PNNL-28130

Pacific Northwest NATIONAL LABORATORY

Probabilistic Look-ahead Contingency Analysis Integrated with GE EMS Tool

September 2018

Yousu Chen
Huiying Ren
Jun Wu (GE Grid Solutions)
Zhangshuan Hou

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical
Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-28130

Probabilistic Look-ahead Contingency Analysis Integrated with GE EMS Tool

September 2018

Yousu Chen
Huiying Ren
Jun Wu (GE Grid Solutions)
Zhangshuan Hou

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Abstract

This report documents the implementation of an advanced smart sampling-based probabilistic look-ahead contingency analysis algorithm, the integration with General Electric (GE) Grid Solutions' commercial energy management system (EMS) tool, along with the case study results based on real-world data from the Bonneville Power Administration (BPA).

The probabilistic look-ahead contingency analysis algorithm goes beyond traditional deterministic approaches by incorporating forecast errors of renewable energy and load into contingency analysis functions. This enables a predictive capability, providing a more comprehensive understanding and revealing the potential violations that may not detected by conventional methods.

To assess its performance under practical environments, the probabilistic look-ahead contingency analysis algorithm has been seamlessly integrated with the dominant GE EMS tool as a proof-of-concept. Adapting an extreme value distribution (EVD) algorithm ensures compatibility with the GE EMS tool's violation-only output format. Real-world data from BPA were used to test its performance, and the results clearly demonstrated the effectiveness of the developed algorithm, enhancing situational awareness.

This collaboration brings together the Department of Energy, a national laboratory (PNNL), a commercial vendor (General Electric), and a major utility (Bonneville Power Administration). It serves as a notable example of successful evaluating a lab-developed research tool in a commercial tool environment using real-world models and data. This work emphasizes the effectiveness of the developed algorithm, indicates a path for technology transition, and underscores the importance of collaboration with diverse team members.

Acknowledgments

This project is funded by the U.S. Department of Energy Office of Electricity Delivery and Energy Reliability (DOE-OE) Advanced Grid Modeling (AGM) program and the Bonneville Power Administration Technology Innovation (BPA TI) program. The project team wants to especially thank Mr. Gil Bindewald, Dr. Ali Ghassemian, DOE-OE Program Managers and Mr. Gordon Matthews and Dr. Thong Tring from BPA for their continuing support. The project would also not have been possible without the significant in-kind cost share provided by GE Grid Solutions.

The project team appreciates technical support from GE Grid Solutions staff Dr. Sajjad Sheikh, Dr. Weiguo Wang, and Mr. Christopher Johnson.

The project team appreciates contributions of the following Pacific Northwest National Laboratory staff:

- Dr. Henry Huang, Analytics Subsector Manager
- Dr. Frank Tuffner, Peer Reviewer
- Mr. Dale King, Project Management Office Director
- Mr. Carl Imhoff, Electricity Infrastructure Market Sector Manager
- Dr. Tony Nguyen, Power System Engineer
- Dr. Pavel Etingov, Power System Engineer
- Mr. Greg Dayley, Computer and Software Support
- Mr. Jeff Banning, Computer and Software Support
- Ms. Sue Arey, Project Specialist
- Ms. Mary Jiang, Project Support

Acronyms and Abbreviations

AGM	Advanced Grid Modeling
BPA TI	Bonneville Power Administration’s Technology Innovation
CA	Contingency Analysis
CDF	Cumulative Distribution Function
CTGS	Contingency Definition
DC	Derivative Classifier
DOE	U.S. Department of Energy
DPF	Decoupled Power Flow
EMS	Energy Management System
EVD	Extreme Value Distribution
GE	General Electric
GEV	Generalized Extreme Value
GRS	General Random Sampling
HPC	High-performance Computing
MCA	Massive Contingency Analysis
MCS	Monte Carlo Sampling
MPI	Message Passing Interface
NERC	North American Electric Reliability Corporation
NETMON	Network Monitor
OE	Office of Electricity Delivery and Energy Reliability
PDF	Probability Density Function
PNNL	Pacific Northwest National Laboratory
PWRFLOW	Power Flow
QMC	Quasi-Monte Carlo
RASMOM	Remedial Action Scheme Monitor
SMTNET	Multiple-timepoint Study
STNET	Steady State Network Study

Contents

Figures

Tables

1.0 Introduction

This section introduces a review of contemporary contingency analysis functions and explores the necessity for a probabilistic contingency analysis function. This is particularly crucial as a mechanism to help meet the requirements of more dynamic, stochastic grid operations and planning.

1.1 The role of contingency analysis

Contingency analysis (CA) is a critical energy management system (EMS) function that is widely used to evaluate the electric grid's status if various combinations of component failures occur in the grid. The outputs of CA provide information to assess grid health to ensure that there are no operating limit violations in the system. In North America, power companies are required to perform CA for grid operations and planning studies to meet the North American Electric Reliability Corporation (NERC) operation standards. Typically, a power company selects a list of contingencies to perform contingency analysis within its operational cycle, typically in a 5-minute cycle. The list can include an *N*-1 contingency that contains one, or multiple, device failures and/or an *N*-2 contingencies. The choice of *N*-1 and *N*-2, as well as the components selected, depends on the company's grid status. Contingency analysis is also critical for power system planning and market computation.

Today's contingency analysis functions in the control centers are dominated by deterministic methods. It uses state estimation outputs as the system states to perform power flow/contingency analysis studies. When there are violations, operators are required to take actions to maintain a system's reliability according to NERC standards. With the fast growth of power grids and deployment of newer, more variable technologies, power companies are under pressure to take more contingencies into consideration as the demand of power supply pushes boundary constraints of the grid. Therefore, a real-time contingency analysis application has been utilized, with the aid of advanced computing technology, such as the multi-thread technology used in General Electric (GE) Grid Solutions' Habitat EMS tools.

In addition to examining the current operational cycle, some companies also conduct look-ahead contingency analysis functions to foresee the system's health condition in next operational intervals, e.g., in next 1 hour with a 15-minute interval. Commercial vendors, such as GE Grid Solutions' Multiple-timepoint study (SMTNET) tool, have the look-ahead contingency analysis capability developed. However, they are using exact forecast values for renewable energies, as well as generation schedules, load forecast, and outage plans, as the inputs to perform contingency analysis; this is a deterministic approach. With the penetration of renewable generation technology, the need for look-ahead contingency analysis that incorporates forecast errors becomes necessary.

1.2 Need for probabilistic look-ahead contingency analysis

Uncertainties brought by renewable energy resources and loads can play a large role in contingency analysis and other EMS functions. However, these uncertainties have not been considered in today's control centers. Therefore, the outputs of contingency analysis might not be accurate because they do not consider the stochastic nature of the forecast in the computation. With the increasing capacity of renewable energy resources and system demands, the impacts of variable generation and load uncertainties cannot be ignored.

Some critical violations could be missed if traditional deterministic approach is used. Thus, these impacts should be studied to maintain system reliability, which requires the function of probabilistic look-ahead contingency analysis to provide a more complete picture to operators.

1.3 The objective of study

This report is the outcome of a joint effort supported by the U.S. Department of Energy (DOE) Office of Electricity Delivery and Energy Reliability (OE) Advanced Grid Modeling program and the Bonneville Power Administration's Technology Innovation (BPA TI) program. In fiscal year 2016 (FY16) and FY17, the Pacific Northwest National Laboratory (PNNL) developed a probabilistic look-ahead contingency analysis algorithm using smart sampling and probabilistic analysis approaches. These developments were tested with real BPA data with encouraging results. In FY18, PNNL continues to work with GE Grid Solutions to integrate the PNNL look-ahead contingency analysis algorithm with GE Grid Solutions' Habitat EMS tools. A prototype tool has been developed to validate the performance of this algorithm on the BPA system and show the advantages of the approach.

1.4 Report structure

The remainder of this report is organized as follows: Section 2 introduces the architecture of the probabilistic look-ahead contingency analysis algorithm that consists of smart sampling algorithm, high-performance computing (HPC) setup, and the extreme value decomposition algorithm. Section 3 briefly introduces the real-world data received from BPA. Section 4 presents how real-world data was extracted, mapped, and processed in this research. Section 5 shows the effort GE Grid Solutions made to integrate with PNNL's algorithm. Section 6 presents the case study results with the BPA data and system. Section 7 concludes the paper with a discussion of future work.

2.0 Probabilistic Look-ahead Contingency Analysis Approach

This section will describe the framework of the probabilistic look-ahead contingency analysis algorithm and the main functions of the smart sampling algorithm, the parallelization setup, and the extreme value distribution algorithm for contingency violation analysis.

