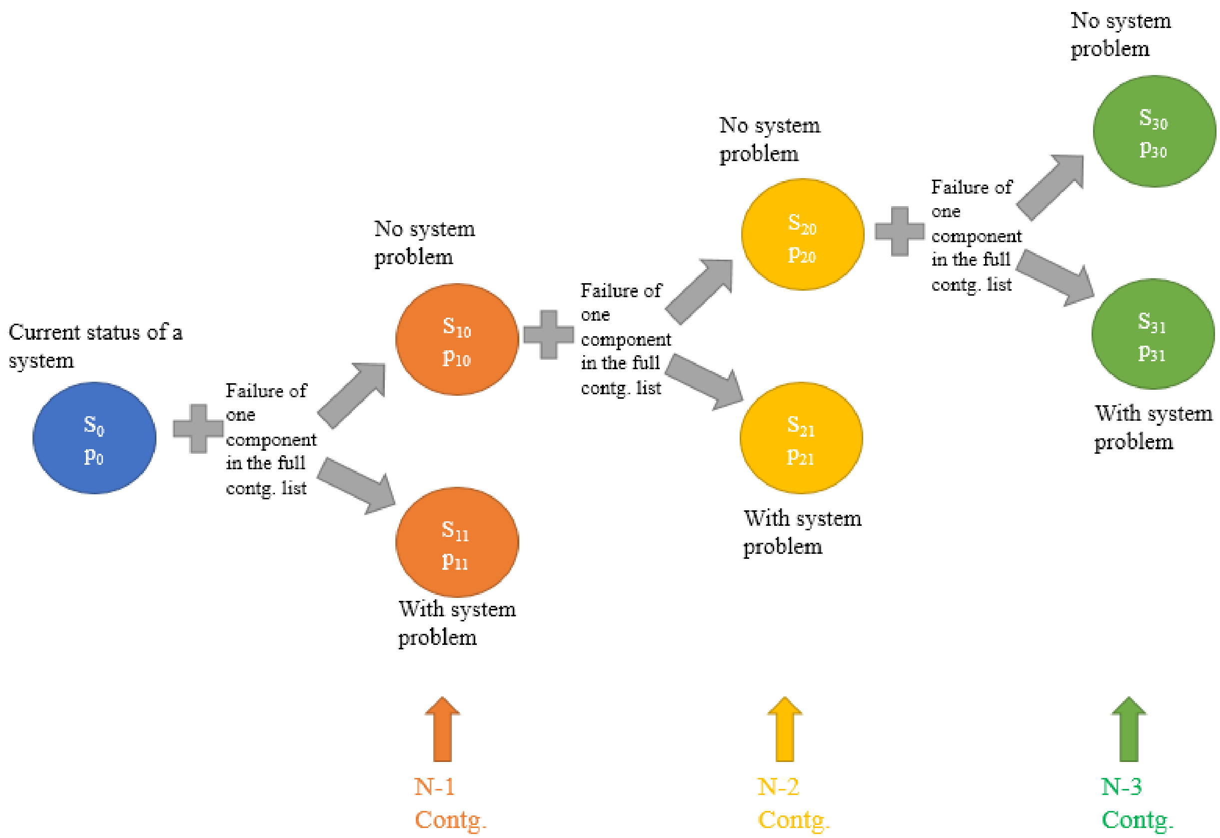

Figure 6. 13: ***Schematic of N-1, N-2, N-3 contingencies and well-being***

6.2.6 Parallel vs. Sequential Execution of Contingency Analysis

The HPC enhancement was investigated using the example 23-bus system in the previous study [17]. It concluded that the time savings of parallel processing of the contingencies is very trivial in this study because only a very small number of contingencies need to be evaluated. The major computation effort was spent on the reliability assessment using a large number of samples. According to the release notes for PSS/E 33.4 [53], on a typical four-core computer, a three-fold improvement can be achieved in terms of computational times needed for performing (N-1-1) contingency analysis. A similar improvement can also be observed for parallel ACCC analysis, as claimed in [5].

For the WECC system example, parallel implementation of the contingency analyses was studied. In the parallel mode of ACCC, the number of threads is set to 10. For Cases 5, 6, and 5v2, there are 3,330, 2,529, and 61,069 contingencies.

A DELL Precision 3630 Tower desktop computer with Intel Core i7-8700, 6 Core 3.20GHz, 32 GB RAM was used for this study. The time needed for completing Cases 5, 6, and 5v2 are tabulated in Table 6.10. As shown, for Cases 5 and 5v2, the sequential ACCC is faster while for Case 6 the parallel ACCC is slightly faster. Based on our simulation results, parallel ACCC is sometimes faster, but in some cases can also be slower than the sequential ACCC; the difference is not significant. Therefore, either one can be used.

Table 6. 10: ***Contingency Analysis Computational Time (Parallel vs. Sequential)***

	ACCC time	
	parallel	sequential
Case 5	106.2 min	92.9 min
Case 6	62.5 min	64.0 min
Case 5v2	1,728.8 min	1,253.2 min

It is clear that the benefit of using the parallel AC contingency analysis capability built into PSS/E version 33.4 cannot be justified even with a large utility-scale system. On the contrary, in the three case studies, only Case 6 showed a minor improvement in terms of computational time. For the other two cases, the parallel processing time is even longer that the sequential calculation, especially for Case 5v2, which has a much larger number of contingencies. This might be improved in the new versions, but they are currently unavailable and could not be evaluated.

