Probabilistic Power Flow Studies to Examine the Influence of Photovoltaic Generation on Transmission System Reliability

by

Miao Fan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2012 by the
Graduate Supervisory Committee:

Vijay Vittal, Chair
Gerald Thomas Heydt
Raja Ayyanar
Jennie Si

ARIZONA STATE UNIVERSITY

December 2012

ABSTRACT

Photovoltaic (PV) power generation has the potential to cause a significant impact on power system reliability since its total installed capacity is projected to increase at a significant rate. PV generation can be described as an intermittent and variable resource because its production is influenced by ever-changing environmental conditions.

The study in this dissertation focuses on the influence of PV generation on transmission system reliability. This is a concern because PV generation output is integrated into present power systems at various voltage levels and may significantly affect the power flow patterns. This dissertation applies a probabilistic power flow (PPF) algorithm to evaluate the influence of PV generation uncertainty on transmission system performance. A cumulant-based PPF algorithm suitable for large systems is used. Correlation among adjacent PV resources is considered. Three types of approximation expansions based on cumulants namely Gram-Charlier expansion, Edgeworth expansion and Cornish-Fisher expansion are compared, and their properties, advantages and deficiencies are discussed. Additionally, a novel probabilistic model of PV generation is developed to obtain the probability density function (PDF) of the PV generation production based on environmental conditions.

Besides, this dissertation proposes a novel PPF algorithm considering the conventional generation dispatching operation to balance PV generation uncertainties. It is prudent to include generation dispatch in the PPF algorithm since the dispatching strategy compensates for PV generation injections and influences the uncertainty results. Furthermore, this dissertation also proposes a probabilistic optimal power dispatching strat-

egy which considers uncertainty problems in the economic dispatch and optimizes the expected value of the total cost with the overload probability as a constraint.

The proposed PPF algorithm with the three expansions is compared with Monte Carlo simulations (MCS) with results for a 2497-bus representation of the Arizona area of the Western Electricity Coordinating Council (WECC) system. The PDFs of the bus voltages, line flows and slack bus production are computed, and are used to identify the confidence interval, the over limit probability and the expected over limit time of the objective variables. The proposed algorithm is of significant relevance to the operating and planning studies of the transmission systems with PV generation installed.

DEDICATION

To my father Guorui Fan, my mother Yixi Liu and my lovely girlfriend Xiao Liu, for their endless love and support through all these years.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation and thanks to my advisor, Dr. Vijay Vittal, for all of his significant guidance, help and encouragement in academics and the research development. I also thank him for generous financial support during my time at Arizona State University. I appreciate his skill and enthusiasm as a teacher and mentor.

I would like to thank Dr. Gerald Heydt. He gave me a lot of invaluable technical advice about the probabilistic power flow. The discussions with him enlightened me so much in my research.

Furthermore, I am grateful to Dr. Raja Ayyanar and Dr. Jennie Si for serving on my committee. They gave important insights into the research problem. Their comments and suggestions throughout the development of this work have been valuable and are greatly appreciated. I am very grateful for their time and commitment to me and my work. I also want to thank Xiaolin Mao, Yacine Chakhchoukh, and Hui Zhang. Their valuable suggestions contributed significantly to my research.

Finally, I would like to thank all my friends in the power and energy engineering group at Arizona State University for the joyful days we spent together.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

NOMENCLATURE

A	Total area of the PV module
a_{ik}	Sensitivity coefficients of sensitivity matrix S
ARMS	Average root mean square
AZ	Arizona
B_{ik}	Imaginary part of the admittance matrix of branch ik
b_{jk}	Sensitivity coefficients of sensitivity matrix L
CDF	Cumulative distribution function
$C_i(P_{gen,i})$	Generation cost function
c_i	The i^{th} constant coefficient of Gram-Charlier expansion
cov	covariance
FFT	Fast Fourier Transforms
$F(v)$	Cumulative distribution function of voltage variable v
$F(x)$	Cumulative distribution function of variable x
$F_{ij}(v_i)$	Cumulative distribution function of voltage v_i at bus i at hour j
$f_x(x)$	Probability density function of variable x
$f_y(y)$	Probability density function of variable y
$f_z(z)$	Probability density function of variable z
G	Jacobian matrix of power injection equations
G_{ik}	Real part of the admittance matrix of branch ik
$g(x)$	Power injection function

$H_i(x)$	The i^{th} Hermite polynomial
$h(x)$	Line flow function
i	Bus index
K	Sensitivity matrix of x
k	Temperature coefficient of PV generation
L	Sensitivity matrix of z
l	Number of branches
$L(v)$	Complementary distribution function of voltage variable v
$L_{ij}(v_i)$	Complementary distribution function of voltage v_i at bus i at hour j
MCS	Monte Carlo simulation
Ns	Number of sampling in MCS
n	Number of buses
P	active power output of PV generation
PDF	Probability density function
P_{gen}	Active power of conventional generation
P_i	Injected active powers at bus i
P_{ij}	Overvoltage probability of voltage v_i at bus i at hour j
PPF	Probabilistic power flow
P_{slack}	Active power of slack bus
$P_{system,j}$	Overvoltage probability of the power system at hour j
PV	Photovoltaic

$Q_{branch\ i}$	Line flow reactive power of the i^{th} branch connected with the slack bus
Q_i	Injected reactive powers at bus i
Q_{ik}	Reactive power of the line flow through branch ik
Q_{slack}	Reactive power of slack bus
R	$R = rA\eta$
r	Solar irradiance
$R_{\max}$	Maximum value of R
$r_{\max}$	Maximum solar irradiance
S_i	Apparent power of branch i
sl	Number of branches connected with the slack bus
SPF	Stochastic power flow
T	Sensitivity matrix of conventional generation
t_{ik}	t_{ik} is the transformation ratio of branch ik
WECC	Western Electricity Coordinating Council
x	State variable vector of bus voltages and angles
$x(q)$	Quantile of the probability q
z	Output variable vector of line flows
α_v	The v^{th} moment of distribution $F(x)$
β_v	The v^{th} central moment of distribution $F(x)$
κ_v	The v^{th} cumulant of $F(x)$

ΔP_{gen}	Uncertainty variable vector of conventional generation
ΔP_{PV}	Uncertainty variable vector of PV generation
ΔT	Uncertainty of PV cell temperature
Δx	Uncertainty variable vector of x
Δy	Uncertainty variable vector of y
Δz	Uncertainty variable vector of z
δ	Current sun declination
η	PV generation efficiency
ε	Residual of the linear model of PV generation production
θ_S	Solar elevation angle
θ_i	Voltage angle at bus i referred to slack bus
θ_{ik}	Difference in voltage angles between bus i and k
ρ	Correlation coefficient
σ	Standard deviation
ϕ	Local latitude
$\phi(x)$	Cumulative distribution function of standard normal distribution
$\varphi(x)$	Probability density function of standard normal distribution
$\xi(q)$	Quantile of the standard normal distribution
$\psi(t)$	Characteristic function of variable x
ω	Solar time

Chapter 1

INTRODUCTION

1.1 Overview

In power system operating and planning studies, the reliability of the system should be assessed and monitored. The uncertainty affects both the long and medium term system planning, and the day-ahead operation. Photovoltaic (PV) power generation has become an increasingly important renewable energy resource due to its ability to produce electric power at varying capacity levels and at varying voltage levels including both distributed generation and dedicated commercial sized plants. Some US state mandates have set a goal of 20% renewable generation by 2020 [1]. Increasing amounts of PV generation are beginning to have an impact on power transmission systems.

Compared to conventional generation, PV generation has several differing characteristics. First, PV generation is a variable resource because its production is influenced by ever-changing weather conditions, such as solar irradiance and temperature. Therefore, PV generation has large variation, which may influence power system operation and performance. As a result of this large variation, probabilistic studies are required for the steady state analysis of the power systems integrated with PV generation. Second, residential roof-top PV generation is not allowed to provide voltage control as mandated by IEEE Std 1547. As a consequence conditions could arise due to this restrictions that could result in voltage magnitude violations (see Appendix A). In addition, the output of PV resources at adjacent locations may be strongly statistically correlated owing to common effects such as insolation, temperature and other environmental factors. This correlation is considered and statistically characterized. Besides, the stochastic behavior

of PV generation usually does not follow a normal distribution. Due to these characteristics, PV generation can potentially cause various power quality issues in power systems [2]–[4], such as fluctuation of bus voltage magnitudes and line flows, voltage violations and unbalanced power flows. A typical PV generation production daily curve is shown in Fig. 1.1.

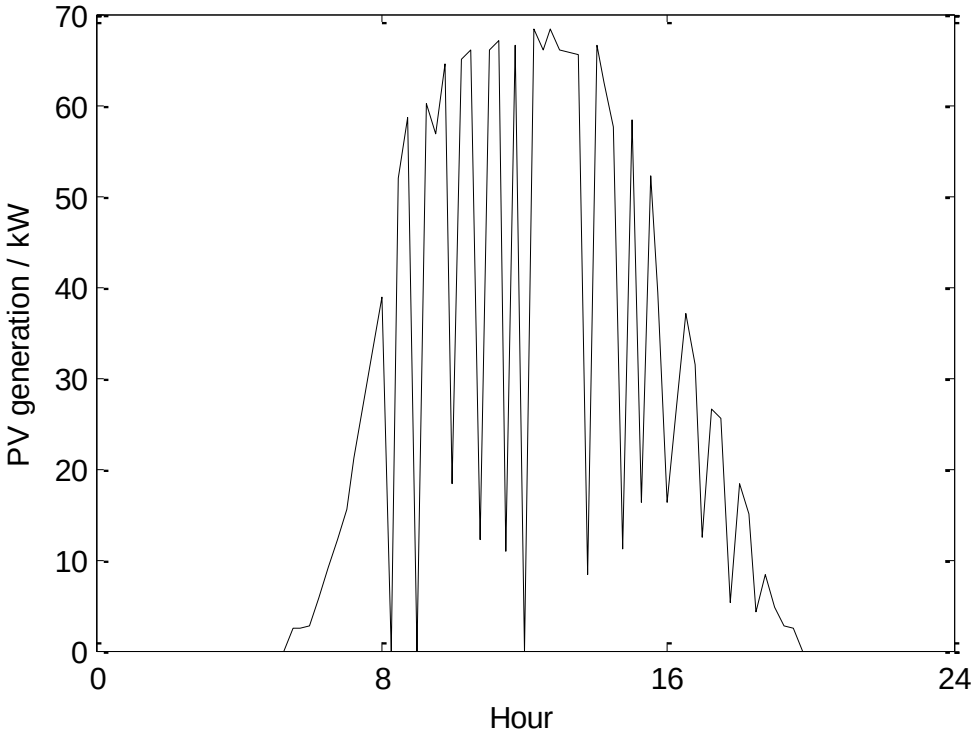

Fig. 1.1 PV generation output daily curve

One important aspect to analyze is the uncertainty of the power system with PV generation installed. Uncertainty deals with characterizing the possible distribution of the expected variable based on its historical statistical data [5]. The probabilistic power flow (PPF) algorithm has been successfully utilized for calculating the effect of the propagation of data inaccuracies through the power flow calculation, thus obtaining a range of values for each output quantity that, to a high degree of probability, bounds the operating

conditions of the system [5], [6]. This probabilistic algorithm takes into consideration the uncertainty of the system information and is able to provide strict bounds for the determined variables of the system. As an input this approach requires information regarding the range of all possible system conditions that might be encountered as a result of expected uncertainties of loads and other system variables. Assuming that the variable can exist anywhere within a precisely restricted interval, the variable is described to vary according to some probability distribution. Thus, the system variables can be characterized not by a single number but by a range of values [7].

Typical steady state studies always treat the peak power demands as the worst case conditions. Periods of light load are also critical in the assessment of the possible state of a power system. While heavy load conditions are generally associated with overload, low voltage and generation deficiency, light load conditions may give rise to overvoltage and undesirable reactive power requirements at generation sites. Thus the scheduling for the worst case scenario is not easily defined [8]. Therefore, the steady state analysis should not just consider the peak power demand cases.

Another important aspect of the power system with PV generation is that the system can adjust its total conventional generation to match the PV generation changes with automatic generation control (AGC) or the day-ahead generation scheduling settlement. In actual system operation, the influence of conventional generation should also be considered, since these resources are essential to balance the power supply and demand, and relieve the uncertainty problem. It is not a realistic assumption to let the slack bus generation balance the change of all the system power injections [43]. The system dispatch constraints should be taken into account to compensate for varying PV generation output

and to enhance the operational performance of power systems. This dispatching operation depends on the change of PV generation and the dispatching strategy. The impact of PV generation uncertainty is limited with the generation dispatching operation and should not be neglected in system analysis.

1.2 Literature Review

Compared to the deterministic power flow, the probabilistic method characterizes the uncertainty in system information by describing the variation in terms of a suitable probability distribution. Work related to the probabilistic analysis in power system power flow first appeared in 1974. Borkowska [6] first proposed the concept of PPF and implemented an algorithm based on convolution. In 1975, Dopazo, Klitin and Sasson created another method of applying the probabilistic approach to power system problems called stochastic power flow (SPF) [5]. Since then, there have been two primary probabilistic analytical approaches proposed, PPF [9]–[12] and SPF [13]–[16]. The inclusion of solar resources in the generation mix offers a potentially ideal application of these probabilistic algorithms.

Both the aforementioned methods apply the linearization model to bus voltages and line flows as a linear combination of load injections. References [9], [10] extended probabilistic analysis techniques to handle AC power flow by modifying the linearization formulation. The main difference between the two cited analytical methods, PPF and SPF, is the representation of input variables (power injections). In the SPF study, the input variables are represented as a mean value vector and a covariance matrix. While in the PPF study, the input variables are described by their probability distribution functions (PDFs). Comparing the two methods, the SPF method requires less computational time

and can provide a rough estimate of the uncertainty impact. In particular, the SPF method is suitable for the case of the input variables following a normal distribution. However, for general cases, the PPF method is more accurate for system operating studies, especially for systems with PV generation which is non-normally distributed. In this case the method obtains the PDFs of the result variables.