2.1 Framework of probabilistic look-ahead contingency analysis

The framework of the probabilistic look-ahead contingency analysis is shown in Figure 2.1. This framework is developed based on an early version [1] that was implemented with an in-house massive contingency analysis (MCA) tool [2] [3]. In this report, the framework is modified to integrate the look-ahead contingency analysis algorithm with the GE Grid Solutions' Habitat EMS tool. The core function in this framework is a smart sampling approach (shown in blue color), which factors forecast errors into look-ahead contingency analysis applications. The inputs for smart sampling techniques are historical forecast information and historical actual information (shown in yellow color). The outputs of smart sampling techniques are a reduced and representative set of realizations (scenarios) that consider the forecast errors efficiently, which leads to a reduced computational burden. More information on smart sampling will be described later in this section.

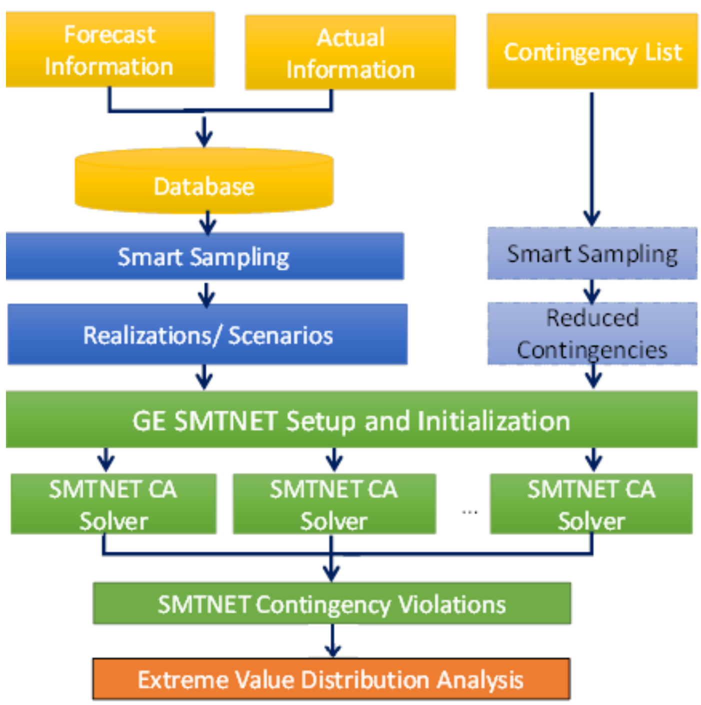


Figure 2.1 Probabilistic look-ahead contingency analysis methodology framework.

In addition to help address forecast uncertainties, the smart sampling algorithm can also be used on contingency lists, (i.e., a reduced set of contingencies, instead of full list, is assigned to each realization). These reduced and representative sets of realizations are then written in GE SMTNET format. SMTNET then reads this information in and perform multi-thread contingency analysis computations (shown in the green color in Figure 2.1). Unlike the previously utilized in-house tool, the SMTNET only outputs contingency violations. Therefore, the extreme value distribution (EVD) algorithm is adapted and applied to this framework (shown in the orange color in Figure 2.1 SMTNET) for branch violations.

2.2 Smart sampling algorithm

The purpose of the smart sampling algorithm is to address slow convergence of the traditional Monte Carlo methods with good coverage and less redundancy. This helps to generate a reduced and representative dataset that covers the forecast errors for look-ahead type of analysis that needs to involve forecast information.

Two sampling methods are introduced in this section: general random sampling (GRS)[4] and Quasi-Monte Carlo (QMC) [5-7]. The GRS method is a "standard" Monte-Carlo-based sampling and is described to help provide a basis to demonstrate the improvements of the QMC method. The QMC method is the smart sampling algorithm that has been integrated in the look-ahead contingency analysis framework.

Monte Carlo-based uncertainty analysis plays a central role in characterizing and quantifying uncertainty, with applications found in virtually all engineering fields [8], [9]. GRS, also known as Monte Carlo sampling (MCS), generates pseudo-random numbers as needed for trials. Typically, GRS involves three steps for each trial. First, random numbers, usually uniformly distributed between 0 and 1, are generated. Next, these numbers are transformed into the proper distribution for each variable. Finally, the values are plugged into the function to obtain the result. Fast random number generators exist, making the first step efficient. The computation times for the second and third steps depend on the complexity of the distributions and the function. The number of random numbers required for each trial can also vary [10, 11].

Based on a *k*-dimensional unit cube $\mathrm{I^k} = [0,1]^{\mathrm{k}}$, and with a (multivariate) bounded integration function f on $\mathrm{I^k}$, the Monte Carlo approximation of integral of f over $\mathrm{I^k}$ is defined by

$$\int_{\mathrm{I^k}} \mathrm{f(x)dx} \approx \frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{f(X_i)}$$

where $(\mathrm{X_i})_{1\leq \mathrm{i} \leq \mathrm{n}}$ are independent random points from $\mathrm{I^k I^d}$. The strong law of large numbers almost completely ensures the convergence of the approximation [26].

However, due to the typically large number of samples required, Monte Carlo sampling is often computationally expensive and/or requires a significantly simplified model. In that respect, minimizing the number of required Monte Carlo runs with representative data characteristics is essential."

Quasi-Monte Carlo (QMC) integration is a method of numerical integration that operates similarly to standard Monte Carlo Integration but utilizes sequences of quasi-random numbers, which exhibit a more uniform behavior to compute the integral. Quasi-random numbers are algorithmically generated by computers and are akin to pseudo-random numbers. However, quasi-random numbers possess the additional important property of being deterministically chosen based on equally distributed sequences to help minimize errors. The main difference between Monte-Carlo method and QMC method is that we no longer use random points $(\mathrm{X_i})_{1\leq \mathrm{i} \leq \mathrm{n}}$, but deterministic points. In our study, the Sobol sequence [12] has been used to generate the quasi-random numbers. The Sobol sequence $X_n = (X_{n,1}; : : : ; X_{n,k})$ is generated from a set of binary functions of length ω bits ($\nu_{i,j}$ with $i = 1, \ldots, \omega$ and $j = 1, \ldots k$. $\nu_{i,j}$ is called direction numbers, which are numbers related to primitive polynomial over field [0, 1]. To generate the jth dimension, the primitive polynomial in dimension j is assumed to be

$$p_j(x) = x^q + a_1 x^{q-1} + a_2 x^{q-2} + \cdots + a_{q-1} x + 1$$

Then the following q-term recurrence relation on integers $(M_{i,j})_i$ is

$$M_{i,j} = 2a_1 M_{i-1,j} \oplus 2^2 a_1 M_{i-2,j} \oplus \ldots 2^q a_q M_{i-q,j} \oplus M_{i-q}. \text{ where } i > q.$$

This allows to compute direction numbers as

$$\nu_{i,j} = M_{i,j}/2^j$$

This recurrence is initialized by the set of arbitrary odd integers $\nu_{1,j} 2\omega$, ... , $\nu_{,j} 2^q \omega$, which are smaller than 2, …,2^q , respectively. The jth dimension of the nth term of the Sobol sequence is

$$x_{n,j} = b_1 \nu_{1,j} \oplus b_2 \nu_{2,j} \oplus \ldots \oplus \nu_{\omega,j},$$

where b_k are the bits of integer $n = \sum_{k=0}^{\omega-1} b_k \, 2_k$. [5-7]. The requirement is to use a different primitive polynomial in each dimension.

Figure 2.2 shows the comparison of samples generated using the two different sampling approaches (GRS and QMC). The red points are sampling points obtained by each approach for two input variables, and the blue points are the projections of these sampling points onto the two dimensions. An even distribution of blue dots along the variable axis (space-filling) is expected for a good sampling approach. Moreover, without prior information on the relative weights of the different parameters (or dimensions), it is vital that no two design points coincide when projected onto a lower number of dimensions (i.e., non-collapsing). The “space-filling” and “non-collapsing” properties are important criteria to evaluate the performance of sampling approaches.

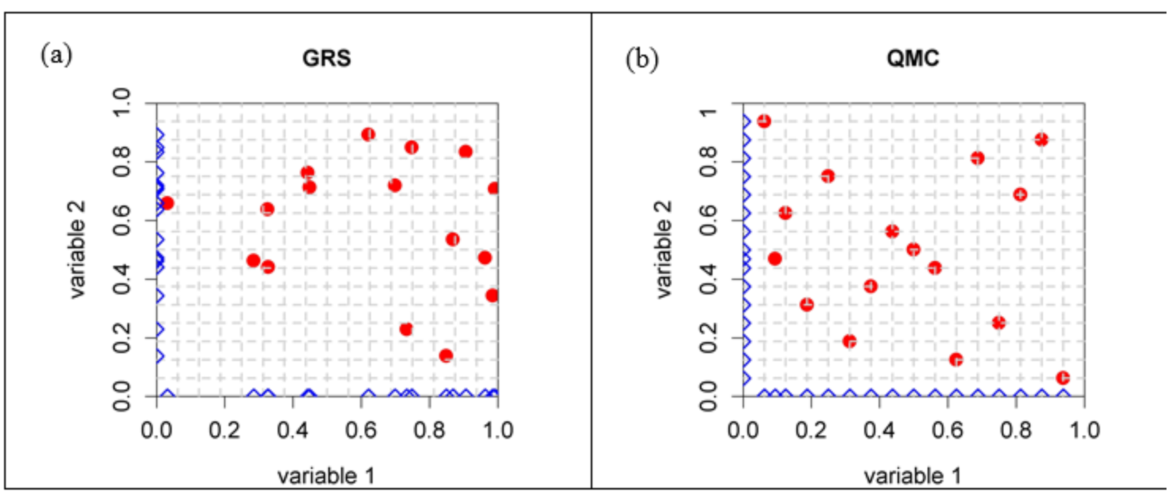


Figure 2.2: Sampling designs for GRS (a) and QMC (b). The red markers are the actual sampling points; the blue markers are their projections onto the two dimensions.