7. Conclusions and Future Work

This study investigated potential outages that are induced by the intermittency of renewable generation, which have not been reported previously. Sample rampings from real data indicate that such rampings can be valid outages. Probabilistic models and parameterization of the models for these outages were developed to provide input data on frequencies and durations of different outage modes. This data can be used as input to contingency analyses. Historical solar and wind generation data were collected to analyze the statistics of these outage modes, e.g., how often a reduction in wind generation at a certain level occurs and how long this reduction lasts. Compared to outages of conventional generators, intermittency induced outages (IIOs) from renewables can be more frequent, but typically have shorter duration times, as indicated by the results in this study. The renewable outages might become dominant for very high penetration levels of wind and solar generation.

This study focuses on demonstrating a method for extracting s from operating data and performing statistical analyses of the extracted data. As indicated, the CMOs pose a much more severe threat to the system than traditional outages and need to be addressed in contingency analyses. To include IIOs in PCA studies, on top of the loss of generation due to wind turbine failures, additional outage modes for wind generation need to be created first by discretizing the generation variation, e.g., 90% or 50% increase or decrease from initial generation level. These new outages need to be added to the contingency list and, together with the outage parameters derived in this paper, can be readily fed into PCA studies by using, e.g., an enhanced Python-driven PCA tool [44] to calculate the probabilistic reliability metrics. The parameterization of CMO models and IIO applications in a PCA will be reported separately.

Probabilistic studies are not only about methods or techniques. The importance of data to the probabilistic studies can never be over-emphasized. Without quality data and rigorous analysis of the data, the outcomes of the studies can only provide misleading or even erroneous information for making decisions. This study developed a repository of outage data that has been extensively collected from different publicly or commercially available sources. The variability of outage data was investigated using a formal statistical test procedure and captured using statistical distributions, which were used to propose a generic enhancement scheme for existing planning tools. Using the existing PCA capability in PSS/E as an example, a demonstration was performed and results clearly validate the importance of capturing data variability. PCA practitioners can either choose to use the raw data in tabular form, or statistical distributions developed using the raw data and the methods presented in this report. If more outage data is collected, it can be integrated into the data repository developed here to provide additional statistical analysis accuracy.

Demonstration of an enhanced PCA method using a large-scale system is considered an immediate next step. With the developed tool, this will be a straightforward exercise when large-scale system data is available. Another important future study will be the inclusion of renewable generation in the probabilistic contingency analysis to address the challenges posed by increasing renewables in the grid. This can be done by following the method briefly discussed in Section 5.3. The study will need extensive renewable generation data collection and more statistical analysis.

This study has provided a number of important insights related to probabilistic contingency analyses, including the following:

1. The study shows that data poolability is indeed an issue that needs to be resolved first before performing a probabilistic contingency analysis (PCA). As shown in this study, outage data for many different types of grid components cannot be pooled and their statistics need to be described by distributions.
2. The results also show that a majority of outage data from conventional generators can be pooled. For the remaining conventional generators, distributions of their outage statistics are relatively narrow compared to other grid components. These results also are an indicator that environmental conditions have significant impacts on grid component outages.
3. Based on analysis of actual renewable generation data, results show that generation from a single wind farm can ramp up and down significantly within a very short time (i.e., tens of minutes). In

addition, this type of ramping event can happen concurrently for multiple wind farms and can cause significant impacts on grid operation. The impacts can be similar to conventional generator outages, and such fast ramps should be considered as outages.

4. Case studies performed in this project show that Monte Carlo simulations implemented in the enhanced PCA are capable of computing the true mean values and providing statistics on system problem indices. This can provide valuable information for developing mitigation actions in the planning process.
5. Increasing penetration levels of wind generation will definitely have increasingly larger, but different impacts on grid reliability. Due to the nature of outages related to the intermittency of renewable generation, system problem frequencies will increase and their durations will decrease.
6. A "well-being" approach can be implemented in the enhanced PCA framework and additional information can be provided to facilitate decision-making.
7. The performance of parallel processing cannot be justified and needs to be further investigated.