Monte Carlo simulation (MCS) [17] is a numerical method which involves repeated deterministic power flow solutions to render a probabilistic description of the output variables. The number of simulations needed increases with the degrees of freedom of the solution space, and therefore to obtain accurate results, thousands of simulations (or many more) are usually required. This makes MCS potentially unattractive for large systems The Monte Carlo method is utilized in a wide range of engineering areas for comparison and validation purposes. The advantage of MCS is that once it has converged, all the PDFs are simultaneously obtained. However, MCS also has some limitations in controlling the convergence of the process for large-scale systems, and this method may face some difficulties in dealing with rare events, e.g., to assess very small values of indices [18].

In various PPF studies, the conventional convolution technique is adopted to obtain the resulting PDFs [9]–[12]. The major problem associated with this approach is that the process requires a large amount of storage and computation time. In large systems, many convolution operations are involved, because integration of PDFs over the range of bus voltages and line flows need to be computed. Reference [7] mentions this problem and has applied the discrete Fourier transform (DFT) to reduce the computational burden.

However, the computation may still be extensive. Note that the DFT is most efficiently calculated using the fast Fourier transform (FFT), an exact representation of the DFT.

To avoid the convolution operation that appears in the calculation of the PDF of a linear combination of several random variables, the concept of cumulants is applied. By using cumulants, it is possible to obtain the resulting PDF of a linear combination of several random variables by a simple arithmetic process instead of convolution. If the cumulants of the distribution are known, there are several types of expansions like the Gram-Charlier expansion, Edgeworth expansion, and Cornish-Fisher expansion that can fit the profile of the PDF curve. Reference [19] proposes a new method for DC PPF in large power systems, and this innovative method combines the concept of cumulants and Gram-Charlier expansion theory to obtain PDFs of transmission line flows. Also, [20] uses the combined cumulants and the Cornish-Fisher expansion method compared with the Gram-Charlier expansion method to show better performance, but [21] analyzes the approximating problem of the Cornish-Fisher Expansion and Edgeworth expansion. The cumulant method has significantly reduced the computational time while maintaining accurate solutions. It enables the operators and planners to obtain the possible ranges of power flow and the probability of occurrence quickly since a simple arithmetic process is used instead of the complex convolution calculation. However, all the expansions applied in the cumulant method may have errors compared to the true PDF curve, since the applied terms of the expansions is finite. Therefore, the approximating accuracy with regard to the order of the expansions needs to be evaluated.

In most PPF studies, it is assumed that loads are independent variables, it is however, known that considerable correlation can exist between the various loads [25]. The

correlation among PV generations in the neighboring area also cannot be ignored. This is particularly the case when the time scale of interest is associated with system planning studies. Omission of this correlation can lead to misleading results and create a PDF of the output variables that is either narrower or broader than the true PDF. This problem can be overcome if a reasonable representation of the correlation is included in the probabilistic power flow analysis.

One method for treating the correlation among bus loads has been proposed in [16]. The model assumes Gaussian distribution of bus loads and a linearized economic dispatch model. Reference [5] proposed a method which models the correlation between loads using the total load constraint, and assumes that bus voltages and line flows are Gaussian distributed, thus only the variance must be computed. Monte Carlo simulations indicate that it is unrealistic to assume Gaussian distributions of bus voltages and line flows. For this reason, Sauer and Heydt [26] have proposed the use of higher moments for accurate representation of the probability distribution functions. Reference [27] proposed a linear dependence model of electric loads, which is suitable for the totally dependent cases. Using a linearized power flow model, a method is proposed which combines Monte Carlo simulation and convolutions. Reference [28] also treated the correlation between loads as approximately linear, and then eliminated the correlation variables. In [25] and [29], the load variable is divided into two random variables, a totally dependent variable and a totally independent variable, and this method is suitable for the case of small correlations. In [30], joint moments and joint cumulants are applied to solve the correlation problems.

In the PPF algorithm, the PDFs of the input variables need to be obtained. Therefore, the probabilistic models of input variables are required to be established, and then historical data is used to calculate the model parameters. The PPF algorithm is used in [31] to analyze the influences of wind power plants and solar energy plants on voltage quality in distribution systems, but the influence of line flow has not been taken into account. References [32], [33] describe a general probabilistic model of PV generation, which assumes that the PV generation active power has a linear relationship with solar irradiance. However, references [8], [34] proposed that the power output of a PV array is a function of the insolation, ambient temperature and prevailing wind speed. The consideration of these factors can make the PV generation model more reasonable and accurate.

For the probabilistic model of conventional generation uncertainty, some references [10] and [11] treat it as an independent variable with binomial distribution or discrete distribution to consider generation outage. However, under general conditions, the change of the conventional output depends on the change of other power injections. Small uncertainties can be compensated by the conventional generators with automatic generation control (AGC). The large power changes of power demands and PV generations need to be considered in the day-ahead generation scheduling settlement.

Significant research has been conducted dealing with different generation scheduling approaches in deterministic power flow studies. In [35], the sensitivities of reactive power generation with respect to demand were calculated to determine the generators to be rescheduled for the purpose of enhancing voltage stability. To minimize the control and operating cost, the corrective and preventive actions were determined in [36] using optimization. In [37], the direct equilibrium tracing approach was used to examine the

voltage stability of the system. Generation scheduling and load curtailment were used in [38] as preventive control against voltage instability. The amount of control actions was found by using optimization [39]–[42] with system operating constraints to determine the correct amount of rescheduling. For the system uncertainty analysis, a realistic generation dispatching law is really complex and nearly impossible to handle. In search for a workable compromise to represent utility's dispatching policy, [43] proposed simplified dispatching law function using a linear model.

1.3 Problem Statement

In this dissertation, the PPF algorithm is applied to analyze the uncertainty issues in the transmission system with PV generation installed. Since this dissertation focuses on large transmission systems, the cumulant method and the linearized model of the transmission system are applied to save storage and computational time. The proposed algorithm considers the uncertainties of both PV generation and loads. Three different types of expansions, Gram-Charlier expansion, Edgeworth expansion, and Cornish-Fisher expansion are considered in the proposed algorithm, and their properties, advantages and deficiencies are compared. The correlation of PV generations is analyzed in this algorithm, and an approach utilizing the joint moments and joint cumulants is applied.

A novel probabilistic model of PV generation is also established in this dissertation. The probabilistic model takes into account the performance of PV generation and the factors which influence PV generation production. The actual data of PV generation obtained from a location in Arizona is analyzed to evaluate the accuracy of the proposed probabilistic model. The probabilistic model of load is also established based on actual historical load data.

This dissertation also proposes a novel PPF algorithm considering the conventional generation dispatch which can balance the variations of PV generation resources. A linear model of the change of conventional generation is developed based on the dispatching law. To take into account the characteristics of generation dispatching utilizing conventional economic dispatch, the linear model utilized in the PPF algorithm is revised. Based on the proposed PPF algorithm, a novel dispatching strategy called probabilistic optimal power dispatching strategy is proposed to consider the overload probability constraint in the economic dispatching strategy.

The accuracy of the results obtained is compared with the MCS method. The convergence and accuracy of MCS is quantified. It is commonly accepted that the PPF algorithm can be used to assess the probability of a line flow being greater than its thermal rating and the probability of a bus voltage magnitude being outside its operational constraints. These are extremely useful parameters in conducting operating studies of transmission systems.

1.4 Dissertation Organization

The dissertation is organized as follows: in Chapter 2, some basic concepts in statistics are defined. The convolution method and cumulant method are presented. Three approximation expansions are introduced, and their properties are discussed.

Chapter 3 gives the outline of the PPF algorithm, which applies the cumulant method to obtain the CDFs and PDFs of the determined variables. The probabilistic formulation for power systems is presented, and the probabilistic formulation for the slack bus is also established. The method to deal with the correlation between each variable is

analyzed. A novel probabilistic model of PV generation is proposed, and the load probabilistic characteristic is also discussed.

In Chapter 4, the probabilistic models of PV generation and load are obtained by using actual history data. A novel probabilistic model of PV generation is proposed, and the load probabilistic characteristic is also discussed.

Chapter 5 develops the PPF algorithm by considering the generation dispatching operation. The probabilistic model of power systems is revised. A novel generation dispatching strategy is proposed to consider both the economic cost and the uncertainty influence.

In Chapter 6, the proposed approach is illustrated on the Arizona area of the WECC system, and the simulation results of the PPF algorithm are discussed.

Conclusions are included in Chapter 7, and future work is also described.

Chapter 2

STATISTICAL BACKGROUND

Before the introduction of the PPF algorithm, some basic concepts of statistics are introduced. The core mathematical techniques of cumulant method are presented in this section. The characteristics of cumulants are the basis of the proposed PPF algorithm. To deduce the cumulant method, convolution techniques are necessary to be introduced.

2.1 Convolution Techniques

Let *x* and *y* be independent random variables with known PDFs $f_x(x)$ and $f_y(y)$. Let $z = x + y$, the PDF $f_z(z)$ is obtained by convolution as follows,

$$f_z(z) = \int_{-\infty}^{+\infty} f_x(x) f_y(z-x) dx \tag{2-1}$$

$$f_z(z) = f_x(x) * f_y(y) \tag{2-2}$$

where * denotes the convolution operation.

Let $X_1, X_2, \cdots, X_n$ be n independent variables with known PDFs $f_{X_1}(X_1), f_{X_2}(X_2), \cdots, f_{X_n}(X_n)$. $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, $a_1, a_2, \cdots, a_n$ are sensitivity coefficients for Z. Let $Y_1 = a_1X_1$, $Y_2 = a_2X_2$, $\cdots, Y_n = a_nX_n$, so that the PDFs of Y_1, Y_2, $\cdots, Y_n$ are $\frac{1}{|a_1|} f_{X_1}(\frac{Y_1}{a_1}), \frac{1}{|a_2|} f_{X_2}(\frac{Y_2}{a_2}), \cdots, \frac{1}{|a_n|} f_{X_n}(\frac{Y_n}{a_n})$. Then, the PDF of Z is given by,

$$f_Z(z) = \frac{1}{|a_1|} f_{X_1}(\frac{Y_1}{a_1}) * \frac{1}{|a_2|} f_{X_2}(\frac{Y_2}{a_2}) * \cdots * \frac{1}{|a_n|} f_{X_n}(\frac{Y_n}{a_n}) \tag{2-3}$$

It is often of interest to characterize variables for the probability parameters as the expected value μ and the standard deviation σ. The expected values of $X_1, X_2, \cdots, X_n$

are $\mu_{X_1}, \mu_{X_2}, \cdots, \mu_{X_n}$, and the standard deviations are $\sigma_{X_1}, \sigma_{X_2}, \cdots, \sigma_{X_n}$. The expected value and standard deviation of Z are given by,

$$\begin{aligned} \mu_Z &= a_1\mu_{X_1} + a_2\mu_{X_2} + \cdots + a_n\mu_{X_n} \\ \sigma_Z^2 &= a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \cdots + a_n^2\sigma_{X_n}^2 \end{aligned} \tag{2-4}$$

2.2 Cumulant Method

The main objective of using the cumulant method is to avoid convolution calculations between the PDFs of input variables and to replace them with a simple arithmetic process [19]. The main idea in this approach is to transform the convolution equations to linear equations between each variable. In order to motivate the proposed approach, some statistical definitions are introduced as follows.

2.2.1 Characteristic Function (Chapter 15 of [21])

The main use of the characteristic function is to transform the convolution relationship between the PDFs of input variables to a multiplicative relationship. The definition of the characteristic function is as follows.

If the random variable x has a cumulative distribution function (CDF) $F(x)$, the expected value of the particular function e^{itx} can be written as,

$$\psi(t) = E(e^{itx}) = \int_{-\infty}^{+\infty} e^{itx} dF(x) \tag{2-5}$$

This function of the real variable t, and the imaginary unit i is called the characteristic function of the variable x. The second characteristic function is defined as $ln\psi(t)$.

Let x and y be independent random variables with the CDFs $F_x(x)$ and $F_y(y)$. Let $z = x + y$, so that the CDF $F_z(z)$ of z is obtained by convolution,

$$F_z(z) = \int_{-\infty}^{+\infty} F_x(z-y)dF_y(y) = \int_{-\infty}^{+\infty} F_y(z-x)dF_x(x)$$
$$F_z(z) = F_x(x) * F_y(y) \tag{2-6}$$

Let $\psi_x(t)$, $\psi_y(t)$ and $\psi_z(t)$ denote the characteristic functions of *x*, *y* and *z*, respectively.

$$\psi_Z(t) = E[e^{it(x+y)}] = E(e^{itx})E(e^{ity}) = \psi_x(t)\psi_y(t) \tag{2-7}$$

For $X_1, X_2, \cdots, X_n$ of *n* independent variables with the CDFs $F_{X_1}(X_1), F_{X_2}(X_2), \cdots, F_{X_n}(X_n)$, the CDF $F_X(X)$ of the sum $X = X_1 + X_2 + \cdots + X_n$ is given by,

$$F_X(X) = F_{X_1}(X_1) * F_{X_2}(X_2) * \cdots * F_{X_n}(X_n) \tag{2-8}$$

Then, for $X_1, X_2, \cdots, X_n$ of *n* independent variables with the characteristic functions $\psi_{X_1}(t), \psi_{X_2}(t), \cdots, \psi_{X_n}(t)$, the characteristic function $\psi_X(t)$ of the sum $X = X_1 + X_2 + \cdots + X_n$ is given by,

$$\psi_X(t) = \psi_{X_1}(t)\psi_{X_2}(t)\cdots\psi_{X_n}(t) \tag{2-9}$$

The multiplication theorem for characteristic functions is given by,

$$ln\psi_X(t) = ln\psi_{X_1}(t) + ln\psi_{X_2}(t) + \cdots + ln\psi_{X_n}(t) \tag{2-10}$$

2.2.2 Moments and Central Moments

For a variable *x* and a positive integer *v*, the function x^v is integrable with respect to the CDF *F(x)* over $(-\infty, +\infty)$, the integral

$$\alpha_v = E(x^v) = \int_{-\infty}^{+\infty} x^v dF(x) \tag{2-11}$$

is called the moment of order *v* or the v^{th} moment of the distribution *x*.