2.3 Parallelization setup

A dynamic load-balancing-scheme-based massive contingency analysis tool was employed in the original algorithm development [2]. However, applying the dynamic load balancing scheme to the GE Grid Solutions' commercial EMS tool is not straightforward. In particular, the EMS tool's built-in multithread option needs to be utilized for parallel computations, but this function can only be applied to contingency levels; the computation of each time point remains sequential. This limitation necessitates further modification to the GE Grid Solutions EMS tool to enable parallelization at the time point level to achieve scalability comparable to the original implementation in Chen et al. [1].

2.4 Probabilistic analysis algorithms (extreme value distribution)

The original probabilistic analysis algorithm is a probability density function (PDF)-based algorithm. This is enabled by the in-house massive contingency analysis tool providing full solutions to all base case and contingency scenarios [1] [2] [3].

However, the outputs of GE SMTNET look-ahead contingency analysis function contain only violation information, so the PDF-based approach cannot be utilized. Therefore, the extreme value distribution (EVD) algorithm has been innovatively adapted to this study.

Extreme value theory, which studies extreme deviations from the median of probability distributions in statistics, had been widely applied in a variety of areas of risk analysis [13] [14]. Extreme events occur when a risk takes values from the tail of its possibility distribution.

Suppose X_1, X_2 , X_3… , are independent random variables with the common cumulative distribution function (CDF) *F*, and a new dataset, $M_n = \max(X_1, X_2, X_3 \dots X_n)$, can be created by including only the first *n* maximum values. This dataset can only be represented by extreme value distribution models. In statistics, the Fisher-Tippett-Gnedenko theorem is a general result in extreme value theory for the behavior of the maxima of independent, identically distributed (iid) sequences. The continuous probability distribution based on this theory is a generalized extreme value (GEV) distribution. This continuous version is developed by combining Gumbel [15], Fréchet [16] and Weibull [17] distribution families. The single GEV distribution can be represented using three parameters[18] [19]:

$$\mathrm{F}(x;\ \mu, \xi, \varphi) = exp\left\{-\left(1 + \xi\frac{x-\mu}{\varphi}\right)^{-\frac{1}{\xi}}\right\},$$

where μ is a location parameter, φ is a scale parameter and ξ is the shape parameter. The three parameters can subdivide into three separate distribution families. When shape parameter $\xi = 0$, the GEV corresponds to the Gumbel distribution, also called type I extreme value distribution. This represents the exponential tail, $\xi > 0$ to Fréchet distribution, also called type II extreme value distribution. Such a type II distribution is associated with the long-tailed cases. Setting parameter $\xi < 0$ corresponds to Weibull distribution, which is a type III extreme value distribution that is related to short-tailed case. The ranges of interest for the three extreme value distributions are different; Gumbel distribution is unlimited, while Fréchet distribution has a lower limit. A simplification to the Gumbel family is also often justified by the argument that there are many standard distributions of the $(X_1, X_2, X_3 \dots X_n)$ for which the Gumbel distribution is the appropriate limit for M_n [20].

In this study, the Gumbel distribution is used to analyze the distribution of the violation rate, which is defined as the percentage of power flows over the normal limits. Each power flow containing contingencies is considered as an extreme event. Because the probability distribution of violation is estimated with uncertainties, a 95% confident interval was used in the GEV distribution fitting to represent the model uncertainties.

3.0 Real-world Data Extraction

This section introduces the real-world data received from the Bonneville Power Administration (BPA). The data includes historical forecast data of wind and load, BPA operational cases in PowerWorld format and GE Grid Solution EMS format, as well as a BPA planning model.

3.1 Description of received real-world data

In the original implementation of the probabilistic contingency analysis methods, BPA kindly provided historical operational node-breaker models for 2014 in a 15-min interval structure. These were formatted in PowerWorld .aux format with different database versions. BPA also provided wind farm information and a planning model in PSS/E PTI format. The operational models are used to extract historical actual load and wind generation information. Extensive effort has been expended to prepare a mapping table between the operational models and planning model to identify the actual values at each wind farm. This historical actual data and the historical forecast data are available at [21].

Studies on the historical data were needed to apply the smart-sampling algorithm discussed in the previous section of the report. The studies resulted in realizations of the smart-sampling algorithm for testing the probabilistic look-ahead contingency analysis function using an in-house massive contingency analysis tool. The study results were very promising and provided meaningful value to the final users, as reported in Chen et al. [1].

To obtain more realistic simulation results using real data with a commercial tool, additional requests were submitted to BPA. These requests sought BPA-solved STNET (steady state network study) and RTCA save cases in GE Grid Solutions' format for selected days, including key modules such as NETMON (network monitor), CTGS (contingency definition), DPF (decoupled power flow), RASMOM (remedial action scheme monitor), and PWRFLOW (power flow). Additionally, outage files and resource plans were requested. Details about the BPA database definition and the size of all records were also requested to facilitate the transition from BPA's database (GE Habitat EMP 2.5 version) to the SMTNET function-capable EMP 3.1 and above versions. BPA provided base cases, three contingency save cases, and four RASMOM save cases, all of which were successfully solved in the STNET after mitigation.

3.1.1 BPA wind data

BPA also provided wind forecasting data for the same interval, which includes data from 33 wind farms at a 1-hour resolution spanning from March to December 2014. The actual wind data was obtained through the BPA operational models after data extraction and mapping to the planning model. Due to limited forecasting data for the wind farm "PTW" and unsuccessful mapping for the wind farm "TRW," they are excluded from our study. Consequently, a total of 31 wind farms were considered in the first mapping stage.

Additional mapping from the BPA system to the GE SMTNET system was necessary, but some bus numbers were not successfully mapped into the GE SMTNET model. Certain wind farms were represented by multiple station identifiers and unit identifiers in the GE SMTNET model. For these wind farms, the forecasted wind power was distributed to each bus based on the ratio of the individual generator capacity. A mapping table was created to store data from these 31 wind farms, including their corresponding bus numbers in the planning model, their mapping station and unit identifiers in the GE SMTNET model, as well as their capacities. The unmapped wind farms and bus numbers were disregarded.

3.1.2 BPA load data

The load area forecast for the BPA area was obtained from the website: http://transmission.bpa.gov/business/operations/wind/. This data was used with the actual load information extracted from operational models for creating a database for the smart sampling algorithm. The BPA load forecast data has a 5-minute resolution and its quality is high with only a few missing points.

3.2 The process of data extraction

There are two issues associated with the actual wind and load forecast data: (1) the sampling rate was inconsistent for the actual wind data and forecast data; and (2) there are gaps in the actual data because of unsolved state estimation cases. To tackle these issues, down-sampling techniques with gap filling were performed. Because the BPA load forecast has excellent data quality with a few missing points, linear interpolation is applied to fill the small gaps. Then down-sampling was applied to extract data in 15-minute resolution with time stamps at 0, 15, 30, 45 minutes for the total load.

To align the actual data and forecast wind data, the hourly wind forecast data was up-sampled (interpolated) to 15-minute resolution using a cubic spline interpolation algorithm [22]. From the two datasets, the down-sampled load forecast data served as the temporal base line, with wind data associated with the closest available load forecast timestamp. To match actual wind data with wind forecast data, the actual wind data at the bus level are summed to obtain an aggregated actual value. An example of processed forecast data at a 15-minute resolution, derived from the original hourly forecast data together with the closest matched actual wind data for a wind farm in April 2014, is shown in Figure 3.1.

Similar to the procedure above, the approach of finding the closest time stamp between the actual and forecast load data is used to match load forecast and actuals. The actual load data is available as a bus-level quantity. There are about 800 load buses in the BPA area. The forecast load is only for the total BPA area, so the actual load data is the summation of the individual actual load values. This aggregation will be treated as a single variable/time-series for the smart sampling algorithms. Once the smart sampling realizations are created, the total load will be re-distributed to each load bus based on a participation factor. As a starting point, a fixed participation factor obtained based on the current load composition is used to calculate the load in the future look-ahead intervals. In the future, the variance of participation factor at different time periods will be evaluated and the smart sampling algorithm can include such information in generating realizations. For load forecasting values, they match with actual load pretty well, except for those time stamps with missing data.