As a follow-up to the current project, additional studies are proposed to further refine the tool and transfer the technologies to industry. The following studies are recommended for future activities:

1. To supplement the data collected in the study, outage data from other sources will be sought, such as Canadian experience, and analyzed to update and expand the reliability data repository.
2. Solar PV generation is another important source of renewable generation. It will be included in the PCA by collecting and analyzing solar generation related data.
3. The interval of the time series wind generation data used in this study is 5 minutes. The algorithm for extracting fast ramping events is relatively simple, i.e., it considers the monotonic increase or decrease in the ramping events. There exist time series data of different temporal resolutions (e.g., 2 second intervals) for wind generation or wind speed. Therefore, fluctuations of wind generation may occur during a ramping event and the event may still qualify as an outage. A generic algorithm will be developed to extract fast ramping events from data of different temporal resolutions to support a PCA study using the enhanced PCA tool.
4. The PCA built into existing software, including PSS/E, calculates probabilistic indices of system problems approximately based on a rare event approximation. When the probabilities of contingencies are relatively large, the rare event approximation can no longer be used, and calculated probabilities may deviate significantly from the real values, or even be larger than 1.0. This is especially true for higher order contingencies since they are not mutually exclusive. Theoretically, the total probability can only be calculated exactly using the inclusion-exclusion principle, which may be difficult to implement because large computational efforts are required when the number of contingencies is large. A new quantification scheme is needed and will be developed to more precisely calculate the probabilistic indices in the PCA.
5. We will reach out to more utilities to further exercise and refine the enhanced PCA tool and demonstrate the capabilities of the tool.

It should be emphasized that probabilistic planning is intended to complement the conventional deterministic process rather than replace it. Deterministic criteria such as the (N-1) and (N-1-1) design rules still need to be followed. Probabilistic planning simply provides more information to help improve decision-making.

The use cases developed in this study demonstrated the usefulness of probabilistic reliability metrics and the complementary enhancement of the deterministic counterpart considering renewable generation in the emerging PCA. The prototype tool development in the current project has raised strong interest from several utilities in terms of data, tool, and applications of probabilistic metrics in grid planning. The outcome of this current project has also been used to support multiple GMLC projects including the "Extreme Event Modeling" and "Metrics" projects.

8. References

[1] R. Billinton, “Canadian Experience in the Collection of Transmission and Distribution Component Unavailability Data,” International Conference on Probabilistic Methods Applied to Power Systems, Ames, Sept. 2004.

[2] B. Silverstein and D. M. Porter, “Contingency Ranking for Bulk System Reliability Criteria, IEEE Transactions on Power Systems,” Vol. 7, No. 3, August 1992.

[3] R. B. Adler et al, “An IEEE Survey of U.S and Canadian Overhead Transmission Outages at 230kV and Above,” IEEE Transactions on Power Delivery, Vol. 9, No. 1, January 1994.

[4] M. G. Lauby et al, “MAPP Bulk Transmission Outage Data Collection and Analysis,” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-103, No. 1, January 1984.

[5] A. A. Chowdhury, L. Bertling, and B. P. Glover, “Causal and Seasonal Analysis of Bulk Transmission Line Forced Outages Using the MAPP Outage Database,” 9th International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, June 11- 15, 2006.

[6] “Transmission Availability Data System Automatic Outage Metrics and Data,” NERC – 2009 Report, June 2014, 2010.

[7] “PSS/E 33.4 Program Operation Manual,” March 2013 Siemens Industry Inc.

[8] M. Yue and X. Wang, "Outage Data Collection and Parameter Estimation for an Improved Probabilistic Contingency Analysis," Proc. 2015 IEEE Power Engineering Society General Meeting, 315-320.

[9] R. Sevlian and R. Rajagopal, “Detection and Statistics of Wind Power Ramps,” *IEEE Transactions on Power Systems*, Vol. 28, No. 4, November 2013.

[10] North America Electric Reliability Corporation, “Transmission Availability Data System (TADS) DEFINITIONS,” May 2010. Online available.

[11] A. Malvaldi et al, “A Spatial and Temporal Correlation Analysis of Aggregate Wind Power in an Ideally Interconnected Europe,” Wind Energy, 2017 (20).

[12] M. Yue and X. Wang, “Grid Inertial Response-Based Probabilistic Determination of Energy Storage System Capacity under High Solar Penetration,” *IEEE Transactions on Sustainable Energy,* Vol. 6, No. 3, July 2015.

[13] M. Yue and X. Wang, "Enhancing PSS/E Probabilistic Contingency Analysis for Development of Risk-Informed Decision-Making Capability," Technical Report to U.S. DOE, Office of Electricity Delivery and Energy Reliability, June 18, 2015.

[14] M. Yue, S.W. Kang, C.L. Jin, and J. Matevosjana, “An Investigation of Potential Intermittency Induced Outage Modes for Wind Generation,” 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), May 2018.

[15] C. L. Atwood et al, “Handbook of Parameter Estimation for Probabilistic Risk Assessment,” NUREG/CR-6823, 2002.

[16] A. Yoshida, "Tests for the poolability of panel data in the presence of unknown heteroskedasticity: an asymptotic extension of anova," *The Japanese Economic Review*, Vol. 50, No. 3, September 1999.

[17] M. Yue and X. Wang, “An Enhanced Probabilistic Contingency Analysis for Development of Risk-Informed Transmission Planning,” 2016 IEEE Power Engineering Society General Meeting.

[18] “pc-GAR for Windows, Software User Guide,” Version 2.06 v08, January, 2009.

[19] “Transmission Availability Data System 2008 FRCC Updated Report,” June 14, 2010.

[20] “Transmission Availability Data System, FRCC – 2009 Report,” June 14, 2010.

[21] “Transmission Availability Data System 2008 MRO Updated Report,” June 14, 2010.