The moments about the mean value μ of *x*, are often called the central moments,

$$\beta_v = E[(x-\mu)^v] = \int_{-\infty}^{+\infty} (x-\mu)^v \, dF(x) \tag{2-12}$$

If *x* is a sample discrete variable written as $x_1, x_2, \cdots, x_n$, the v^{th} moment and central moment of *x* are,

$$\alpha_v = E(x^v) = \frac{1}{n}\sum_{i=1}^{n} x_i^v \tag{2-13}$$

$$\beta_v = E[(x-\mu)^v] = \frac{1}{n}\sum_{i=1}^{n} (x_i-\mu)^v \tag{2-14}$$

It can be demonstrated that

$$\beta_v = \sum_{k=0}^{v} \binom{v}{k} (-1)^{v-k} \alpha_k {\alpha_1}^{v-k} \tag{2-15}$$

where $\binom{v}{k} = \frac{v!}{k!(v-k)!}$. For a linear function $Y = aX + b$, the v^{th} moment of the variable Y is given by,

$$\alpha'_v = E[(aX+b)^v] = a^v \alpha_v + \binom{v}{1} a^{v-1} b \alpha_{v-1} + \cdots + b^v \tag{2-16}$$

2.2.3 Cumulants

If the k^{th} moment of the distribution exists, the characteristic function in terms of the moments α_v, can be developed in terms of a Taylor series centered at zero for small values of *t*,

$$\psi(t) = 1 + \sum_{1}^{k} \frac{\alpha_v}{v!} (it)^v + o(t^k) \tag{2-17}$$

$$ln\psi(t) = \sum_{1}^{k} \frac{\kappa_v}{v!} (it)^v + o(t^k) \tag{2-18}$$

The coefficients κ_v are called the cumulants or semi-invariants of the distribution. The cumulants κ'_v of a linear function $Y = aX + b$ are obtained as follows,

$$ln[e^{bit}\psi(at)] = \sum_{1}^{k} \frac{\kappa'_v}{v!}(it)^v + o(t^k) \tag{2-19}$$

where $\kappa'_1 = a\kappa_1 + b$, and $\kappa'_v = a^v\kappa_v$ for $v > 1$.

According to the properties of characteristic functions, if $X = X_1 + X_2 + \cdots + X_n$ and X_1, X_2, ⋯, X_n are independent, the characteristic function $\psi_X(t)$ is given by,

$$\psi_X(t) = \psi_{X_1}(t)\psi_{X_2}(t)\cdots\psi_{X_n}(t) \tag{2-20}$$

$$ln\psi_X(t) = ln\psi_{X_1}(t) + ln\psi_{X_2}(t) + \cdots + ln\psi_{X_n}(t) \tag{2-21}$$

Therefore,

$$\kappa_{X,v} = \kappa_{X_1,v} + \kappa_{X_2,v} + \cdots + \kappa_{X_n,v} \tag{2-22}$$

Furthermore, according to (2-22), the following equations can be obtained,

$$\mu_X = \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n} \tag{2-23}$$

$$\sigma_X^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \cdots + \sigma_{X_n}^2 \tag{2-24}$$

The discussion can be extended to any number of independent variables. If

$$Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n \tag{2-25}$$

Then, the v^{th} cumulant Z is given as follows,

$$\kappa_{Z,v} = a_1^v\kappa_{X_1,v} + a_2^v\kappa_{X_2,v} + \cdots + a_n^v\kappa_{X_n,v} \tag{2-26}$$

2.2.4 Relationship between Cumulants and Moments

The relationship between the moments and cumulants can be deduced by,

$$ln[1+\sum_{v=1}^{k}\frac{\alpha_v}{v!}(it)^v]=\sum_{v=1}^{k}\frac{\kappa_v}{v!}(it)^v+o(t^k) \tag{2-27}$$

It is seen that κ_n is a polynomial in $\alpha_1,\cdots,\alpha_n$ and conversely α_n is a polynomial in $\kappa_1,\cdots,\kappa_n$. The cumulants are related to the moments by the following recursion formula,

$$\kappa_n=\alpha_n-\sum_{m=1}^{n-1}\binom{n-1}{m-1}\kappa_m\alpha_{n-m} \tag{2-28}$$

In particular,

$$\begin{aligned}
&\kappa_1=\alpha_1=\mu\\
&\kappa_2=\alpha_2-\alpha_1^2=\sigma^2\\
&\kappa_3=\alpha_3-3\alpha_1\alpha_2+2\alpha_1^3\\
&\kappa_4=\alpha_4-3\alpha_2^2-4\alpha_1\alpha_3+12\alpha_1^2\alpha_2-6\alpha_1^4\\
&\kappa_5=\alpha_5-5\alpha_4\alpha_1-10\alpha_3\alpha_2+20\alpha_3\alpha_1^2+30\alpha_2^2\alpha_1-60\alpha_2\alpha_1^3+24\alpha_1^5\\
&\kappa_6=\alpha_6-6\alpha_5\alpha_1-15\alpha_4\alpha_2+30\alpha_4\alpha_1^2-10\alpha_3^2+120\alpha_3\alpha_2\alpha_1-120\alpha_3\alpha_1^3\\
&\qquad+30\alpha_2^3-270\alpha_2^2\alpha_1^2+360\alpha_2\alpha_1^4-120\alpha_1^6\\
&\cdots\cdots
\end{aligned} \tag{2-29}$$

The expressions of the central moment β_v related to the cumulants become,

$$\begin{aligned}
&\beta_1=0\\
&\beta_2=\kappa_2=\sigma^2\\
&\beta_3=\kappa_3\\
&\beta_4=\kappa_4+3\kappa_2^2\\
&\beta_5=\kappa_5+10\kappa_2\kappa_3\\
&\beta_6=\kappa_6+15\kappa_2\kappa_4+10\kappa_3^2+15\kappa_2^3\\
&\cdots\cdots
\end{aligned} \tag{2-30}$$

2.2.5 Joint Moments and Joint Cumulants

In the above analysis, all the input variables are assumed to be independent (un-correlated). However, in an actual system, the power injections from nearby PV resources

are likely to be highly correlated owing to common effects such as weather conditions and human-behavior patterns. Thus, the independence assumption of input variables can become less valid, and the power may be correlated in some cases. Therefore, in the PPF algorithm it is more realistic to include correlation to statistically characterize the dependence.

For two random variables *x* and *y*, the covariance between *x* and *y* is given by,

$$\mathrm{cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] \tag{2-31}$$

By using the linearity property of expectations, (2-31) can be simplified to,

$$\mathrm{cov}(x, y) = E(xy) - \mu_x \mu_y \tag{2-32}$$

If *x* and *y* are two scalar random variables with a joint probability density function $f(x, y)$, the covariance between *x* and *y* is,

$$\mathrm{cov}(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dxdy - \mu_x \mu_y \tag{2-33}$$

If *x* and *y* are two sample discrete variables with two series of *n* measurements written as $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$, the mean value of *x* and *y* is $\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i$, and the covariance between *x* and *y* is given by,

$$\mathrm{cov}(x, y) = E[(x - \mu_x)^T (y - \mu_y)] = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y) \tag{2-34}$$

The correlation coefficient between *x* and *y* is given by,

$$\rho = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y} \tag{2-35}$$

where $|\rho| \leq 1$. If $\rho = 0$, *x* and *y* are independent. If $\rho > 0$, *x* and *y* are positively dependent. If $\rho < 0$, *x* and *y* are negatively dependent. The concept of correlation is readily extended

to the vector case. Let X_1, X_2, …, X_n be n elements of the vector X. The joint probability density function is $f(X_1, X_2, \ldots, X_n)$, and the joint characteristic function $\psi(t_1, t_2, \ldots, t_n)$ is given by,

$$\begin{aligned}\psi(t_1,t_2,\cdots,t_n) &= E(e^{it^T X}) \\ &= \underbrace{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}}_{n} e^{it^T X} f(X_1, X_2,\ldots, X_n)dX_1 dX_2 \ldots dX_n\end{aligned} \tag{2-36}$$

where $t = (t_1, t_2, \cdots, t_n)^T$, $X = (X_1, X_2, \cdots, X_n)^T$.

Similar to the calculation of moments in (2-11), one of the v^{th} order joint moments of the n variables is,

$$\begin{aligned}\alpha_{v1,v2,\cdots,vn} &= E(X_1^{v1} X_2^{v2} \cdots X_n^{vn}) \\ &= \underbrace{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}}_{n} X_1^{v1} X_2^{v2} \cdots X_n^{vn} f(X_1, X_2,\ldots, X_n)dX_1 dX_2 \ldots dX_n\end{aligned} \tag{2-37}$$

where $v1 + v2 + \cdots + vn = v$.

Similar to (2-17) and (2-18), the Taylor series expansion of the joint characteristic function is,

$$\psi(t_1,t_2,\cdots,t_n) = \sum_{v1,v2,\cdots,vn=0}^{\infty} \alpha_{v1,v2,\cdots,vn} \frac{(it_1)^{v1}}{v1!}\frac{(it_2)^{v2}}{v2!}\cdots\frac{(it_n)^{vn}}{vn!} \tag{2-38}$$

$$ln\psi(t_1,t_2,\cdots,t_n) = \sum_{v1,v2,\cdots,vn=0}^{\infty} \kappa_{v1,v2,\cdots,vn} \frac{(it_1)^{v1}}{v1!}\frac{(it_2)^{v2}}{v2!}\cdots\frac{(it_n)^{vn}}{vn!} \tag{2-39}$$

where $\kappa_{v1,v2,\cdots,vn}$ is the v^{th} order joint cumulant of the n variables. The relationship between joint moments and joint cumulants is rather complex, and the theoretical derivation of this relationship is shown in [30]. As an example, the joint cumulants among four random variables is given by,

$$
\begin{aligned}
\kappa_{v1,v2} &= \alpha_{v1,v2} - \alpha_{v1}\alpha_{v2} \\
\kappa_{v1,v2,v3} &= \alpha_{v1,v2,v3} - \alpha_{v1,v2}\alpha_{v3} - \alpha_{v1,v3}\alpha_{v2} - \alpha_{v2,v3}\alpha_{v1} + 2\alpha_{v1}\alpha_{v2}\alpha_{v3} \\
\kappa_{v1,v2,v3,v4} &= \alpha_{v1,v2,v3,v4} - \alpha_{v1,v2,v3}\alpha_{v4} - \alpha_{v1,v2,v4}\alpha_{v3} - \alpha_{v1,v3,v4}\alpha_{v2} \\
&\quad - \alpha_{v2,v3,v4}\alpha_{v1} - \alpha_{v1,v2}\alpha_{v3,v4} - \alpha_{v1,v3}\alpha_{v2,v4} - \alpha_{v1,v4}\alpha_{v2,v3} \\
&\quad + 2(\alpha_{v1,v2}\alpha_{v3}\alpha_{v4} + \alpha_{v1,v3}\alpha_{v2}\alpha_{v4} + \alpha_{v1,v4}\alpha_{v2}\alpha_{v3} + \alpha_{v2,v3}\alpha_{v1}\alpha_{v4} \\
&\quad + \alpha_{v2,v4}\alpha_{v1}\alpha_{v3} + \alpha_{v3,v4}\alpha_{v1}\alpha_{v2}) - 6\alpha_{v1}\alpha_{v2}\alpha_{v3}\alpha_{v4}
\end{aligned}
\tag{2-40}
$$

In particular, for two random variables *x* and *y*, the second order joint cumulants are,

$$
\begin{aligned}
\kappa_{2,0} &= \sigma_x^2 \\
\kappa_{0,2} &= \sigma_y^2 \\
\kappa_{1,1} &= \mathrm{cov}(x, y) = \rho\sigma_x\sigma_y
\end{aligned}
\tag{2-41}
$$

For *n* input variables, the number of the v^{th} order of self and joint cumulants among the input variables can be determined as [15],

$$
N_{n,v} = \binom{n+v-1}{v} = \frac{(n+v-1)!}{v!(n-1)!} \tag{2-42}
$$

Thus, the number of self and joint cumulants for different numbers of input variables is shown in Table 2.1. It is observed that the number of self and joint cumulants increases dramatically as the degrees of freedom increases (i.e., when more input variables and more orders of joint cumulants are considered).

Table 2.1 Number of self and joint cumulants

No. of input variables	No. of v^{th} order of self and joint cumulants				
	1st	2nd	3rd	4th	5th
1	1	1	1	1	1
10	10	55	220	715	2002
100	100	5050	171700	4.42E06	9.20E07
1000	1000	500500	1.67E08	4.19E10	8.42E12

2.2.6 Cumulants of a Linear Combination of Variables

Section 2.2.3 discusses the cumulants of a linear combination of independent variables (see (2-26)), and the correlated case is discussed in this section.