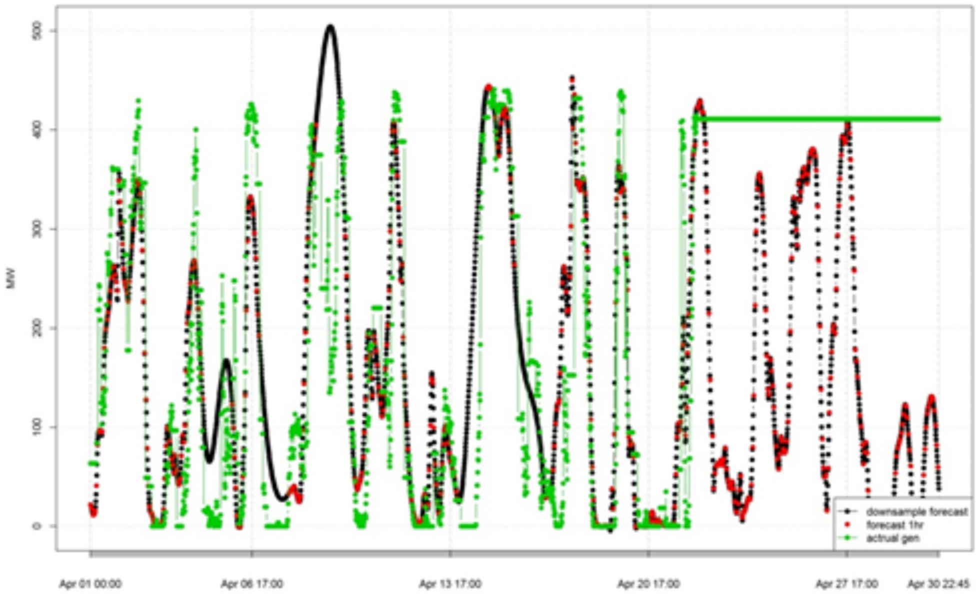


Figure 3.1: An example of down sampled forecast data, original forecast data and actual data for a wind farm in April.

4.0 Smart Sampling Application with Real-world Data

This section will discuss how the BPA wind and load data are selected to create realizations that are inputs for the GE Grid Solutions' SMTNET function and how the output data were processed by EVD algorithm.

4.1 Data selection

Applying the smart sampling techniques to the forecasted and actual wind data should ideally occur during a good data period in both data sets. The key is to find a good time period of data with the following considerations: (1) fewer missing data, (2) a good total wind power generation, and (3) a good range of wind power generation. After exploring the available data, a good time frame candidate for applying the smart sampling algorithm was between April 18th and April 20th. This 2-day data period was selected for the prediction and smart sampling realization generation that are discussed in the next subsections.

4.2 Realization creation

After matching and preprocessing all of the BPA wind and load data, 32 variables, including 31 wind farms and the total BPA load, were included in the smart sampling realization generation. These realizations were converted to the format that is suitable for the GE SMTNET model.

A sensitivity study based on the number of realizations and the efficiency of the QMC was conducted to find a reasonable number of required realizations. The sensitivity results are shown in Figure 4.1.

Based on a set of sensitivity analysis, 144 realizations were used as smart sampling outputs, considering the fact of 48 realizations for one SMTNET run (3 runs of 48 realizations each). The 144 realizations were divided into separate groups to facilitate task-parallel computing and mitigate a computing memory issue encountered with the software. Contingencies could also be sampled and saved to the GE SMTNET format.

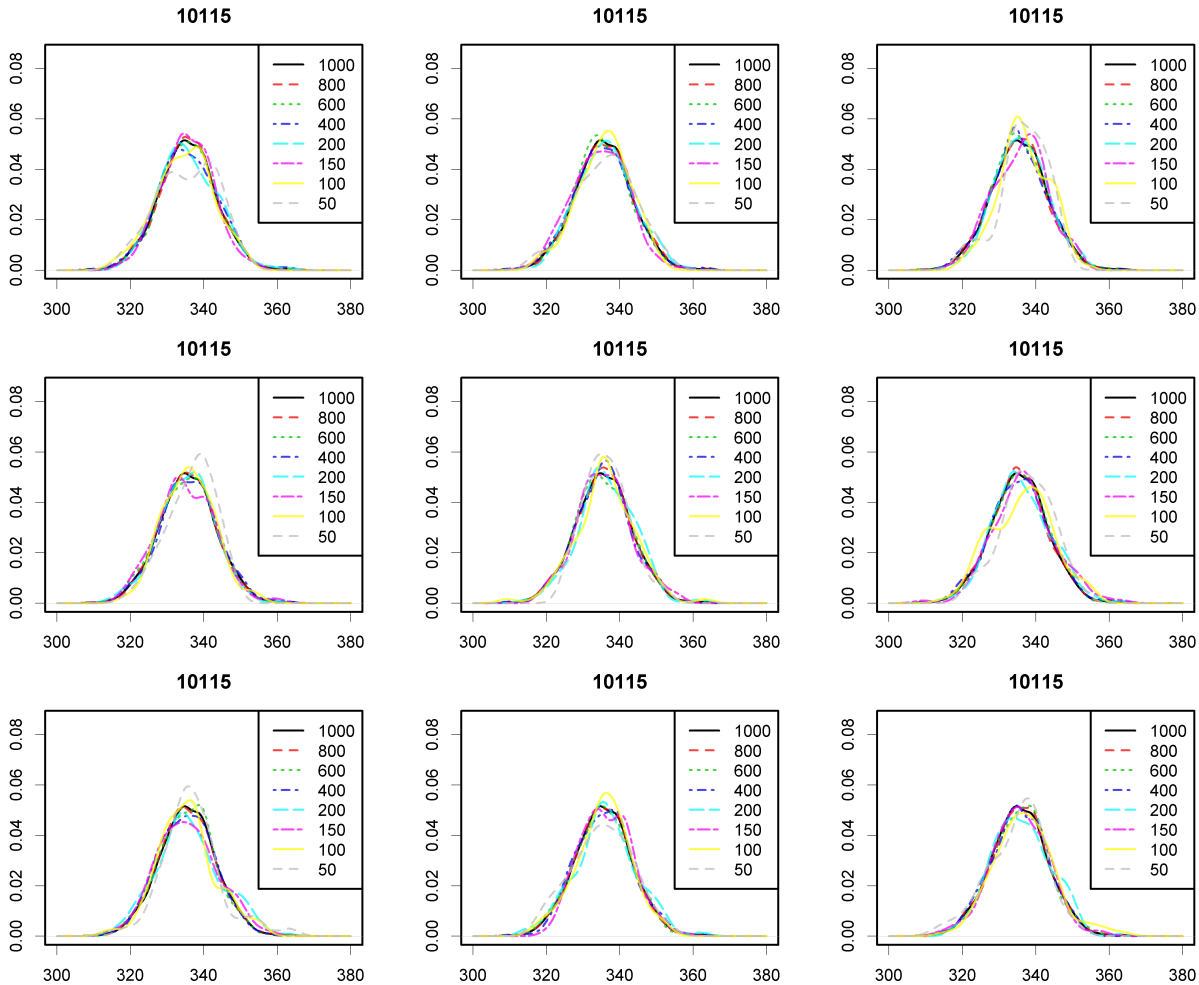


Figure 4.1: A sensitivity study results to find the number of realizations required for look-ahead contingency analysis study.

There were over 4,500 unit data inputs in a single SMTNET run (48 time intervals). An example of the range of these unit data can be seen at Figure 4.2.