[22] “Transmission Availability Data System, MRO – 2009 Report,” June 14, 2010.

[23] “Transmission Availability Data System 2008 NPCC Updated Report,” June 14, 2010.

[24] “Transmission Availability Data System, NPCC – 2009 Report,” June 14, 2010.

[25] “Transmission Availability Data System 2008 RFC Updated Report,” June 14, 2010.

[26] “Transmission Availability Data System, RFC – 2009 Report,” June 14, 2010.

[27] “Transmission Availability Data System 2008 SERC Updated Report,” June 14, 2010.

[28] “Transmission Availability Data System, SERC – 2009 Report,” June 14, 2010.

[29] “Transmission Availability Data System 2008 SPP Updated Report,” June 14, 2010.

[30] “Transmission Availability Data System, SPP – 2009 Report,” June 14, 2010.

[31] “Transmission Availability Data System 2008 WECC Updated Report,” June 14, 2010.

[32] “Transmission Availability Data System, WECC – 2009 Report,” June 14, 2010.

[33] “Transmission Availability Data System 2008 TRE Updated Report,” June 14, 2010.

[34] H. G. Saddock et al, “Common Mode Forced Outages of Overhead Transmission Lines,” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-95, No. 3, May/June, 1976.

[35] The PACME Working Group of the IEEE PES Reliability, Risk, and Probability Applications (RRPA) Subcommittee, “Effects of Dependent and Common Mode Outages on the Reliability of Bulk Electric System – Part II: Outage Data Analysis,” IEEE PES General Meeting, 2014.

[36] J.W. Hickman et al., “PRA Procedures Guide: A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants,” available online at http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr2300/vol1/cr2300v1-b.pdf.

[37] IEEE Standard Definitions for Use in Reporting Electric Generating Unit Reliability, Availability, and Productivity, IEEE Std 762™-2006.
[38] D. Lurie, L. Abramson, and J. Vail, “Applied Statistics,” NUREG-1745, March, 2011.
[39] The PACME Working Group of the IEEE PES Reliability, Risk, and Probability Applications (RRPA) Subcommittee, “Effects of Dependent and Common Mode Outages on the Reliability of Bulk Electric System – Part II: Outage Data Analysis,” IEEE PES General Meeting, 2014.
[40] M. Yue and T. L. Chu, “Probabilistic Reliability Modeling for Digital Component Using a Hierarchical Bayesian Method,” the Proceedings of PSAM 8, 2006.
[41] J. Widén C. Gallego, A. Cuerva, and A. Costa, “Detecting and Characterizing Ramp Events in Wind Power Time Series,” Journal of Physics, Conference Series 555 (2014) 012030.
[42] R. Sevlian and R. Rajagopal, “Wind Power Ramps: Detection and Statistics,” *Proc. 2012 IEEE Power Engineering Society General Meeting*
[43] M Cui et al, “An Optimized Swinging Door Algorithm for Wind Power Ramp Event Detection,” *IEEE Transactions on Sustainable Energy*, Vol. 7, No. 1, Jan. 2016.
[44] https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
[45] M.C. Alexiadis, P.S. Dokopoulos, and H.S. Sahsamanoglou, Wind speed and power forecasting based on spatial correlation models, *IEEE Transactions on Energy Conversion*, vol. 14, no. 3, pp. 836-842, 1999.
[46] M. Yue, “An Enhanced Probabilistic Contingency Analysis for Transmission Planning under High Penetration of Renewables: Data and Tool,” *Proc. 2017 IEEE Power Engineering Society Generation Meeting.*
[47] J.W. Hickman et al., “PRA Procedures Guide: A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants,” available online at http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr2300/vol1/cr2300v1-b.pdf.
[48] Sandra Bums and George Gross, “Value of Service Reliability,” IEEE Transactions on Power Systems, Vol. 5, No. 3, August 1990.
[49] W. Li and P. Choudhury, “Probabilistic Transmission Planning,” IEEE Power & Energy Magazine, September/October 2007.
[50] R. Allan and R. Billinton, “Probabilistic Assessment of Power Systems,” Proceedings of the IEEE, Vol. 88, No. 2, February 2000.
[51] R. Barlow and F. Proschan, Statistical Theory of Reliability and Life Testing: Probability Models, Holt, Reinhart and Winston, Inc. 1981.
[52] L. H. Fink and K. Carlsen, “Operating under stress and strain,” IEEE Spectrum, Vol. 15, No. 3, March 1978.
[53] [5] Siemens Industry Inc., “Release Notes for PSS/E 33.4,” March 1, 2013.

9. Appendix

A snapshot of each of the input data files for the enhanced PCA study is given here to better understand the format and contents of these files.

- **Outage Statistics Data File**

A snapshot of the Outages Statistics Data file (*.prb) is given in Figure A.1. Note that in the following snapshots, the numbers in gray on the left-hand side are the index of the line number but not the contents of the file. The *.prb file follows the format of the PSS/E Outage Statistics Data File. A detailed explanation of this data file is available in Section 6.15.2 of the PSS/E 33.4 Program Operation Manual.