For $Z = a_1X_1 + a_2X_2$, it could be demonstrated [20] that the v^{th} order cumulants of Z are given by,

$$\kappa_{Z,v} = \sum_{v1=0}^{v} \frac{v!}{v1!(v-v1)!} a_1^{v1} a_2^{v-v1} \kappa_{v1,(v-v1)} \tag{2-43}$$

Without loss of generality for $Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, the equations (2-29), (2-37) and (2-40) can be used to evaluate the cumulants of Z as,

$$\begin{aligned}
\kappa_{Z,1} &= \mu_Z = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \kappa_{X_i,1} = \sum_{i=1}^{n} a_i \mu_{X_i} \\
\kappa_{Z,2} &= \sigma_Z^2 = E\left[\left(\sum_{i=1}^{n} a_i X_i\right)^2\right] - \left[E\left(\sum_{i=1}^{n} a_i X_i\right)\right]^2 \\
&= \sum_{i=1}^{n} a_i^2 \kappa_{X_i,2} + 2\sum_{i=1,i<j}^{n} a_i a_j \kappa_{X_i,X_j} \\
&= \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + 2\sum_{i=1,i<j}^{n} a_i a_j \rho \sigma_{X_i} \sigma_{X_j} \\
\kappa_{Z,3} &= E\left[\left(\sum_{i=1}^{n} a_i X_i\right)^3\right] - 3E\left(\sum_{i=1}^{n} a_i X_i\right) E\left[\left(\sum_{i=1}^{n} a_i X_i\right)^2\right] + 2\left[E\left(\sum_{i=1}^{n} a_i X_i\right)\right]^3 \\
&= \sum_{i=1}^{n} a_i^3 \kappa_{X_i,3} + 3\sum_{i=1,i\neq j}^{n} a_i^2 a_j \kappa_{X_i^2,X_j} + 6\sum_{i=1,i<j<k}^{n} a_i a_j a_k \kappa_{X_i,X_j,X_k} \\
&\cdots
\end{aligned} \tag{2-44}$$

In particular, if $X_1, X_2, \cdots, X_n$ are independent, the v^{th} order cumulant of Z is given by,

$$\kappa_{Z,v} = a_1^v \kappa_{X_1,v} + a_2^v \kappa_{X_2,v} + \cdots + a_n^v \kappa_{X_n,v} \tag{2-45}$$

For the random variable $z = x + y$, according to (2-44), the mean value of z is $\mu_z = \mu_x + \mu_y$, and the standard deviation of z is $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y}$. Three cases of different correlations between x and y are discussed as follows.

A. Totally independent case

If the two variables x and y are totally independent, the correlation coefficient between x and y is $\rho = 0$, and the standard deviation of z is $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$.

B. Totally dependent case

For the positively correlated condition between x and y, the correlation coefficient between x and y is $\rho = 1$, so the standard deviation of z is $\sigma_z = \sigma_x + \sigma_y$. Since $\sigma_x + \sigma_y \geq \sqrt{\sigma_x^2 + \sigma_y^2}$, the standard deviation of the sum of the totally positive correlated variables is no less than that of the totally independent variables. It means that the resulting PDF of z is flatter and broader.

For the negatively correlated condition between x and y, the correlation coefficient between x and y is $\rho = -1$, so the standard deviation of z is $\sigma_z = |\sigma_x - \sigma_y|$. Since $|\sigma_x - \sigma_y| \leq \sqrt{\sigma_x^2 + \sigma_y^2}$, the standard deviation of the sum of two totally negative correlated variables is no more than that of two totally independent variables.

C. Partially dependent case

If $0 < |\rho| < 1$, the variables x and y are partially correlated, and the standard deviation of z is $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y}$.

If $0 < \rho < 1$, *x* and *y* are partially positive correlated.

$$\sqrt{\sigma_x^2 + \sigma_y^2} \le \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y} \le \sigma_x + \sigma_y \qquad (2\text{-}46)$$

The above equation indicates that the standard deviation of the sum of the partially positive correlated variables is no less than that of the totally independent variables but no greater than that of the totally positively dependent variables.

If $-1 < \rho < 0$, *x* and *y* are partially negative correlated.

$$\left|\sigma_x - \sigma_y\right| \le \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y} \le \sqrt{\sigma_x^2 + \sigma_y^2} \qquad (2\text{-}47)$$

Thus, the standard deviation of the sum of the partially negative correlated variables is no greater than that of the totally independent variables but no less than that of the totally negatively dependent variables.

According to the above analysis, the correlation among the input variables influences the uncertainty results significantly and cannot be ignored.

2.3 Approximation Expansions of CDF and PDF

When the moments and cumulants of the variables are known, the next step is to obtain the CDF and PDF of the variables. There are different approaches by using various types of series expansions to approximate the true function. Most of these series expansions are based on orthogonal functions and their properties. Thus the effectiveness of the expansions and their truncated forms depends on how similar the actual variables behave as compared to the orthogonal functions used. The coefficients in the expansions can be computed from the moments or cumulants of the distribution through the utilization of basic properties of orthogonal functions.

Three different approximation expansions are introduced as follows. The variable *x* in the three expansions should be rendered to *standard measure* (mean value of zero and standard deviation of unity). If the variable *x* has mean value μ and standard deviation σ, the variable should be normalized as $x^* = \frac{x-\mu}{\sigma}$, and the cumulant should be normalized as $\kappa_v^* = \frac{\kappa_v}{\sigma^v}$.

2.3.1 Gram-Charlier Type A Series Expansion

If a random variable *x* is normalized, according to Gram-Charlier Type A series expansion theory [22] [45], the CDF and the PDF of *x* can be written as,

$$\begin{aligned} F(x) &= \sum_{i=0}^{n} \frac{c_i}{i!} \phi^{(i)}(x) \\ f(x) &= \sum_{i=0}^{n} \frac{c_i}{i!} \varphi^{(i)}(x) \end{aligned} \tag{2-48}$$

where *n* is the total order number of the Gram-Charlier expansion; $\phi(x)$ and $\varphi(x)$ represent the CDF and PDF of the standard normal distribution with $\mu = 0$ and $\sigma = 1$;

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{2-49}$$

i is index corresponding to the order of Gram-Charlier expansion; c_i is the constant coefficient of the Gram-Charlier expansion,

$$c_i = (-1)^i \int_{-\infty}^{\infty} f(x) H_i(x) dx \tag{2-50}$$

where $H_i(x)$ is the Hermite polynomial.

In this case the orthogonal series expansion is in terms of the classical orthogonal functions, the Hermite polynomials, defined as,

$$H_i(x) = (-1)^i e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \tag{2-51}$$

Thus, the Hermite polynomials are calculable as,

$$H_i(x) = (-1)^i \frac{d^n \varphi(x)}{dx^n} \frac{1}{\varphi(x)} \tag{2-52}$$

$$H_n(x) = n! \sum_{k=0}^{[n/2]} \frac{(-1)^k x^{n-2k}}{k!(n-2k)!2^k} \tag{2-53}$$

The first ten probability Hermite polynomials are given by,

$$\begin{aligned}
H_0(x) &= 1 \\
H_1(x) &= x \\
H_2(x) &= x^2 - 1 \\
H_3(x) &= x^3 - 3x \\
H_4(x) &= x^4 - 6x^2 + 3 \\
H_5(x) &= x^5 - 10x^3 + 15x \\
H_6(x) &= x^6 - 15x^4 + 45x^2 - 15 \\
H_7(x) &= x^7 - 21x^5 + 105x^3 - 105x \\
H_8(x) &= x^8 - 28x^6 + 210x^4 - 420x^2 + 105 \\
H_9(x) &= x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x \\
H_{10}(x) &= x^{10} - 45x^8 + 630x^6 - 3150x^4 + 4725x^2 - 945
\end{aligned} \tag{2-54}$$

The expressions for the c_v constant coefficients are given by,

$$
\begin{aligned}
c_0 &= 1 \\
c_1 &= c_2 = 0 \\
c_3 &= -\frac{\beta_3}{\sigma^3} \\
c_4 &= \frac{\beta_4}{\sigma^4} - 3 \\
c_5 &= -\frac{\beta_5}{\sigma^5} + 10\frac{\beta_3}{\sigma^3} \\
c_6 &= \frac{\beta_6}{\sigma^6} - 15\frac{\beta_4}{\sigma^4} + 30 \\
c_7 &= -\frac{\beta_7}{\sigma^7} + 21\frac{\beta_5}{\sigma^5} - 105\frac{\beta_3}{\sigma^3} \\
c_8 &= \frac{\beta_8}{\sigma^8} - 28\frac{\beta_6}{\sigma^6} + 210\frac{\beta_4}{\sigma^4} - 315 \\
c_9 &= -\frac{\beta_9}{\sigma^9} + 36\frac{\beta_7}{\sigma^7} - 378\frac{\beta_5}{\sigma^5} + 1260\frac{\beta_3}{\sigma^3} \\
&\cdots\cdots
\end{aligned}
\tag{2-55}
$$

The convergence properties of the Gram-Charlier Type A expansion depends on the form of the variables and their similarity to the Hermite polynomial functions. Some research efforts assess the convergence in certain given applications as poor (Chapter 17 of [22]). The concept of convergence is that the approximation expansion in (2-48) converges to $F(x)$ if the approximation expansion tends to $F(x)$ for every x as $n \rightarrow \infty$. That is, the truncation of the series expansion is at the n^{th} term. It is proved that the infinite series in (2-48) converges for every x if the integral

$$
\int_{-\infty}^{\infty} \exp(x^2 / 4) dF(x) \tag{2-56}
$$

is convergent, and if $F(x)$ is of bounded variation in $(-\infty, \infty)$. The inference is that the density function $f(x)$ must fall to zero faster than exp $(x^2/4)$ for the series to converge which is often too restrictive for practical applications.

2.3.2 Edgeworth Expansion

If a random variable x is normalized, the Edgeworth series expansion may be of value, and this expansion for order n can be written as,

$$f(x)=\varphi(x)\left\{1+\sum_{n=1}^{\infty}\sum_{\{k_m\}}H_{n+2r}(x)\prod_{m=1}^{n}\frac{1}{k_m!}\left(\frac{\kappa_{m+2}}{(m+2)!}\right)^{k_m}\right\} \tag{2-57}$$

where $\{k_m\}$ in the inner summation consists of all non-negative integer solutions of the equation $k_1 + 2k_2 + \cdots + nk_n = n$, and $r = k_1 + k_2 + \cdots + k_n$. In order to obtain the corresponding expansion for $F(x)$, $\varphi(x)$ is replaced by $\phi(x)$. The detailed theoretical derivation can be found in [21] and [22]. The Edgeworth expansion in powers of $n^{-3/2}$ is given as follows. The terms of order $n^{-v/2}$ contain the cumulants $\kappa_3,\cdots, \kappa_{v+2}$.

$$\begin{aligned} f(x)=\varphi(x)&-\frac{\kappa_3}{3!}\varphi^{(3)}(x)+\frac{\kappa_4}{4!}\varphi^{(4)}(x)+\frac{10}{6!}\kappa_3^2\varphi^{(6)}(x)\\ &-\frac{\kappa_5}{5!}\varphi^{(5)}(x)-\frac{35\kappa_3\kappa_4}{7!}\varphi^{(7)}(x)+\frac{280}{9!}\kappa_3^3\varphi^{(9)}(x)\cdots \end{aligned} \tag{2-58}$$

Reference [22] shows that under fairly general conditions the Edgeworth expansion has asymptotic properties, which means that the difference between the true function and the partial series of the first N terms is of a lower order than the N^{th} term in the sum [21].

2.3.3 Cornish-Fisher Expansion

The Cornish-Fisher expansion is used to approximate the variable's quantile, which is the inverse function of the CDF. The quantile x of the probability q is the root of the distribution function $F(x) = q$. The method is based on the cumulants of the variable and the quantiles of the standard normal probability distribution. The theoretical derivation is

given in [23]. If the variable is normalized, considering the first five orders of cumulants, the expansion is given by,

$$\begin{aligned} x(q) = \xi(q) &+ \frac{\xi^2(q)-1}{6}\kappa_3 + \frac{\xi^3(q)-3\xi(q)}{24}\kappa_4 \\ &- \frac{2\xi^3(q)-5\xi(\alpha)}{36}\kappa_3^2 + \frac{\xi^4(q)-6\xi^2(q)+3}{120}\kappa_5 \\ &- \frac{\xi^4(q)-5\xi^2(q)+2}{24}\kappa_3\kappa_4 + \frac{12\xi^4(q)-53\xi^2(q)+17}{324}\kappa_3^3 \end{aligned} \tag{2-59}$$

where q is the probability, $x(q)$ is the quantile of the variable, $\xi(q)$ is the quantile of the standard normal distribution, $x(q) = F^{-1}(q)$, $\xi(q) = \phi^{-1}(q)$.

When the quantiles of each probability are calculated, the CDF curve of the variable x can be drawn by using $x(q)$—q curve.

2.3.4 Comparison among the Three Expansions

Since the three expansions are all related to the normal distribution, the approximation is expected to be more accurate to fit a nearly normal distribution. The deficiencies of the three expansions are summarized as follows [21]-[24].

- The convergence and accuracy of the three approximations do not necessarily improve with increasing orders of truncation of the series.
- The three expansions do not guarantee monotonicity and convergence.
- The CDF of the Gram-Charlier and Edgeworth expansions do not necessarily exist in the range 0~1.
- Edgeworth and Cornish-Fisher expansion becomes less and less reliable when the probability is near 0 or 1.

Although both Edgeworth and Cornish-Fisher expansions also have convergence problems, the convergence properties of the Gram-Charlier expansion depend on the application. As a result, the Gram-Charlier series has limited applicability except for nearly normal distribution of the variables. In practical applications, the primary concern in not whether the infinite series is convergent, but whether a finite number of terms suffice to give a satisfactory approximation of the PDF [22]. It is possible that the first few terms of Gram-Charlier expansion give a good fit even though the infinite series diverges. Reference [21] gives an example of the χ_v^2 distribution, in which the Gram-Charlier expansion is divergent while Edgeworth expansion gives a better result. For $x > -\sqrt{v/2}$, the PDF of χ_v^2 distribution is shown as follows.

$$f(x) = \sqrt{2v}\,\frac{\left(\sqrt{2v}x + v\right)^{v/2-1} \exp\left[-\left(\sqrt{2v}x + v\right)/2\right]}{2^{v/2}\Gamma\left(v/2\right)} \tag{2-60}$$

where $\Gamma(x)$ is gamma function given by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \tag{2-61}$$

Let $v = 5$. Considering the first six orders of cumulants, the PDF and CDF curves of χ_v^2 distribution approximated by different types of expansions are shown Fig. 2.1 and Fig. 2.2. It shows that the Gram-Charlier expansion is divergent, since the distribution does not meet the convergence requirements that $\int_{-\infty}^{\infty} \exp(x^2/4)\, dF(x)$ must be convergent, and the density function $f(x)$ falls to zero slower than exp ($x^2/4$).