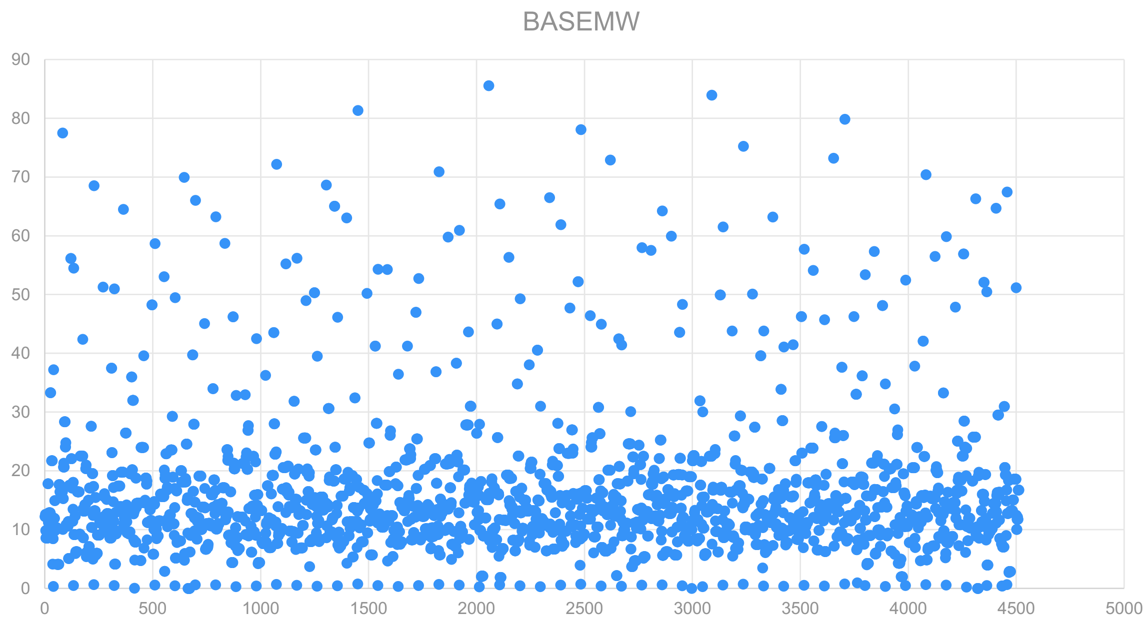


Figure 4.2: The range of wind generation data in one SMTNET run (48 time points).

4.3 Extreme value distribution analysis

Because GE SMTNET only outputs contingency analysis violations, the extreme valued distribution algorithm was applied to find the probability property of branch violations. As discussed earlier, this type of outputs fits a Gumbel distribution version of the generalized extreme value (GEV). This Gumbel fit was performed on the outputs from the GE SMTNET model. The fitting results were saved in .CSV files for each branch, which contains cumulative probability and violation rates in percentile, as well as the uncertainty ranges with 5% significant levels.

Beyond the Gumble distribution curve, frequency analysis was also performed for each branch at different realizations and contingencies. The frequency table was generated to represent the distribution of the power flows, including the bins/breaks and counts. Histograms can be obtained from the frequency tables to estimate the probability distributions of power flows at each branch. Examples are provided in Section 6.

5.0 GE SMTNET Modifications

The probabilistic look-ahead contingency algorithm is used to predict the power grid's future based on current system status and the forecast values provided at current time step. The GE Grid Solutions' EMS tool did not have the exact capability to perform this type of function. However, the SMTNET, multiple-timepoint study, function was available in GE EMS tool that can do the same work with a modification.

The SMTNET is a platform to perform a look-ahead security analysis for future time points given a set of inputs. It can run multiple time points with multiple network models and summarize violations for each time step. In this study, the realizations created by smart sampling algorithms can be considered as a new network model with different sizes of renewable generation and load.

Recall from earlier that the BPA operational cases needed to be converted from EMP 2.4 to EMP 3.1 or higher to utilize the SMTNET functionality. The conversion process is rather complex, involving database tuning and necessary validation. The first step was to resize the clones of NETMOM, PWRFLOW,CTGS, SMTNET etc. to enable GE's SMTNET/STNET application to run with the provided BPA realtime cases for 48 intervals. The reason for selecting 48 cases is mainly due to the consideration for optimal memory usage. The second step was to migrate BPA cases and execute them under the EMP 3.1 environment, aiming to get similar results to the original EMP 2.4.

5.1 Wind data modification

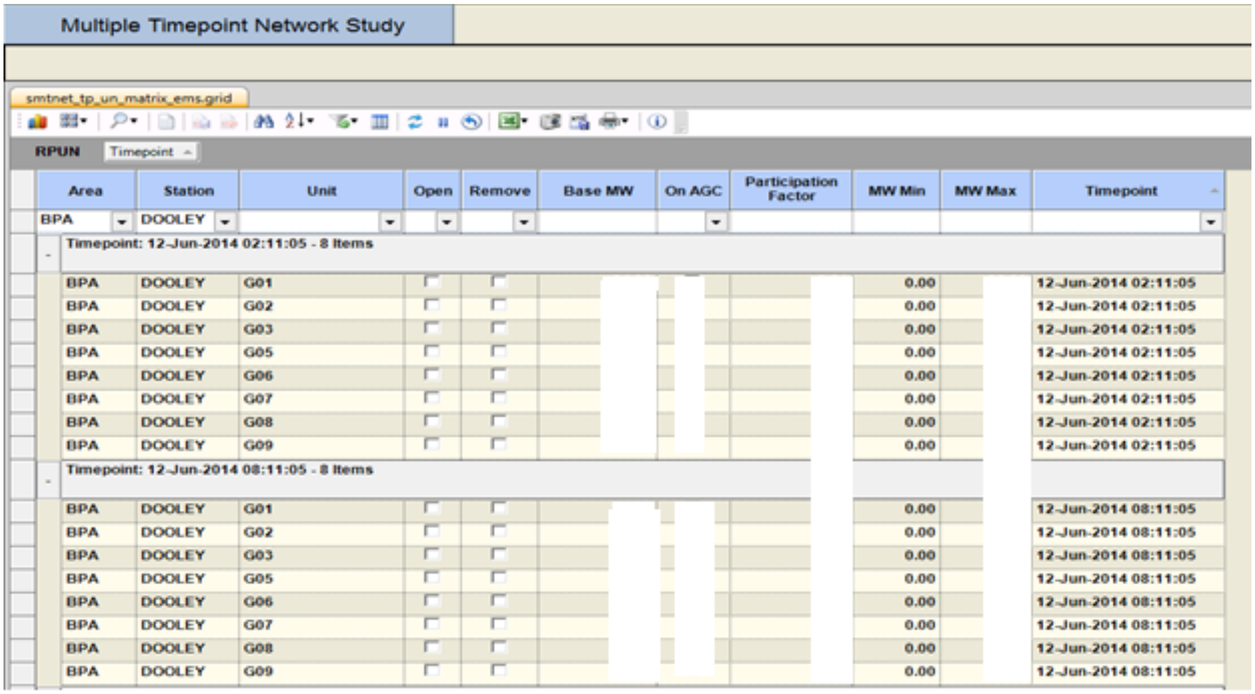
To integrate with the PNNL probabilistic look-ahead contignency analysis function, the SMTNET application was modified to read wind data for each study interval generated from smart sampling algorithms. These wind data overwrite existing wind data in the BPA base case and can be used as new base cases at the corresponding study interval/time points for power flow and contingency analysis.

The format of reading wind data is defined as:
I,UNITDATA,TIME,STATION_ID,UNIT_ID,BASEMW,WMN,WMX,ON AGC,PARTFACTOR

An example input is shown below:

I,UNITDATA,TIME,STATION_ID,UNIT_ID,BASEMW,WMN,WMX,ON AGC,PARTFACTOR
D,UNITDATA,01-Jan-2017 08:00:00,XXX,WND1,15.33,0,33,Y,0.33
D,UNITDATA,01-Jan-2017 08:00:00,XXX,WND2,15.33,0,33,Y,0.33
D,UNITDATA,01-Jan-2017 08:00:00,XXX,WND3,15.79,0,34,Y,0.34
D,UNITDATA,01-Jan-2017 08:00:00,YYY,G01,9.78,0,20,Y,1

A screenshot of wind unit data read in SMTNET is shown in Figure 5.1. For data protection reason, the generator information has been removed.



Multiple Timepoint Network Study

smtnet_tp_un_matrix_ems.grid

RPUN Timepoint

Area	Station	Unit	Open	Remove	Base MW	On AGC	Participation Factor	MW Min	MW Max	Timepoint
BPA	DOOLEY									
Timepoint: 12-Jun-2014 02:11:05 - 8 Items										
BPA	DOOLEY	G01						0.00		12-Jun-2014 02:11:05
BPA	DOOLEY	G02						0.00		12-Jun-2014 02:11:05
BPA	DOOLEY	G03						0.00		12-Jun-2014 02:11:05
BPA	DOOLEY	G05						0.00		12-Jun-2014 02:11:05
BPA	DOOLEY	G06						0.00		12-Jun-2014 02:11:05
BPA	DOOLEY	G07						0.00		12-Jun-2014 02:11:05
BPA	DOOLEY	G08						0.00		12-Jun-2014 02:11:05
BPA	DOOLEY	G09						0.00		12-Jun-2014 02:11:05
Timepoint: 12-Jun-2014 08:11:05 - 8 Items										
BPA	DOOLEY	G01						0.00		12-Jun-2014 08:11:05
BPA	DOOLEY	G02						0.00		12-Jun-2014 08:11:05
BPA	DOOLEY	G03						0.00		12-Jun-2014 08:11:05
BPA	DOOLEY	G05						0.00		12-Jun-2014 08:11:05
BPA	DOOLEY	G06						0.00		12-Jun-2014 08:11:05
BPA	DOOLEY	G07						0.00		12-Jun-2014 08:11:05
BPA	DOOLEY	G08						0.00		12-Jun-2014 08:11:05
BPA	DOOLEY	G09						0.00		12-Jun-2014 08:11:05

Figure 5.1: A screenshot of reading wind data at each time point in SMTNET (amended).