OutageStatistics0.prb

```
CONTINGENCY SGL_0 0.13060000 132.03264962
CONTINGENCY SGL_1 0.17415880 46.71026337
CONTINGENCY SGL_2 0.18766807 44.14453134
CONTINGENCY SGL_3 0.27837314 33.36675259
CONTINGENCY SGL_4 0.17415880 46.71026337
CONTINGENCY SGL_5 0.14328048 54.39159726
CONTINGENCY SGL_6 0.22240618 38.97817479
CONTINGENCY SGL_7 0.63926355 20.77797774
CONTINGENCY SGL_8 0.14328048 54.39159726
CONTINGENCY SGL_9 0.20503712 41.34252714
CONTINGENCY SGL_10 0.19345775 43.15462574
CONTINGENCY SGL_11 0.14328048 54.39159726
CONTINGENCY SGL_12 0.14521037 53.81580683
CONTINGENCY SGL_13 0.14328048 54.39159726
CONTINGENCY SGL_14 0.20696702 41.06022443
CONTINGENCY SGL_15 0.25714429 35.20768370
CONTINGENCY SGL_16 0.14328048 54.39159726
CONTINGENCY SGL_17 0.14521037 53.81580683
CONTINGENCY SGL_18 0.19731754 42.52696152
CONTINGENCY SGL_19 0.14328048 54.39159726
CONTINGENCY SGL_20 0.16064953 49.70750859
CONTINGENCY SGL_21 0.20696702 41.06022443
CONTINGENCY SGL_22 0.14328048 54.39159726
CONTINGENCY SGL_23 0.14328048 54.39159726
CONTINGENCY SGL_24 0.14328048 54.39159726
CONTINGENCY SGL_25 0.14328048 54.39159726
CONTINGENCY SGL_26 0.14328048 54.39159726
CONTINGENCY SGL_27 0.14328048 54.39159726
CONTINGENCY SGL_28 0.14328048 54.39159726
CONTINGENCY SGL_29 0.14328048 54.39159726
```

Figure A. 1: ***A snapshot of the Outages Statistics Data file (*.prb).***

- **Monitored Element Data File**

A snapshot of the Monitored Element Data file (*.mon) is given below. The *.mon file follows the format of the PSS/E Monitored Element Data File. A detailed explanation of this data file is available in Section 8.1.3 of the PSS/E 33.4 Program Operation Manual.

18HS4ap.mon

```
MONITOR VOLTAGE DEVIATION ALL BUSES 0.08 0.08
MONITOR VOLTAGE RANGE SUBSYSTEM VOLHL 0.9 1.1
END

```

*Figure A. 2: **A snapshot of the Monitored Element Data file (*.mon).***

- **Contingency Description Data File**

A snapshot of the Contingency Description Data file (*.con) is given below. This type of file follows the format of the PSS/E Contingency Description Data File. A detailed explanation of this data file is available in Section 8.1.4 of the PSS/E 33.4 Program Operation Manual.

18HS4ap_wind.con

```
CONTINGENCY SGL_0
OPEN LINE FROM BUS 24042 TO BUS 24646 TO BUS 29036 CKT 5
END
CONTINGENCY SGL_1
OPEN LINE FROM BUS 10009 TO BUS 10113 CKT 1
END
CONTINGENCY SGL_2
OPEN LINE FROM BUS 10012 TO BUS 10245 CKT 1
END
CONTINGENCY SGL_3
OPEN LINE FROM BUS 10024 TO BUS 10440 CKT 1
END
CONTINGENCY SGL_4
OPEN LINE FROM BUS 11280 TO BUS 11283 CKT 1
END
CONTINGENCY SGL_5
OPEN LINE FROM BUS 42509 TO BUS 42569 CKT 1
END
CONTINGENCY SGL_6
OPEN LINE FROM BUS 79255 TO BUS 79256 CKT 1
END
CONTINGENCY SGL_7
OPEN LINE FROM BUS 79264 TO BUS 79265 CKT 1
END
CONTINGENCY SGL_8
OPEN LINE FROM BUS 42509 TO BUS 42577 CKT 1
END
CONTINGENCY SGL_9
OPEN LINE FROM BUS 80017 TO BUS 80048 CKT 1
END
```

*Figure A. 3: **A snapshot of the Contingency Description Data file (*.con).***

- **Contngency Description File**

A snapshot of the *_out.con file generated by the API 'rank' is given below. The *_out.con file follows the format of the PSS/E Contingency Description Data File. A detailed explanation of this data file is available in Section 8.1.4 of the PSS/E 33.4 Program Operation Manual.