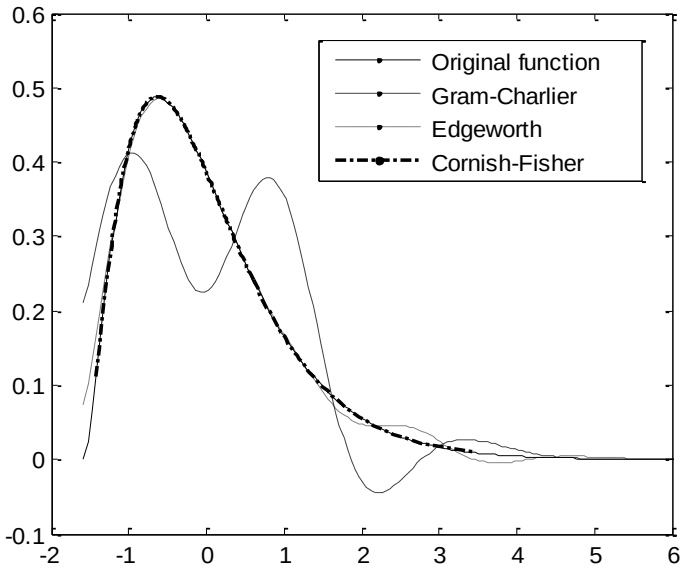

Fig. 2.1 PDF curves of χ_v^2 distribution approximated by different types of expansions

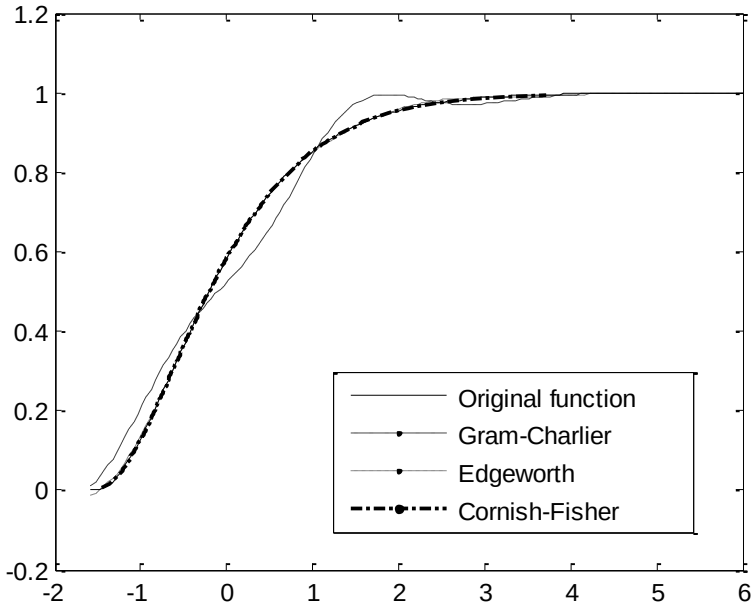

Fig. 2.2 CDF curves of χ_v^2 distribution approximated by different types of expansions

In [20] and [46], some examples show that the Cornish-Fisher expansion gives a better performance for non-normal distribution. However, the Cornish-Fisher and Edgeworth expansions may have an error in the tail regions, e.g. the two ends of the CDF curve.

For example, the Beta distribution PDF is as follows,

$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1},\qquad 0\le x\le 1 \tag{2-62}$$

Let $a = 2$, $b = 0.5$. Considering the first six orders of cumulants, the CDF curves of the three approximations are shown in Fig. 2.3, which indicates the Cornish-Fisher and Edgeworth expansions become unreliable as the probability is close to 1. Reference [24] discussed the tail behavior problem of Cornish-Fisher and Edgeworth expansions.

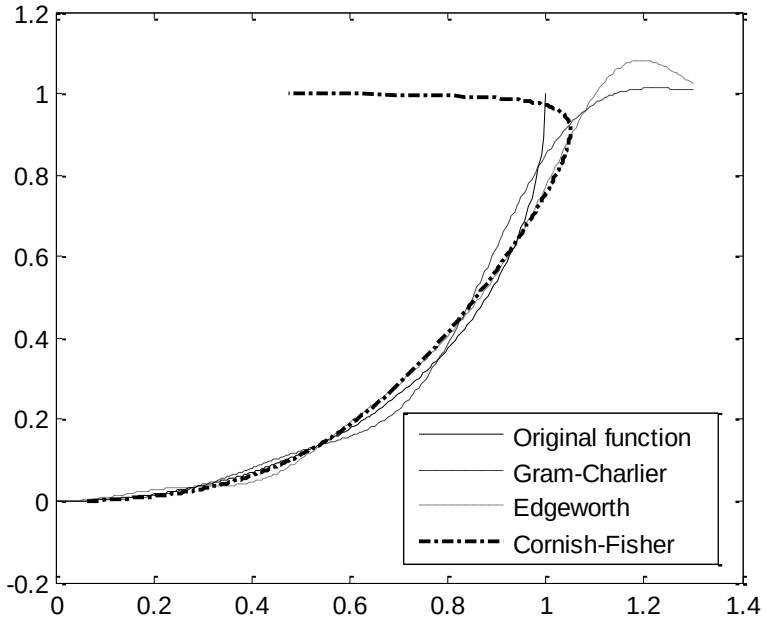

Fig. 2.3 CDF curves of Beta distribution approximated by different types of expansions

In many engineering applications, performance in the tail regions is very important and often the main reason for the study at hand. In short, the application of a particular expansion is dependent on the PDF of the variables.

Chapter 3

PROBABILISTIC POWER FLOW ALGORITHM

In operating and planning studies of power systems, the security of the system should be assessed and monitored. The PPF algorithm can efficiently characterize the impact of variable uncertainties on power system performance. This method permits the power injection variables to vary probabilistically and provides results in terms of probabilistic measures instead of deterministic values which are more realistic.

3.1 Probabilistic Formulation for Power System

In conventional power flow studies, all the variables have deterministic values. In general, the form of the power flow equations is as follows.

For the power injections,

$$\begin{aligned} P_i &= V_i \sum_{k=1}^{n} V_k (G_{ik} \cos\theta_{ik} + B_{ik} \sin\theta_{ik}) \\ Q_i &= V_i \sum_{k=1}^{n} V_k (G_{ik} \sin\theta_{ik} - B_{ik} \cos\theta_{ik}) \end{aligned} \tag{3-1}$$

For the line flows,

$$\begin{aligned} P_{ik} &= -t_{ik} G_{ik} V_i^2 + V_i V_k (G_{ik} \cos\theta_{ik} + B_{ik} \sin\theta_{ik}) \\ Q_{ik} &= t_{ik} B_{ik} V_i^2 - \frac{B_{ik}}{2} V_i^2 + V_i V_k (G_{ik} \sin\theta_{ik} - B_{ik} \cos\theta_{ik}) \end{aligned} \tag{3-2}$$

where P_i and Q_i are injected active and reactive powers at bus i. P_{ik} and Q_{ik} are the active and reactive line flow in branch ik. V_i and V_k are the bus voltage magnitude at bus i and bus k respectively. θ_{ik} is the difference in voltage angles between bus i and k. G_{ik} and B_{ik} are the real and imaginary parts of the admittance matrix of branch ik respectively. t_{ik} is the transformer off nominal turn ratio of branch ik. $t_{ik} = \frac{n_p n_{s0}}{n_{p0} n_s}$. n_p and n_s are the number of

turns of primary and secondary winding. n_{p0} and n_{s0} are the nominal number of turns of primary and secondary winding. If the actual turns ratio is equal to the nominal ratio, t_{ik} =1.

For the transmission system, to calculate the line flows through the transmission lines, the following model is used.

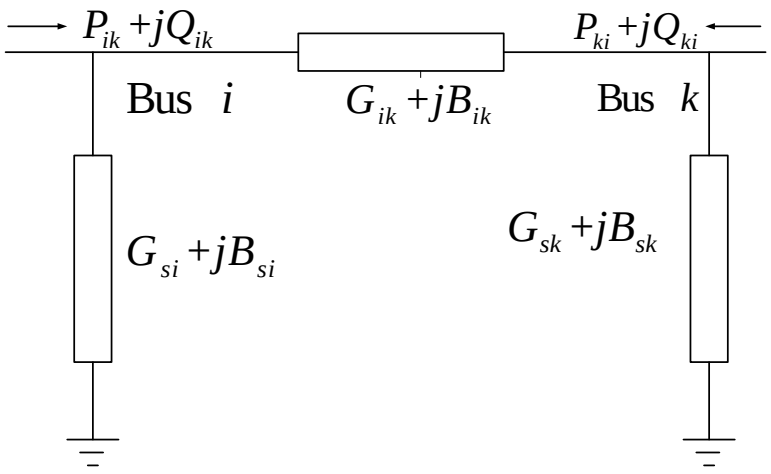

Fig. 3.1 Lumped transmission line model

The transformer model is shown in Fig. 3.2, and the equivalent circuit model of the transformer used in this power flow calculation is given in Fig. 3.3.

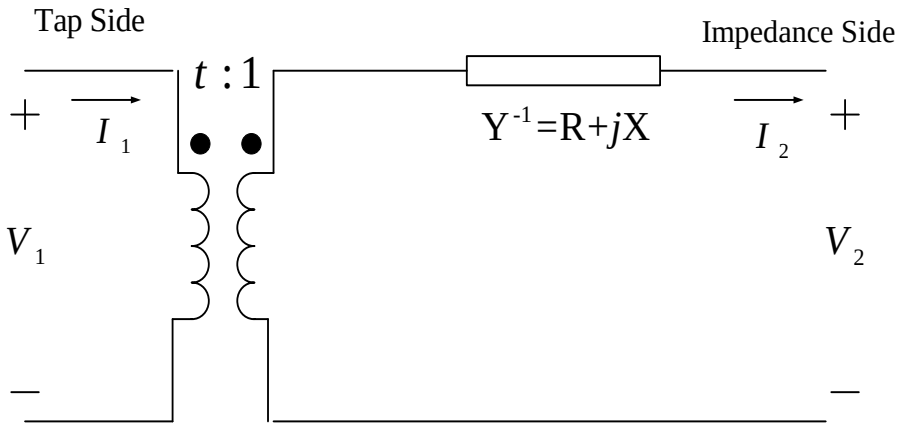

Fig. 3.2 Transformer model

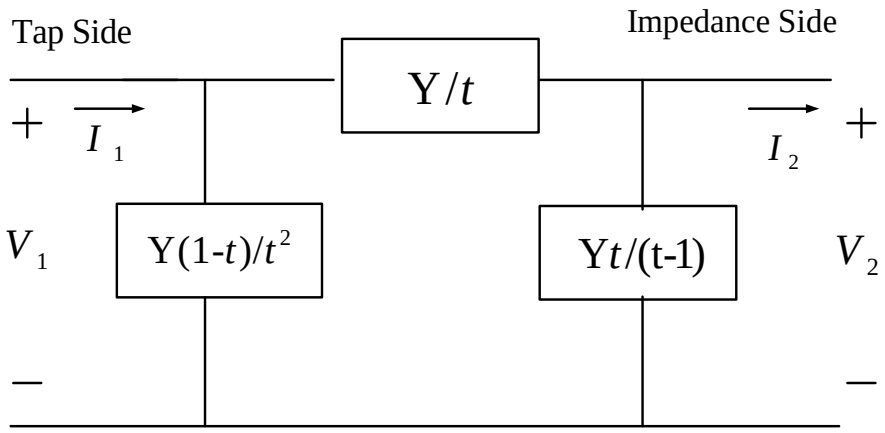

Fig. 3.3 Equivalent circuit model of transformer

In the deterministic power flow studies, the calculation is usually done using the nonlinear Newton-Raphson approach. The known quantities are the injected active power P_i at both PQ buses and PV buses except the slack bus, the injected reactive power Q_i at PQ buses, and the voltage magnitude V_i at PV buses.

Let *y* be the input variable vector for active and reactive power injections, *x* be the state variable vector for bus voltage magnitudes and angles, *z* be the output variable vector for line flow active powers and reactive powers, *g* be the bus power injection function and *h* be the line flow function. The equations for the power injections and line flows (3-1) and (3-2) can be expressed as follows,

$$\begin{cases} y = g(x) \\ z = h(x) \end{cases} \tag{3-3}$$

When the variation of power injection uncertainties is not large, the error due to linearization may be acceptable. Expanding (3-3) around the operating point by using Taylor's series and omitting the terms which are the second order and higher, the equations are shown as follows,

$$\begin{cases} \Delta x = G^{-1}\Delta y = K\Delta y \\ \Delta z = H\Delta x = HK\Delta y = L\Delta y \end{cases} \tag{3-4}$$

where Δx, Δy and Δz are the uncertainty variable vectors of *x*, *y* and *z*; *G* is the Jacobian matrix at the operating point and can be expressed as $G = \left.\frac{\partial g(X)}{\partial X}\right|_{X=X_0}$; *K* is the inverse matrix of *G* and is referred to as a sensitivity matrix; *H* can be expressed as $H = \left.\frac{\partial h(X)}{\partial X}\right|_{X=X_0}$; *L* is the sensitivity matrix of line flows and can be expressed as *L=HK*.