5.2 Load data modification

The SMTNET application was modified to read BPA load data for each study interval, also generated from the smart sampling algorithms described in prior sections. These load data overwrite existing load data in the BPA base case and can be used as new base cases at the corresponding study interval/time points for power flow and contingency analysis.

The format of reading load forecast data is defined as:
I,LDAREA,TIME,AREA_ID, MW_SCHEDULE

An example inputs is shown below:

I,LDAREA,TIME,AREA_ID, MW_SCHEDULE
D,LDAREA, 12-JUN-2014 02:11:05, BPA, 4500
D,LDAREA, 12-JAN-2014 08:11:05, BPA, 4800
D,LDAREA, 12-JAN-2014 14:11:05, BPA, 5000
D,LDAREA, 12-JAN-2014 20:11:05, BPA, 5200

A screenshot of wind unit data read in SMTNET is shown in Figure 5.2. For data protection reason, the generator information has been removed.

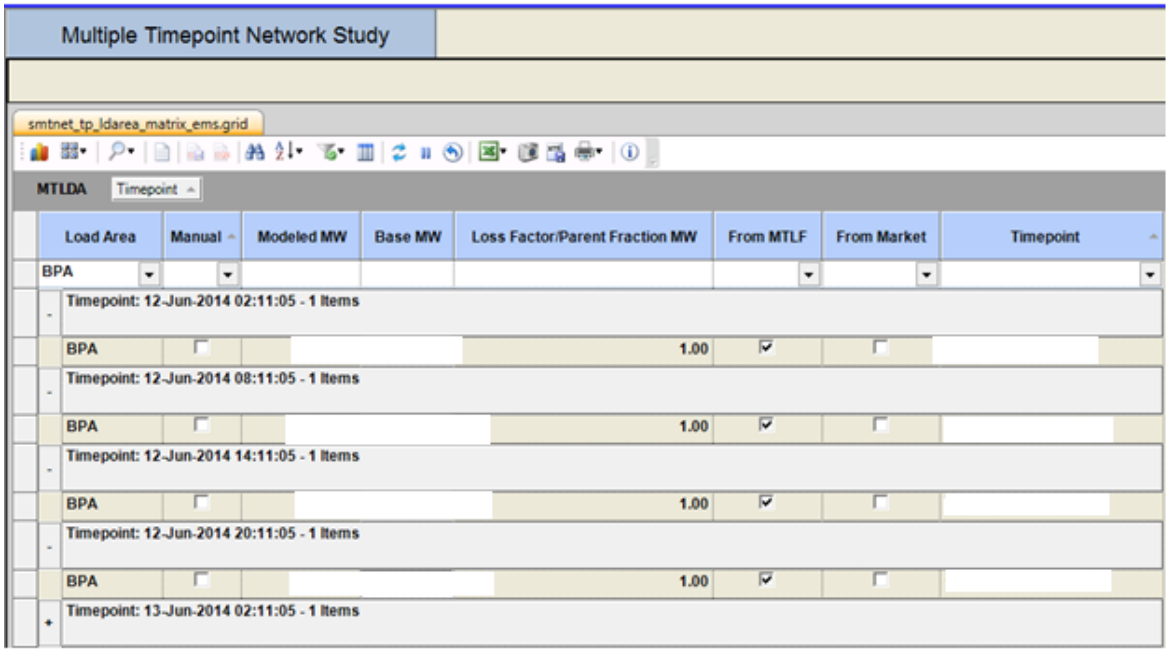


Figure 5.2: A screenshot of reading load data for each time point in SMTNET (amended).

5.3 Contingency data modification

To allow users to apply sampled contingencies to each base case for reduced computational burden, the original implementation also applied the smart sampling algorithm in contingencies. The idea is that instead of running all contingencies for each base cse, a user can run a sampled set of contingencies for each base case. Therefore, the SMTNET application was modified to allow user to activate a list of user-defined contingencies for each realization and deactivate the contingencies that are not in the list but are active in the base case.

An example of input format is show below:

Scenario,contingency ID
04-Jan-2017 04:40:00,03L11008
04-Jan-2017 04:40:00,03L11009
04-Jan-2017 04:40:00,03L11010

5.4 Output data modification

To analyze the outputs of SMTNET, it was modified to output branch flow violations for each contigency at each scenario for the extreme value distribution analysis. The outputs include:

- Branch flow for the violated branch in each contingency for each scenario.
- Normal limit for each violated branch violation in each contingency for each scenario.
- Emergency limit for each violated branch in each contingency for each scenario from sampling files.
- Load shed limit for each violated branch in each contingency for each scenario from sampling files.
- Monitor ID for each violated branch in each contingency for each scenario.

6.0 Performance Evaluation

This section is focused on the performance evaluation of the integrated look-ahead contingency analysis. The evaluation is mainly based on a comparison between the probabilistic-based algorithm and the existing deterministic approach, including a discussion on computational performance.

6.1 Evaluation environment

The GE Grid Solutions Habitat EMP 4.5 virtual machine was running on a server using Windows Server 2012 R2 standard. The processor is an Intel Xeon X7550 @ 2.0 GHz with four processors. The size of installed memory is 24 GB. There were 1000 realizations created by smart sampling algorithms that were used as a base case for a time point. Because SMTNET was setup to run simulation of 48 time points due to consideration of memory usage, a set of 21 unit input files was created for 21 SMTNET runs, each with different values of wind farm and total BPA area load. Each realization was set up to run with the full list of BPA 1,837 contingencies, or a selcted set of contingencies.

6.2 STNET base case contingency results

STNET is the GE EMS tool function used for studying network analysis in a steady state. To ensure a fair comparison, the BPA base case with the original contingency setup was performed in GE Grid Solutions' STNET application. This study focuses on the contingency analysis results rather than the base case results, as the same base case was used to both STNET and SMTNET, except the changes of wind units and BPA load introduced by the smart sampling algorithm. A screenshot of the base case contingency violation results is shown in Figure 6.1. There are 26 violations from 1,241 contingencies, including 18 branch violations. For data protection reason, the detailed violation information has been removed.

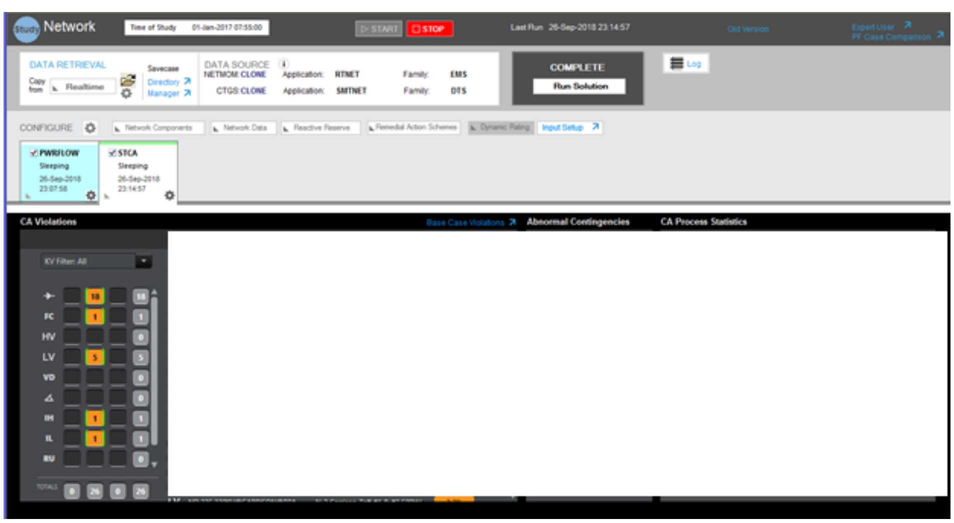


Figure 6.1: The screenshot of STCA results (amended).

The STCA recorded 19 instances of branch violations across 8 contingencies.

6.3 SMTNET results

Recall that SMTNET is a function in the GE EMS tool designed to perform a multiple time point study given a set of inputs. We leveraged this function for the PNNL probabilistic contingency analysis algorithm by assigning different wind units and load realizations to different time points. Subsequently, we conducted a probabilistic analysis on the outputs of SMTNET, focusing on contingency violations. A screenshot of SMTNET graphical user interface is shown in Figure 6.2.