18HS4ap_out.con

```
CONTINGENCY 'UNIT 14932(1)'
REMOVE UNIT 1  FROM BUS  14932  / PI =   0.152448 FROM BUS 'PALOVRD2    24.000'
END
CONTINGENCY 'UNIT 14933(1)'
REMOVE UNIT 1  FROM BUS  14933  / PI =   0.152227 FROM BUS 'PALOVRD3    24.000'
END
CONTINGENCY 'UNIT 14931(1)'
REMOVE UNIT 1  FROM BUS  14931  / PI =   0.152116 FROM BUS 'PALOVRD1    24.000'
END
CONTINGENCY 'UNIT 40063(1)'
REMOVE UNIT 1  FROM BUS  40063  / PI =   0.127723 FROM BUS 'CGS         25.000'
END
CONTINGENCY 'UNIT 36411(1)'
REMOVE UNIT 1  FROM BUS  36411  / PI =   0.126585 FROM BUS 'DIABLO 1    25.000'
END
CONTINGENCY 'UNIT 36412(1)'
REMOVE UNIT 1  FROM BUS  36412  / PI =   0.126585 FROM BUS 'DIABLO 2    25.000'
END
CONTINGENCY 'UNIT 26039(1)'
REMOVE UNIT 1  FROM BUS  26039  / PI =   0.094664 FROM BUS 'INTERM1G    26.000'
END
CONTINGENCY 'UNIT 26040(2)'
REMOVE UNIT 2  FROM BUS  26040  / PI =   0.094664 FROM BUS 'INTERM2G    26.000'
END
CONTINGENCY 'UNIT 623503(1)'
REMOVE UNIT 1  FROM BUS 623503  / PI =   0.090188 FROM BUS 'COLSTRIP GN326.000'
END
CONTINGENCY 'UNIT 623504(1)'
REMOVE UNIT 1  FROM BUS 623504  / PI =   0.089110 FROM BUS 'COLSTRIP GN426.000'
END
```

*Figure A. 4: **A snapshot of the Contingency Description Data file (*_out.con).***

- **Intermittency Induced Outage Data File**

A snapshot of the *.iio file is given below. The *.iio file has four columns that represent machine id, bus number, Pmax in MW, and PGen in MW, respectively, where Pmax denotes the active power upper limit, and Pgen denotes the machine active power output. Note that the bus number is the original bus number instead of the new bus number, as explained in Section 5.3.

wind_penet01.iio

1	10246	132.0000	80.0000
1	10396	300.0000	290.0000
1	10394	150.0000	140.0000
1	10997	204.0000	10.0000
1	10395	150.0000	140.0000
1	12058	268.0000	218.7000
1	10485	110.0000	94.0000
1	14802	235.0000	230.0000
1	10486	149.0000	141.0000
1	14811	235.0000	230.0000
1	10903	143.0000	85.0000
1	14821	290.0000	269.9000
1	10909	102.0000	5.0000
1	14902	299.9000	271.0000
1	11135	142.0000	101.1630
1	15142	257.0000	250.0000
1	14800	160.0000	150.0000
1	15147	255.0000	280.0000
1	14801	160.0000	150.0000
1	15157	250.0000	245.0000
1	14809	160.0000	96.0000
1	15159	250.0000	245.0000
1	14810	160.0000	155.0000
1	15161	250.0000	245.0000
1	14900	119.6000	116.0000
1	15166	293.0000	277.0000
1	14924	113.6000	113.6000
1	15926	282.0000	280.5900
1	14925	113.6000	113.6000
1	16542	235.0000	230.0000
1	14946	140.0000	140.0000
3	18403	252.0000	171.6240
2	14947	140.0000	140.0000
1	18428	214.5000	213.0000
1	14966	173.7000	50.0000
1	18433	246.0000	245.0000
1	14967	173.7000	50.0000
1	18436	255.0000	254.0000
1	14968	183.7000	60.0000
3	19313	297.3000	285.0000
1	14974	154.7000	115.0000
8	20008	261.0000	211.0980
1	14975	154.7000	115.0000
9	20009	261.0000	203.0000
1	14976	154.7000	115.0000
4	20189	290.0000	235.0000
1	14977	154.7000	115.0000
1	22240	299.0000	270.0000
1	14982	177.7000	177.7000
1	22265	225.0000	134.2110

Figure A. 5: ***A snapshot of the *.iio file.***

- **Common Mode Outage Data File: Double Outages**

A snapshot of the *.cmf2 file is given below. The *.cmf2 file has nine columns that represent correlation, machine id, bus number, Pmax in MW, PGen in MW, machine id, bus number, Pmax in MW, and PGen in MW, respectively. Note that the bus number is the original bus number instead of the new bus number, as explained in Section 5.3.