From the linearized models established in (3-4), Δx and Δz establish the linear relationship with Δy as follows,

$$\begin{cases} \Delta x_i = \sum_{k=1}^{n} a_{ik}\Delta y_k \\ \Delta z_j = \sum_{k=1}^{n} b_{jk}\Delta y_k \end{cases} \tag{3-5}$$

where $i = 1,2,\cdots,n$; $j = 1,2,\cdots,l$; *n* is the number of buses; *l* is the number of branches; a_{ik} and b_{jk} are the sensitivity coefficients, which are obtained from *K* and *L*, respectively.

If the PDFs of Δy_k are known as $f_{\Delta y_k}(\Delta y_k)$, let $\Delta y'_k = a_{ik}\Delta y_k$ and $\Delta y''_k = b_{jk}\Delta y_k$, then the convolution method can be applied to obtain the PDFs of Δx_i and Δz_j as follows,

$$\begin{aligned} f_{\Delta x_i}(\Delta x_i) &= \frac{1}{|a_{i1}|} f_{\Delta y_1}\left(\frac{\Delta y'_1}{a_{i1}}\right) * \frac{1}{|a_{i2}|} f_{\Delta y_2}\left(\frac{\Delta y'_2}{a_{i2}}\right) * \cdots * \frac{1}{|a_{in}|} f_{\Delta y_n}\left(\frac{\Delta y'_n}{a_{in}}\right) \\ f_{\Delta z_j}(\Delta z_j) &= \frac{1}{|b_{j1}|} f_{\Delta y_1}\left(\frac{\Delta y''_1}{b_{j1}}\right) * \frac{1}{|b_{j2}|} f_{\Delta y_2}\left(\frac{\Delta y''_2}{b_{j2}}\right) * \cdots * \frac{1}{|b_{jn}|} f_{\Delta y_n}\left(\frac{\Delta y''_n}{a_{jn}}\right) \end{aligned} \tag{3-6}$$

The expected values and standard deviations of Δx_i and Δz_j are given by,

$$\begin{aligned}
\mu_{\Delta x_i} &= a_{i1}\mu_{\Delta y_1} + a_{i2}\mu_{\Delta y_2} + \cdots + a_{in}\mu_{\Delta y_n} \\
\sigma^2_{\Delta x_i} &= a^2_{i1}\sigma^2_{\Delta y_1} + a^2_{i2}\sigma^2_{\Delta y_2} + \cdots + a^2_{in}\sigma^2_{\Delta y_n} \\
\mu_{\Delta z_j} &= b_{j1}\mu_{\Delta y_1} + b_{j2}\mu_{\Delta y_2} + \cdots + b_{jn}\mu_{\Delta y_n} \\
\sigma^2_{\Delta z_j} &= b^2_{j1}\sigma^2_{\Delta y_1} + b^2_{j2}\sigma^2_{\Delta y_2} + \cdots + b^2_{jn}\sigma^2_{\Delta y_n}
\end{aligned} \tag{3-7}$$

This is the central concept of the probabilistic power flow study, namely representing line power flow as a linear combination of random variables (the bus demands and injections).

3.2 Probabilistic Formulation for Slack Bus

Since the power injection equations of the slack bus are not included in the Jacobian matrix, the linear model for the slack bus is not included in Section 3.1. At the slack bus, both the bus voltage angle and voltage magnitude are fixed. The deterministic value of the active power P_{slack} and reactive power Q_{slack} of the slack bus can be calculated from the line flows of the branches connected with the slack bus as follows,

$$\begin{aligned}
P_{slack} &= \sum_{i=1}^{sl} P_{branch\ i} \\
Q_{slack} &= \sum_{i=1}^{sl} Q_{branch\ i}
\end{aligned} \tag{3-8}$$

where $P_{branch\ i}$ and $Q_{branch\ i}$ are the active power and reactive power of the line flow through the i^{th} branch connected with the slack bus; *sl* is the number of branches connected with the slack bus.

Thus, the uncertainty variables ΔP_{slack} and ΔQ_{slack} are represented as follows,

$$\Delta P_{slack} = \sum_{i=1}^{sl} \Delta P_{branch\ i}$$
$$\Delta Q_{slack} = \sum_{i=1}^{sl} \Delta Q_{branch\ i} \tag{3-9}$$

The PDFs of $\Delta P_{branch\ i}$ and $\Delta Q_{branch\ i}$, $f_{\Delta P_{branch\ i}}(\Delta P_{branch\ i})$ and $f_{\Delta Q_{branch\ i}}(\Delta Q_{branch\ i})$ are obtained from Section 3.1.

Thus, the PDFs of ΔP_{slack} and ΔQ_{slack} are given as follows,

$$f_{\Delta P_{slack}}(\Delta P_{slack}) = f_{\Delta P_{branch\ 1}}(\Delta P_{branch\ 1}) * f_{\Delta P_{branch\ 2}}(\Delta P_{branch\ 2}) * \cdots * f_{\Delta P_{branch\ sl}}(\Delta P_{branch\ sl})$$
$$f_{\Delta Q_{slack}}(\Delta Q_{slack}) = f_{\Delta Q_{branch\ 1}}(\Delta Q_{branch\ 1}) * f_{\Delta Q_{branch\ 2}}(\Delta Q_{branch\ 2}) * \cdots * f_{\Delta Q_{branch\ sl}}(\Delta Q_{branch\ sl}) \tag{3-10}$$

It can be observed that when the slack bus uncertainty is large, the uncertainty of the line flows through the branches connected with the slack bus is also large, since the compensation is made through these branches. If the slack bus location is changed to another generation bus, the variance characteristics of the line flows will also be changed.

3.3 Probabilistic Formulation for Apparent Power of Line Flow

In Section 3.1, the probabilistic model for both active power and reactive power through the branches are calculated. Then, the probabilistic model for the apparent power can also be obtained easily. For branch *i*, the deterministic value of the apparent power S_i can be calculated as,

$$S_i = \sqrt{P_i^2 + Q_i^2} \tag{3-11}$$

Then, the uncertainty of the apparent power is represented as a truncated Taylor series,

$$\Delta S_i = \frac{P_i \Delta P_i}{\sqrt{P_i^2 + Q_i^2}} + \frac{Q_i \Delta Q_i}{\sqrt{P_i^2 + Q_i^2}} = \frac{P_i}{S_i} \Delta P_i + \frac{Q_i}{S_i} \Delta Q_i \tag{3-12}$$

When the PDF of ΔP_i and ΔQ_i of branch *i* are obtained, let $\Delta P_i' = \frac{P_i}{S_i}\Delta P_i$ and $\Delta Q_i' = \frac{Q_i}{S_i}\Delta Q_i$, then the PDF of ΔS_i is calculated as follows,

$$f_{\Delta S_i}(\Delta S_i) = \frac{S_i}{|P_i|} f_{\Delta P_i}(\frac{S_i}{P_i}\Delta P_i') * \frac{S_i}{|Q_i|} f_{\Delta Q_i}(\frac{S_i}{Q_i}\Delta Q_i') \qquad (3\text{-}13)$$

3.4 Probabilistic Power Flow based on Cumulants

The foregoing sections introduce the probabilistic power flow based on the convolution method. The main disadvantage of this method is the computationally burdensome convolution operation which may not be suitable for large systems. On the other hand, one assumption of the convolution method is the statistical independence of the input variables. This assumption is not realistic for many actual conditions in power systems. Especially, PV generation productions in adjacent locations may be highly correlated. To solve these problems, the probabilistic power flow based on cumulants is proposed.

Based on the cumulant method mentioned in Section 2.2, the computational procedure to obtain the cumulants of the voltages and line flows consists of the following:

1. Use the conventional power flow calculation from the Newton-Raphson algorithm to compute the expected values of the state variables (bus voltage magnitudes and angles) and the output variables (line flow active powers and reactive powers). The sensitivity matrices *K* and *L* of the bus voltages and the line flows are also obtained, according to (3-4);

2. Compute the self and joint moments of the power injections (loads and PV generations) according to (2-11) and (2-37);

3. Compute the self and joint cumulants of the power injections according to the relationship between the cumulants and moments, using (2-29) and (2-40);

4. Compute the cumulants of the bus voltages and the line flows based on the linear relationship using the sensitivity matrices K and L according to (2-44) (use (2-45) for the independent case);

5. Calculate the central moments of the bus voltages and the line flows based on the relationship between the cumulants and the central moments according to (2-30);

6. Select various types of expansions (Gram-Charlier expansions, Edgeworth expansions, and Cornish-Fisher expansions) to approximate the CDFs and PDFs of the bus voltages and the line flows, and use the cumulants of the variables to calculate the constant coefficients of the selected expansion according to Section 2.3.

3.5 Evaluation of the Accuracy of PPF and MCS Results

In order to evaluate the accuracy of the PPF solution, the results are compared with the reference results. The average root mean square (ARMS) error is computed as the accuracy index [18]. The statistical points are chosen from the range of the CDFs of both PPF and MCS in a certain interval.

$$\mathrm{ARMS} = \frac{\sqrt{\sum_{i=1}^{N} (F_{\mathrm{PPF},i} - F_{\mathrm{Ref},i})^2}}{N} \tag{3-14}$$

where $F_{\mathrm{PPF},i}$ and $F_{\mathrm{Ref},i}$ are the i^{th} value on the CDF curves by the PPF algorithm and the reference, respectively. The parameter *N* is the number of selected points which are chosen from the range of the CDFs within a certain interval.

The Monte Carlo simulation (MCS) results, which can provide considerably accurate results [47] are used as a reference in the case study. The accuracy and convergence of MCS need to be evaluated. [18] uses a parameter as follows.

In MCS, the input variable *x* is a sampled as $x_1, x_2, \cdots, x_{Ns}$. $y = V(x)$. The estimated expected value of *y* is as,

$$\tilde{E}(y) = \frac{1}{Ns}\sum_{i=1}^{Ns} V(x_i) \tag{3-15}$$

where $\tilde{E}(y)$ is not the true mean value but an sample mean value. All the result variables can be estimated by (3-15) by using various function *V*. The variance of the estimated value is given by,

$$\mathrm{var}\left[\tilde{E}(y)\right] = \frac{\mathrm{var}(y)}{Ns} \tag{3-16}$$

where $\mathrm{var}(y)$ is the variance of *y*. Equation (3-16) indicates that the uncertainty of the estimated value depends on the variance of the test function *V* and the sampling number *Ns*. The accuracy of MCS increases with a larger sample number. The uncertainty is usually represented as the coefficient of variation,

$$\beta = \frac{\sqrt{\mathrm{var}\left[\tilde{E}(y)\right]}}{\tilde{E}(y)} \tag{3-17}$$

Substituting (3-16) into (3-17),

$$Ns = \frac{\mathrm{var}(y)}{\left[\beta \tilde{E}(y)\right]^2} \tag{3-18}$$

Equation (3-18) can be used to determine the required number of simulations for a given β. A larger number of simulation reaches low uncertainty level β, but the computational burden also increases.

3.6 Confidence Level and Overlimit Probability

A confidence level is an interval in which a measurement or trial fails corresponding to a given probability. System planning studies are aimed at determining the value of 10% and 90% probability level that the voltage magnitude and line flow will not exceed specified limits, since this would roughly indicate the desired range of the variables [19]. Consequently, accurate estimation of 10% and 90% confidence levels has important meaning to a system planner. The confidence level can be read directly from the CDF curve of the expected variable. Therefore, the confidence level of the bus voltage magnitudes and line flows can be obtained. For example,

Fig. 3.4 shows a CDF curve of a bus voltage magnitude, and the confidence interval is indicated from 1.0460 to 1.0493.

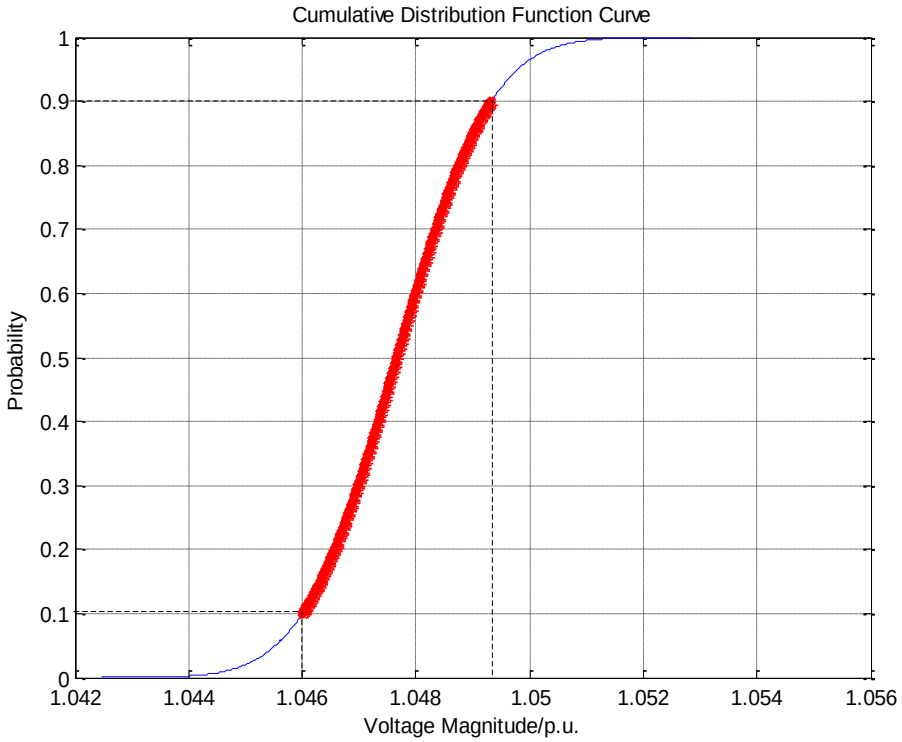

Fig. 3.4 Cumulative distribution function curve

When the CDF $F(v)$ of the variable *v* is computed, the complementary distribution function $L(v)$ can be obtained, which is defined as follows,

$$L(v) = 1 - F(v) = P(V > v) \tag{3-19}$$

The complementary distribution function represents the probability that the variable exceeds a certain value. Therefore, the overlimit probability of the variable *v* is $L(v_{\text{limit}})$. For example, a complementary distribution function curve is shown in Fig. 3.5. From this curve, the overvoltage probability (greater than 1.05) is observed to be 3.216%. It can be used in power system reliability analysis.