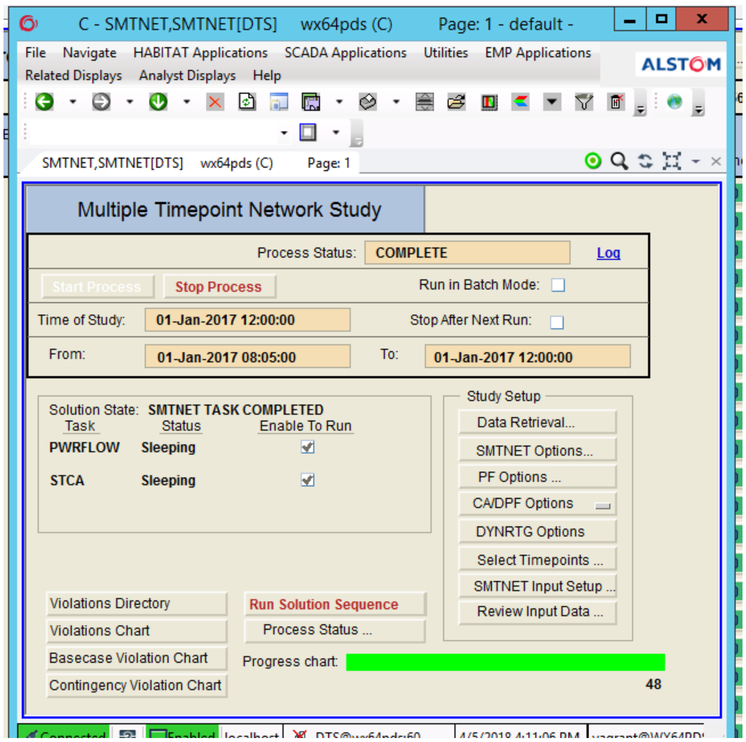


Figure 6.2: A screenshot of SMTNET Graphical User Interface.

In our study, a single SMTNET simulation involved 48 realizations with 1,241 contingencies for each realization, resulting in a total of 59,568 simulations per SMTNET run. Consequently, the probabilistic look-ahead contingency analysis function identified more contingency violations, such as branch flow violations and low voltage violations. For example, at the 08:00:00 time point scenario, the total number of contingency violations was 29, including 20 branch flow violations, compared to the base case, which had 26 violations, including 18 branch flow violations. This difference was caused by the variance of wind unit capacity and load due to forecast errors.

Each SMTNET run created a CA_OUT.TXT file that contains branch violations for each realization and contingency. An R script was developed to read this file and perform EVD analysis to present the probabilistic contingency violations to users. The R script also extracted the frequency and the worst violation for branch overflow violations. Three examples of EVD curve for line ID 455, 1061, and 2475 are shown in Figure 6.3, Figure 6.4, and

Figure 6.5, respectively. These figures provide user insights into the distribution of the branch violations, where the blue solid line is the Gumbel fitting curve, and the red dashed lines are uncertainty bounds with the Gumbel model.

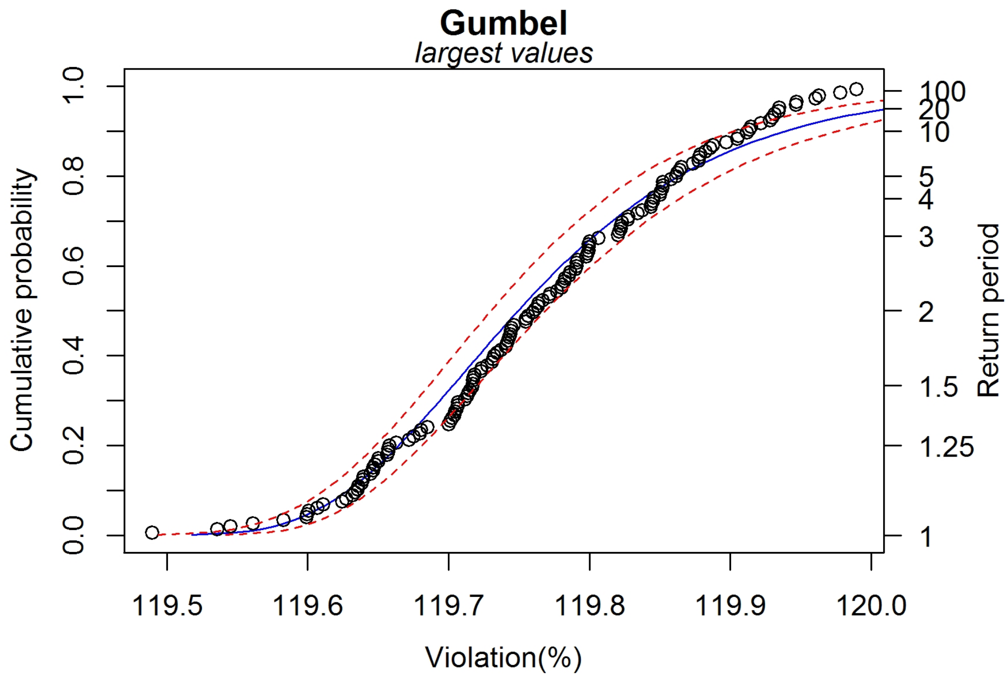


Figure 6.3: EVD Distribution for Line ID 455. The blue solid line is the Gumbel fitting curve; the red dash lines are uncertainty bounds with the Gumbel model.

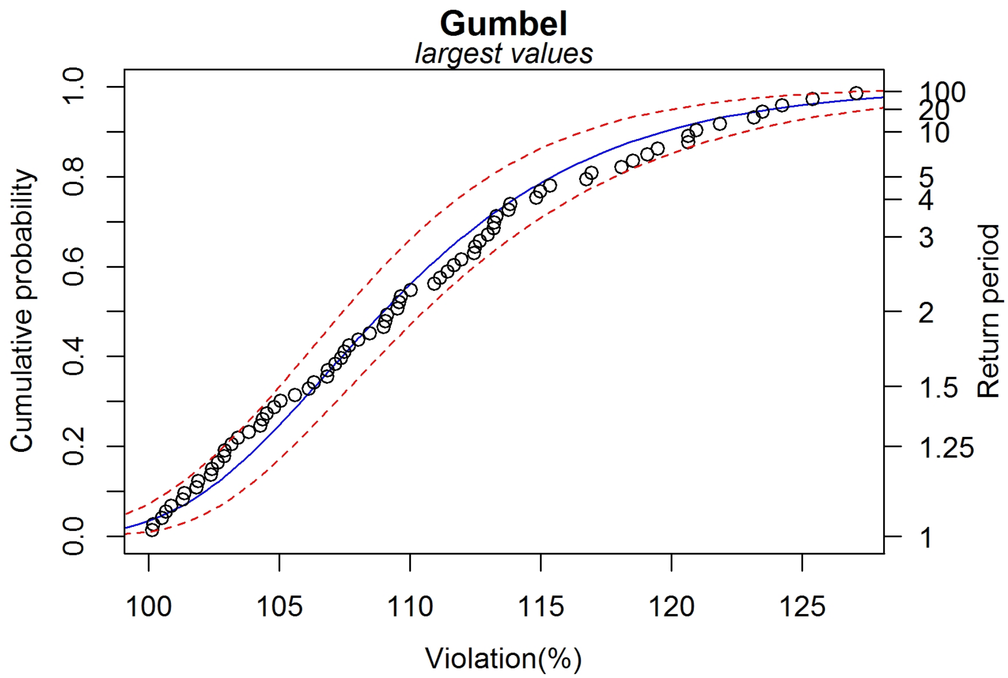


Figure 6.4: EVD Distribution for Line ID 1061. The blue solid line is the Gumbel fitting curve; the red dash lies are uncertainty bounds with Gumbel model.

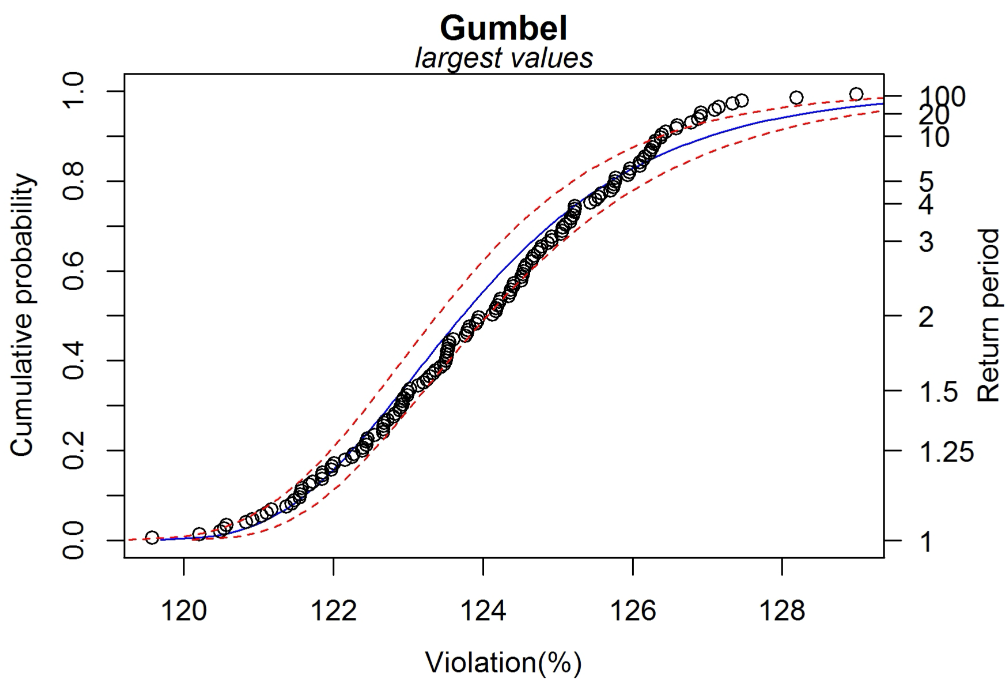


Figure 6.5: The EVD curve at line ID 2475. The blue solid line is the Gumbel fitting curve; the red dash lines are uncertainty bounds with the Gumbel model.