wind_penet01.cmf2

0.8	1	10246	132.0000	80.0000	1	10396	300.0000	290.0000
0.8	1	10394	150.0000	140.0000	1	10997	204.0000	10.0000
0.8	1	10395	150.0000	140.0000	1	12058	268.0000	218.7000
0.8	1	10485	110.0000	94.0000	1	14802	235.0000	230.0000
0.8	1	10486	149.0000	141.0000	1	14811	235.0000	230.0000
0.8	1	10903	143.0000	85.0000	1	14821	290.0000	269.9000
0.8	1	10909	102.0000	5.0000	1	14902	299.9000	271.0000
0.8	1	11135	142.0000	101.1630	1	15142	257.0000	250.0000
0.8	1	14800	160.0000	150.0000	1	15147	255.0000	280.0000
0.8	1	14801	160.0000	150.0000	1	15157	250.0000	245.0000
0.8	1	14809	160.0000	96.0000	1	15159	250.0000	245.0000
0.8	1	14810	160.0000	155.0000	1	15161	250.0000	245.0000
0.8	1	14900	119.6000	116.0000	1	15166	293.0000	277.0000
0.8	1	14924	113.6000	113.6000	1	15926	282.0000	280.5900
0.8	1	14925	113.6000	113.6000	1	16542	235.0000	230.0000
0.8	1	14946	140.0000	140.0000	3	18403	252.0000	171.6240
0.8	2	14947	140.0000	140.0000	1	18428	214.5000	213.0000
0.8	1	14966	173.7000	50.0000	1	18433	246.0000	245.0000
0.8	1	14967	173.7000	50.0000	1	18436	255.0000	254.0000
0.8	1	14968	183.7000	60.0000	3	19313	297.3000	285.0000
0.6	1	14974	154.7000	115.0000	8	20008	261.0000	211.0980
0.6	1	14975	154.7000	115.0000	9	20009	261.0000	203.0000
0.6	1	14976	154.7000	115.0000	4	20189	290.0000	235.0000
0.6	1	14977	154.7000	115.0000	1	22240	299.0000	270.0000
0.6	1	14982	177.7000	177.7000	1	22265	225.0000	134.2110
0.6	1	14983	177.7000	177.7000	1	22607	272.0000	270.0000
0.6	1	15140	166.0000	160.0000	1	23440	206.7200	114.0000
0.6	1	15141	166.0000	160.0000	1	24319	207.0000	200.0000
0.6	1	15145	166.0000	160.0000	1	24923	202.0000	202.0000
0.6	1	15146	166.0000	160.0000	1	24926	202.0000	202.0000
0.6	1	15158	145.0000	140.0000	2	24967	250.0000	150.0000
0.6	1	15160	145.0000	140.0000	1	24969	300.0000	150.0000
0.6	1	15162	145.0000	140.0000	1	24976	250.0000	240.0000
0.6	1	15165	157.0000	100.0000	1	26004	270.0000	193.4080
0.6	1	15167	163.2000	150.0000	2	26005	270.0000	-3.0000
0.6	1	15168	163.2000	150.0000	3	26006	270.0000	-3.0000
0.6	1	15184	150.0000	100.0000	4	26007	270.0000	-3.0000
0.6	1	15902	119.0000	119.0000	1	26026	230.0000	200.0000
0.6	1	15918	103.5000	100.0000	8	26150	201.0000	175.0000
0.6	1	15919	151.0000	148.0000	8	26151	250.0000	189.0000

Figure A. 6: ***A snapshot of the *.cmf2 file.***

- **Common Mode Outage Data File: Tripple Outages**

A snapshot of the *.cmf3 file is given below. The *.cmf3 file has 13 columns that represent correlation, machine id, bus number, Pmax in MW, PGen in MW, machine id, bus number, Pmax in MW, PGen in

MW, machine id, bus number, Pmax in MW, and PGen in MW, respectively. Note that the bus number is the original bus number instead of the new bus number, as explained in Section 5.3.

wind_penet01.cmf3

```
0.8 1  15927  153.5000  150.0000  4  26157  216.0000  209.0000 1  15928  151.0000  150.0000
0.8 ST  26234  224.4000  200.0000  1  15929  153.0000  150.0000 7  29902  205.0000  200.0000
0.8 1  15930  128.0000  125.0000  5  29904  205.0000  200.0000 1  15934  119.0000  95.0000
0.6 1  32900  240.0000  232.7800  1  16503  156.0000  152.0000 1  33108  215.0000  186.0400
0.6 1  16509  104.0000  100.0000  1  33109  215.0000  186.0400 1  16540  160.0000  155.0000
0.6 1  33110  215.0000  186.0400  1  16541  160.0000  155.0000 1  33113  249.0000  229.6200
```

Figure A. 7: ***A snapshot of the *.cmf3 file.***

The probability density function (pdf) of the lognormal distribution is

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{\frac{-(\ln x - \mu)^2}{2\sigma^2}\right\} \quad ; \quad x > 0$$

The mean and variance of the lognormal are defined as follows:

$$\text{mean} = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$\text{var} = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

As an illustration, a sample lognormal distribution is plotted in Figure 9.8 where μ=log(20,000) and σ= 1. The mean value is 33,000, as indicated by the vertical red line in the figure.