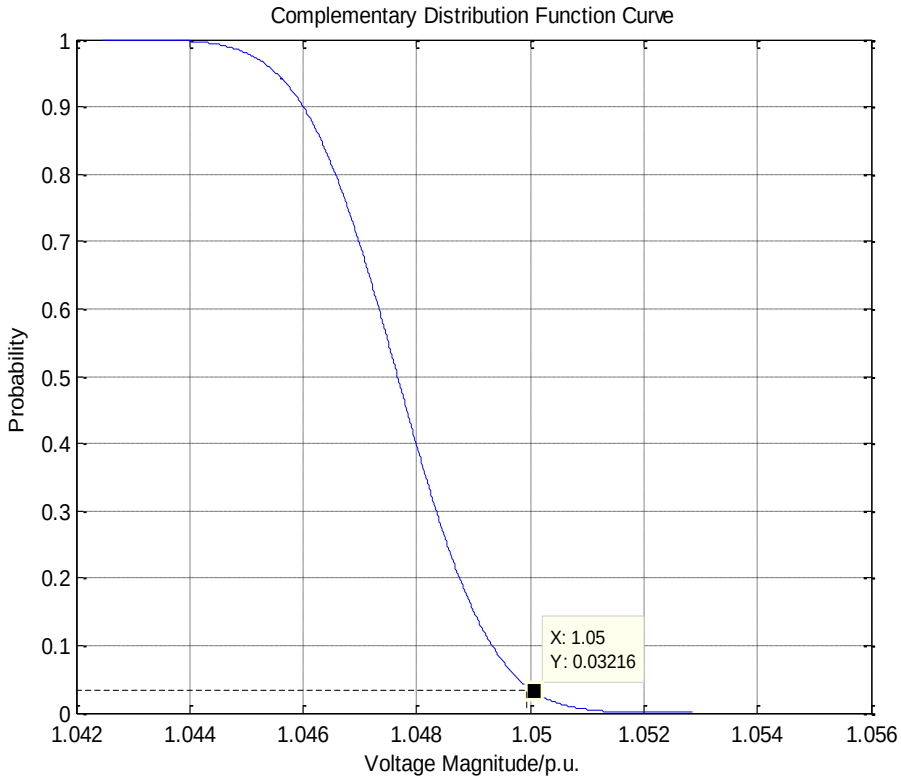

Fig. 3.5 Complementary distribution function curve

Hourly simulations can be automated by repeating the PPF algorithm to obtain the bus *i* voltage CDF $F_{ij}(v_i)$ in hour *j* of the day, and the complementary distribution function is $L_{ij}(v_i) = 1 - F_{ij}(v_i)$. From the complementary distribution function, the overvoltage probability of bus *i* in hour *j* is calculated by the equation $P_{ij} = L_{ij}(V_{\lim it})$. Then, the expected overvoltage time (in hours) for bus *i* per day is obtained by the equation $\sum_{j=1}^{24}(P_{ij} \times 1h)$.

For the whole system, the overvoltage probability in hour j is $P_{system,j} = 1 - \prod_{i=1}^{n}(1 - P_{ij})$, and the expected overvoltage time (in hours) of the system per day is $\sum_{j=1}^{24}(P_{system,j} \times 1h)$.

By repeating the above steps, the daily overvoltage probability and expected overvoltage time for one year can be calculated. Then the annual expected overvoltage time for bus i is $\sum_{k=1}^{365}\sum_{j=1}^{24}\left(P_{ij}^{k} \times 1h\right)$, and the annual expected overvoltage time for the system is $\sum_{k=1}^{365}\sum_{j=1}^{24}(P_{system,j}^{k} \times 1h)$. If it is roughly assumed that the everyday overvoltage probability is the same, the annual expected overvoltage time for bus i is $365 \times \sum_{j=1}^{24}(P_{ij} \times 1h)$, and the annual expected overvoltage time for the system is $365 \times \sum_{j=1}^{24}(P_{system,j} \times 1h)$.

Chapter 4

PROBABILISTIC MODELS OF PV GENERATION AND LOAD

As stated in Chapter 3, the probabilistic models of power injections need to be established before applying the PPF algorithm. Section 4.1 and 4.2 introduce the probabilistic models of both PV generation and load, respectively. The probabilistic characteristics of PV generation and load are evaluated.

4.1 Probabilistic Model of PV Generation

PV generation can only work in the daytime, and its production is easily influenced by the environmental conditions since the PV output depends on the insolation [33]. The quantity and quality of the solar irradiance are variable related to the geographical condition and time, which is predictable. However, meteorological conditions such as cloud and fog are less predictable and act quickly. Therefore, the PV generation output is not easy to be controlled by the system operators. The probabilistic characteristic of PV resource should be different in various locations and time with different weather conditions.

4.1.1 Uncertainty Analysis for Actual History Data of PV Generation

To study the uncertainty of PV generation production, the actual history data of PV generation is observed. A typical PV production curve measured at 12:00 pm in Tempe, Arizona for a period for a period of two years is shown in Fig. 4.1, and the probability density function (PDF) curve is shown in Fig. 4.2.

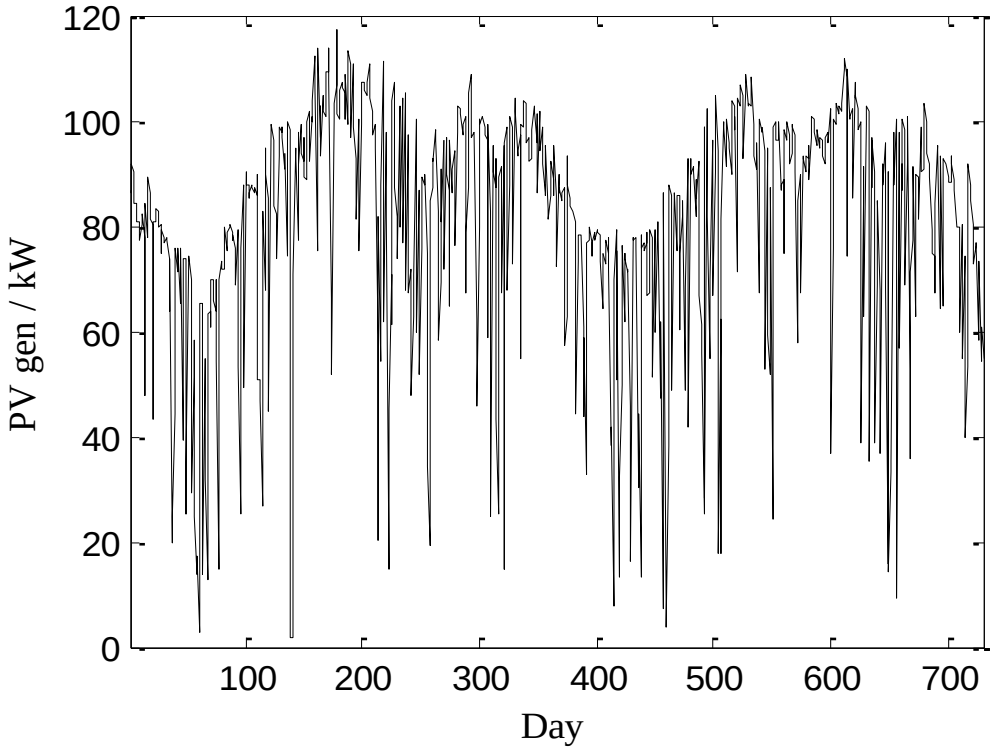

Fig. 4.1 Production fluctuation of PV generation from 2009 to 2010

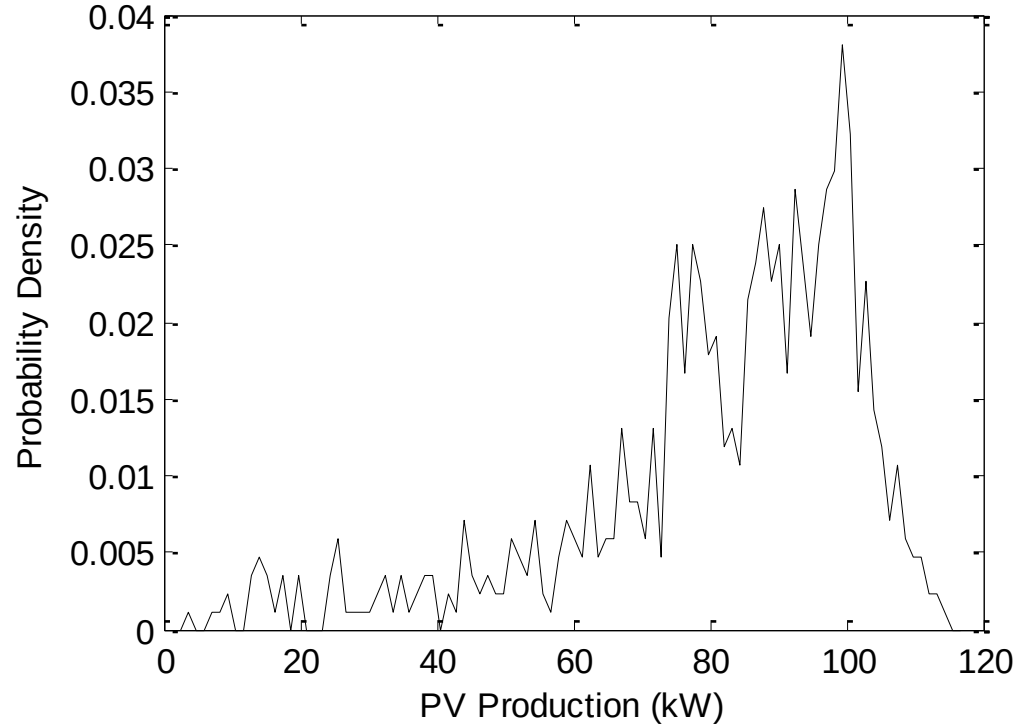

Fig. 4.2 Probability density function curve of the PV production

The actual data of a typical PV generation (shown in Fig. 4.1) is analyzed. It is observed that the PV generation production is periodic. For this reason, the FFT method is applied to obtain the amplitude spectrum in frequency domain. The result excluding the DC component is shown in Fig. 4.3.

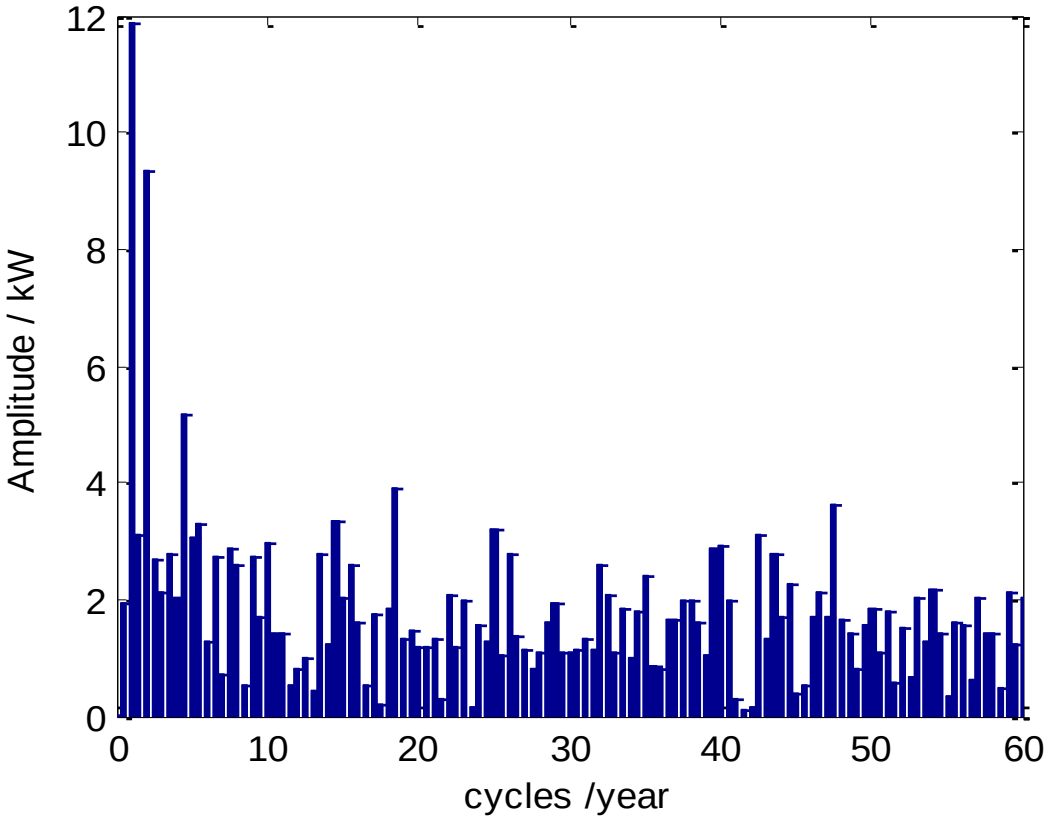

Fig. 4.3 Amplitude spectrum of PV generation production

Fig. 4.3 indicates the PV generation production, which is easily affected by seasonal environmental conditions is highly periodic. The periodic components which can be easily predicted are deterministic in nature. In the absence of meteorological factors, PV generation output is very predictable, since the solar orbit can be determined.

Solar elevation angle for specific location and solar time, which is the angle between the direction of the geometric center of the sun and the horizon, can be determined as follows [56],

$$\sin\theta_S = \cos\omega\cos\delta\cos\phi + \sin\phi\sin\delta \tag{4-1}$$

where θ_S is the solar elevation angle; ω is the solar time; δ is the current sun declination; ϕ is the local latitude.