Table 6.1 summarizes the frequency of branch overflow violations, as well as the worst branch violations happened in all contingency simulations.

Table 6.1: The frequency and worst violations for the branches that contain violations.

Line ID	Frequency	Worst Violation (%)
1061	72	127.1
455	144	120.0
1815	288	135.8
1817	288	132.7
3101	144	105.9
3017	287	119.2
3099	267	117.0
1445	144	129.0
1485	144	115.4
1488	144	120.5
2474	144	124.3
2475	144	129.0
3602	144	154.4
3019	91	105.6
581	80	105.0

The probabilistic look-ahead contingency analysis function revealed additional violations not identified by the original deterministic method. The contingency analysis results revealed 21 branch violations. Upon comparing these against the base case violations, new violations happened at a transformer (associated with line ID 1061 in Table 6.1) for two distinct contingencies, with violation percentages of 113% and 112.5%, respectively. The total occurrence of these violations was 72. Hence, these specific contingency violations should not be overlooked. This discovery highlights the importance of studying new potential violations that go unnoticed in the deterministic method, providing valuable insights for grid operators.

6.4 Computational performance

The original development, utilizing an in-house massive contingency analysis function, demonstrated near-linear speedup, scaling efficiently to over 10,000 cores on a supercluster at PNNL's institutional computing machine [24-26]. However, the GE Grid Solutions EMS tool lacked a comparable capability and computational platform. Consequently, scalability testing was constrained to the SMTNET built-in multi-thread feature. As mentioned earlier, this multi-threading capability exclusively impacts contingency-level analysis. Consequently, simulations with 48 time points (realizations) remain sequential. Each time point took approximately 320 seconds to complete the contingency analysis simulation with 8 threads. The computational time for a full run with 48 time points was approximately 4.3 hours.

Another scalability test was conducted with a different number of active contingencies using six threads. Table 6.2 summarizes the contingency analysis performance study based on the number of contingency cases.

Table 6.2: Test results with different number of active contingencies with six threads.

# of Active Contingencies	Computational time (s)
20	29
65	40
188	96
322	143
391	174
508	220
849	228
1027	329
1171	370
1232	385
1443	461
1837	587

To further evaluate the scalability, a separate scalability test was conducted with different numbers of threads for 1,241 contingencies. The test results are shown in Table 6.3.

Table 6.3: The scalability testing with different numbers of threads for 1,241 contingencies.

Number of thread	Computation time(second)
1	550
2	435
4	388
6	321
8	323

Table 6.3 indicates that it took approximately 6 minutes to complete 1,241 contingencies using six threads. There was no additional gain observed when using more than six threads. The time improvement over a single thread was around 40%. In comparison to the scalability achieved with the in-house massive contingency analysis tool, multiple-thread techniques in the Windows environment have room for improvement.

7.0 Conclusion and Future Work

This project has shown a clear path to integrating the PNNL look-ahead contingency analysis algorithm with GE Grid Solutions' EMS tools. The reported work represents a preliminary study using real model data. The advantages of the developed algorithm have been demonstrated through case studies, revealing violations not previously detectable by traditional deterministic approaches, and providing better situational awareness for grid operation.

The original plan was to deploy the entire package, including GE Habitat EMS tool, on a PNNL cluster machine, so the computational power of the cluster could be utilized to expedite the time-intensive contingency analysis computation. This was expected to involve parallelizing the contingencies and time point through message passing interface (MPI)-based parallelization mechanisms, known for their efficiency compared to the multi-thread-based mechanism built in the GE EMS tool. The potential next step is to collaborate with GE to identify the path forward for implementing parallelization among time points using the MPI-based approach.

Visualizing the outputs of contingency analysis is another key aspect for future work. PNNL has developed a visual analytic platform capable of managing multidimensional data from various databases, including PI and Oracle. With assistance from GE Grid Solutions' help, PNNL has recently connected GE Habitat database to the PI data base. As a result, when SMTNET creates new outputs, they will be automatically stored in the PI data base. A web-based visualization tool can be developed to display the contingency outputs data for better situational awareness.

8.0 References

1. Chen, Y., P. Etingov, H. Ren, Z. Hou, M. Rice and Y Makarov. *A look-ahead probabilistic contingency analysis framework incorporating smart sampling techniques*. In 2016 IEEE *Power and Energy Society General Meeting (PESGM),* 10.1109/PESGM.2016.7741449, 2016. IEEE, Piscataway, NJ..
2. Huang, Z., Y. Chen, and J. Nieplocha. *Massive contingency analysis with high performance computing*. in *Power & Energy Society General Meeting, 2009. PES'09.* 009. IEEE, Piscataway, NJ..
3. Chen, Y., Z. Huang, and M. Rice. *Evaluation of counter-based dynamic load balancing schemes for massive contingency analysis on over 10,000 cores*. in *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*. 2012. IEEE, Piscataway, NJ.
4. Davis, M.W., *Production of conditional simulations via the LU triangular decomposition of the covariance matrix.* Mathematical Geology, 1987. **19**(2): p. 91-98.
5. Dutang, C. and D. Wuertz, *A note on random number generation*. 2009.
6. Niederreiter, H., *Quasi-Monte Carlo methods and pseudo-random numbers.* Bulletin of the American Mathematical Society, 1978. **84**(6): p. 957-1041.
7. Bratley, P. and B.L. Fox, *Algorithm 659: Implementing Sobol's quasirandom sequence generator.* ACM Transactions on Mathematical Software (TOMS), 1988. **14**(1): p. 88-100.
8. Mosegaard, K. and A. Tarantola, *Monte Carlo sampling of solutions to inverse problems.* J. geophys. Res, 1995. **100**(B7): p. 12431-12447.
9. Xu, C., H.S. He, Y. Hu, X. Li and R. Bu. *Latin hypercube sampling and geostatistical modeling of spatial uncertainty in a spatially explicit forest landscape model simulation.* Ecological Modelling, 2005. **185**(2): p. 255-269.
10. Zeeb, C.N. and P.J. Burns, *A Comparison of Failure Probability Estimates by Monte Carlo Sampling and Latin Hypercube Sampling.* Sandia National Laboratories, 1997.
11. Janssen, H., *Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence.* Reliability Engineering & System Safety, 2013. **109**: p. 123-132.
12. Antonov, I.A. and V. Saleev, *An economic method of computing LP τ-sequences.* USSR Computational Mathematics and Mathematical Physics, 1979. **19**(1): p. 252-256.
13. Kotz, S. and S. Nadarajah, *Extreme value distributions: theory and applications*. 2000: World Scientific.
14. Embrechts, P., S.I. Resnick, and G. Samorodnitsky, *Extreme value theory as a risk management tool.* North American Actuarial Journal, 1999. **3**(2): p. 30-41.
15. Gumbel, E.J., *Statistics of extremes*. 2012: Courier Corporation.
16. Bachelier, L., *Théorie de la spéculation*. 1900: Gauthier-Villars.
17. Weibull, W., *Wide applicability.* Journal of applied mechanics, 1951. **103**(730): p. 293-297.
18. Coles, S., et al., *An introduction to statistical modeling of extreme values*. Vol. 208. 2001: Springer.
19. Prescott, P. and A. Walden, *Maximum likeiihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples.* Journal of Statistical Computation and Simulation, 1983. **16**(3-4): p. 241-250.
20. Coles, S. and L. Pericchi, *Anticipating catastrophes through extreme value modelling.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 2003. **52**(4): p. 405-416.
21. *Data for BPA balancing authority total load, wind gen, wind forecast, hydro, thermal, and net interchange*. Available from: http://transmission.bpa.gov/business/operations/Wind/WindGenTotalLoadYTD_2014.xls.
22. Hazewinkel, M., *Spline interpolation.* Encyclopedia of Mathematics, 2001. 1.

Pacific Northwest
NATIONAL LABORATORY

www.pnnl.gov

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

U.S. DEPARTMENT OF
ENERGY