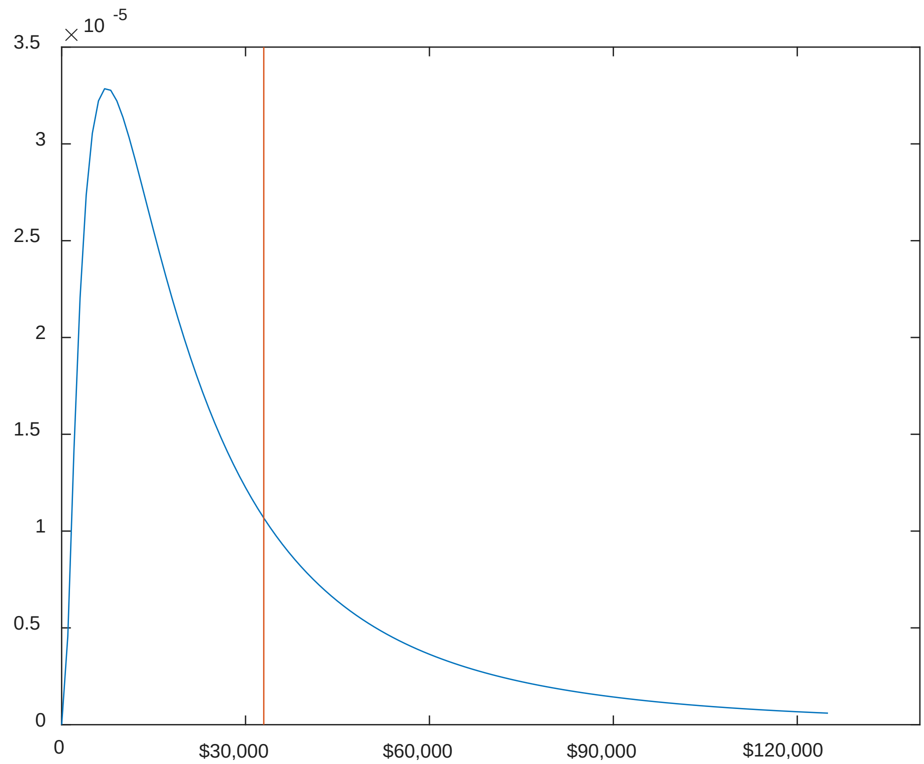

Figure 9.8: Lognormal Distribution probability density function (PDF).

Set O mentioned in Section 6.2.2 contains the following bus numbers from the system:
{12050, 12081, 12091, 14014, 14203, 14229, 14350, 14352, 14353, 14358, 14359, 14362, 14402, 14403, 14408, 14409, 14410, 16114, 19230, 19231, 19310, 19311, 19312, 19313, 19314, 20577, 22361, 22911, 23310, 23311, 23320, 23349, 23350, 23351, 23352, 24273, 24274, 24275, 24276, 24277, 24278, 24306, 24310, 24315, 24317, 24318, 24357, 24805, 24900, 25082, 25117, 25118, 25121, 26960, 26961, 26963, 26970, 27001, 27011, 27018, 27020, 27021, 27031, 29552, 30520, 31087, 31103, 31729, 31733, 32301, 32411, 33118, 33119, 33120, 33629, 33631, 34651, 35067, 35206, 36209, 36866, 37959, 38131, 40084, 40460, 43046, 43048, 43049, 43081, 43082, 43123, 43269, 43485, 43537, 43557, 43558, 43560, 43990, 45539, 46184, 46186, 47447, 47840, 47921, 47924, 47932, 47933, 47948, 47952, 47956, 50197, 50371, 50907, 51176, 51193, 54204, 54282, 54348, 54369, 54405, 54434, 54435, 54666, 54676, 54733, 54767,

54777, 54778, 54878, 54879, 54902, 54920, 54949, 54974, 55041, 55044, 55050, 55052, 55080, 55090, 55125, 55200, 55205, 55206, 55208, 55211, 55212, 55213, 55215, 55216, 55217, 55220, 55221, 55223, 55225, 55236, 55245, 55262, 55263, 55274, 55276, 55277, 55286, 55288, 55289, 55402, 55462, 55605, 55610, 55626, 55655, 55669, 55670, 55677, 55680, 55688, 55689, 55696, 56214, 56291, 57218, 57233, 57612, 59218, 59221, 59602, 59603, 59605, 59741, 64330, 64568, 64569, 64840, 73620, 73621, 84111, 84359, 84602, 84618, 84723, 84822, 84833, 84834, 84840, 84848, 84850, 84860, 84861, 84873, 84880, 84885, 84901, 84908, 84910, 85727, 85728, 85729, 85737, 85738, 85739, 85740, 85744, 85855, 85900, 85903, 85904, 85905, 85906, 85907, 85908, 85909, 85910, 85911, 85913, 85914, 85920, 85934, 85935, 85960, 85961, 85965, 85985, 85997, 300251, 622515, 622521, 622701, 622702, 622703, 622704, 622714, 622715, 623551, 624251, 624331, 624531, 624551, 626902, 627201, 627202}.

Pseudo Codes for N-1 and N-2 Contingency File Generation

The pseudo codes to generate N-1 and N-2 contingencies in *.con and *.prb files are given below.

```
For i in ID_list
   If total generation deviation of contingency i >= threshold value
      Write label & content of contingency i to *.con file.
      Write label, frequency, and duration of contingency i to *.prb file
For i1 in ID_list
   For i2 in ID_list[i1+1:end]
     If total generation deviation of contingencies i1 & i2 >= threshold value
        If wind farms involved in contingencies i1 and i2 have no overlap
            Contingency i1_i2 consists of contingencies i1 & i2
            Calculate the frequency & duration of contingency i1_i2 based on the
         frequency & duration of contingencies i1 & i2
            Write label & content of contingency i1_i2 to *.con file.
            Write label, frequency, and duration of contingency i1_i2 to *.prb file
```