Tempe is located at a latitude of 33.41 ℃N, longitude of 111.91 ℃E. According to (4-1), sin θ_S at 12:00 pm for two years is shown in Fig. 4.4.

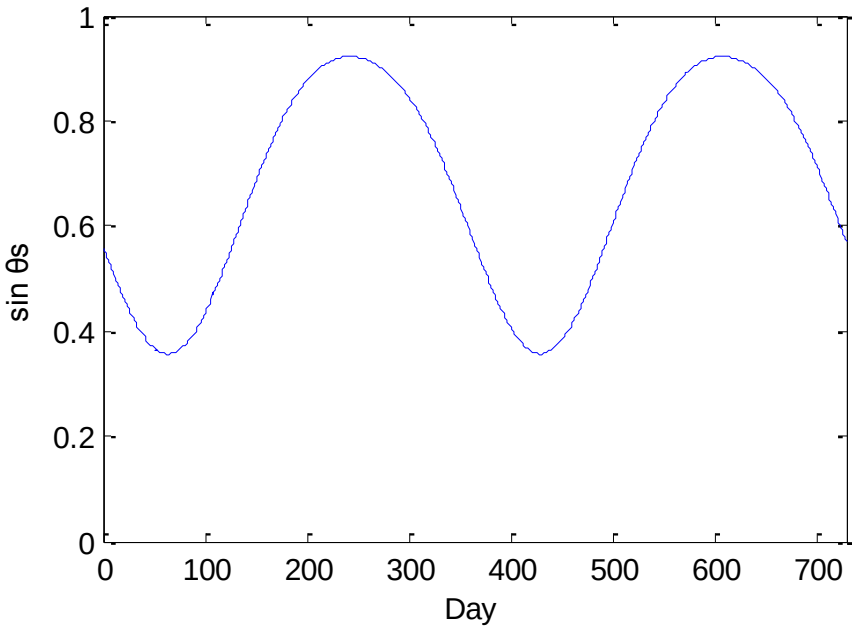

Fig. 4.4 Curve of sin θ_S at 12:00 pm for two years.

The extraterrestrial solar irradiance is approximately in portion of sin θ_S, and this variation component is easily determined. To evaluate the uncertainty of PV generation, the least square method is applied to determine the periodic component which is predict-

able in the system operation. In this method, the PV production is assumed as a linear function of sin θ_S,

$$P=a\sin\theta_S+b \tag{4-2}$$

where a and b are unknown parameter. Let Y be the PV production, and X be sin θ_S, a and b can be determined when the actual data of X and Y are already known.

$$a=\frac{n\sum_n XY-\sum_n X\sum_n Y}{n\sum_n X^2-(\sum_n X)^2},\quad b=\bar{Y}-a\bar{X} \tag{4-3}$$

Without considering meteorological and weather factors, the PV production P = sin θ_S is drawn in Fig. 4.5.

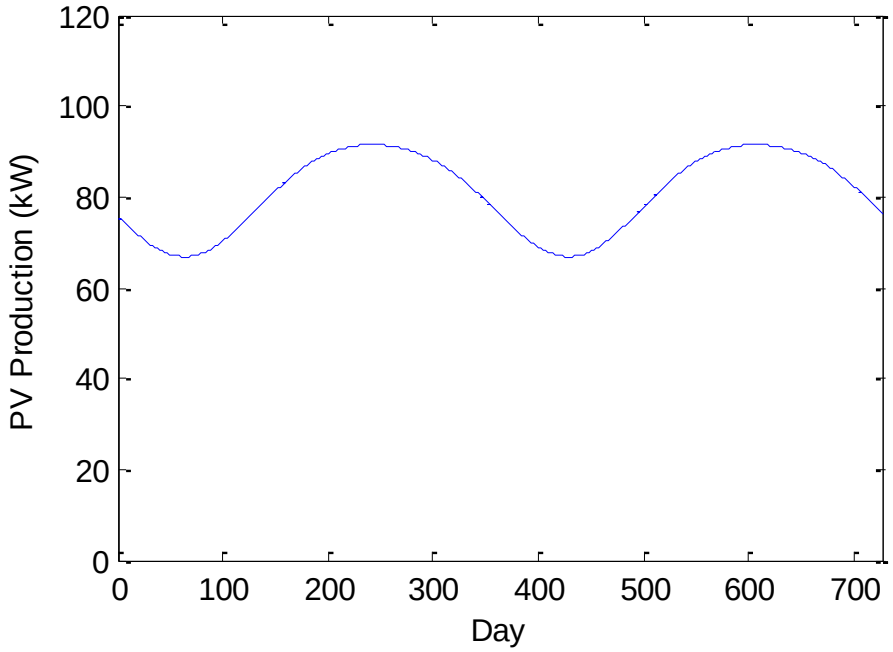

Fig. 4.5 Production curve of PV generation P = sin θ_S

ε is assumed to be a residual of the linear model, which is considered to be the unpredictable component of PV generation uncertainty.

$$\varepsilon = Y - (aX + b) \tag{4-4}$$

By applying the least squares method, the PDF curve of unpredictable component of PV production is shown in Fig. 4.6. According to the result curve, the unpredictable uncertainty is still large after removing the changing solar position impact. It indicates that the changing weather condition is the dominant factor for the PV generation output.

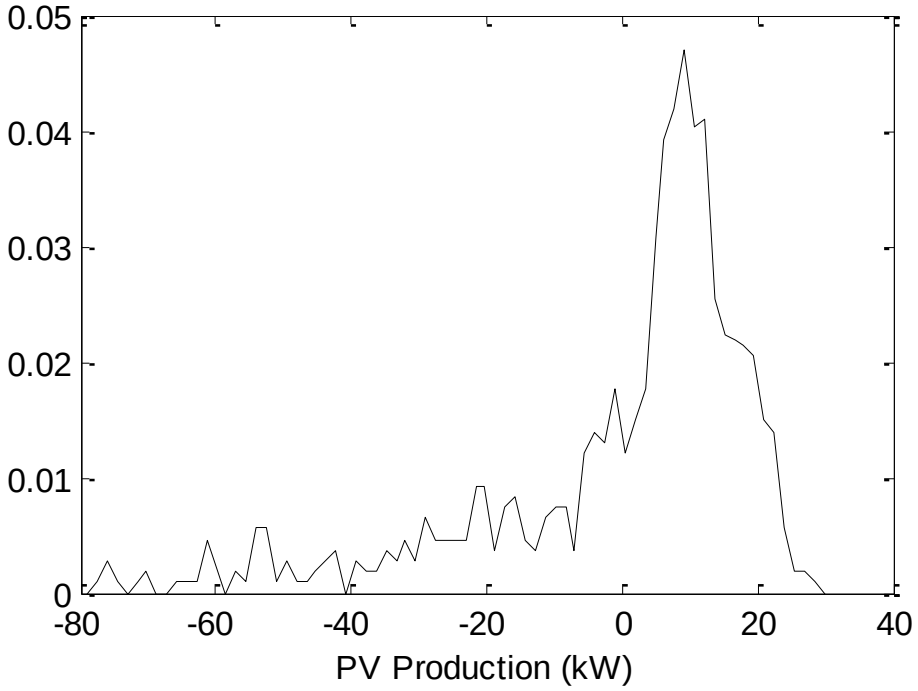

Fig. 4.6 PDF curve of unpredictable component of PV production

This analysis method can be applied to estimate the probabilistic characteristics of PV generations in various locations with different weather conditions.

As a comparison, another result is obtained by simply filtering the periodic components of PV production (shown in Fig. 4.1). The production curve regarded as the uncertainty of PV generation is shown in Fig. 4.7, and the PDF curve is shown in Fig. 4.8. Table 4.1 compares the standard deviation and coefficient of variation (CV) of PV production in different cases.

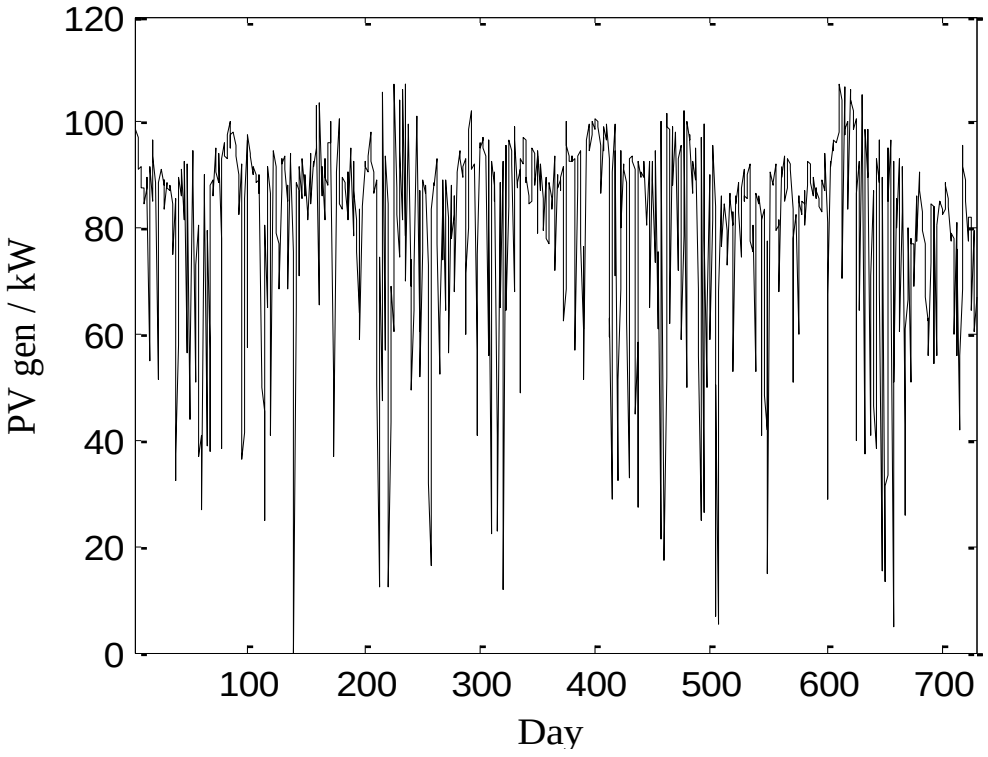

Fig. 4.7 Production fluctuation of PV generation after filtering

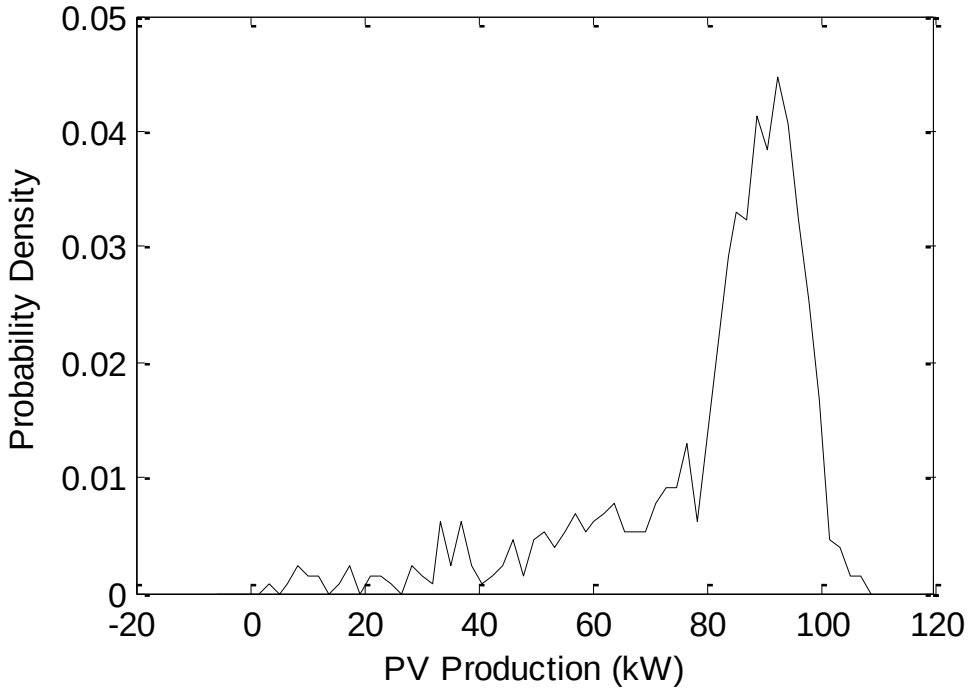

Fig. 4.8 PDF curve of PV production after filtering the periodic components

Table 4.1 Comparison of the uncertainty results of PV production

	Fig. 4.2	Fig. 4.5	Fig. 4.6	Fig. 4.8
Standard deviation (kW)	22.42	8.77	20.64	19.84
Coefficient of variation* (%)	27.81%	10.88%	25.60%	24.61%

* Coefficient of variation – the ratio of the standard deviation to expected value of PV production (80.62 kW).

The results demonstrate that the periodic variation of PV generation is mainly caused by the solar position in accordance with an annual cycle. It can be also observed that PV production has a small and predictable variation on sunny days.

4.1.2 Probabilistic Model of PV Generation

The actual history data of PV generation production discussed in Section 4.1.1 can be used to calculate the moments and cumulants directly. According to the actual data, the probabilistic model can be established to predict the distribution of PV generation production.

In order to establish the probabilistic model of PV generation, the performance model should be studied first. The modeling step requires a thorough understanding of the physical process. When designing a PV generation system, it is necessary to predict its expected energy production. References [33] showed that the total active power P has a linear relationship with solar irradiance *r* as follows,

$$P = rA\eta \tag{4-5}$$

where *r* denotes the solar irradiance; *A* is the total area of the PV module; η is the PV generation efficiency.

However, another reference [8] proposed that the PV generation production is a function of the insolation, ambient temperature, and prevailing wind speed. The last two factors are influential in determining the cell operating temperature which, in turn, affects the output. Reference [34] also gave a practical data curve of the PV generation production shown in Fig. 4.9.

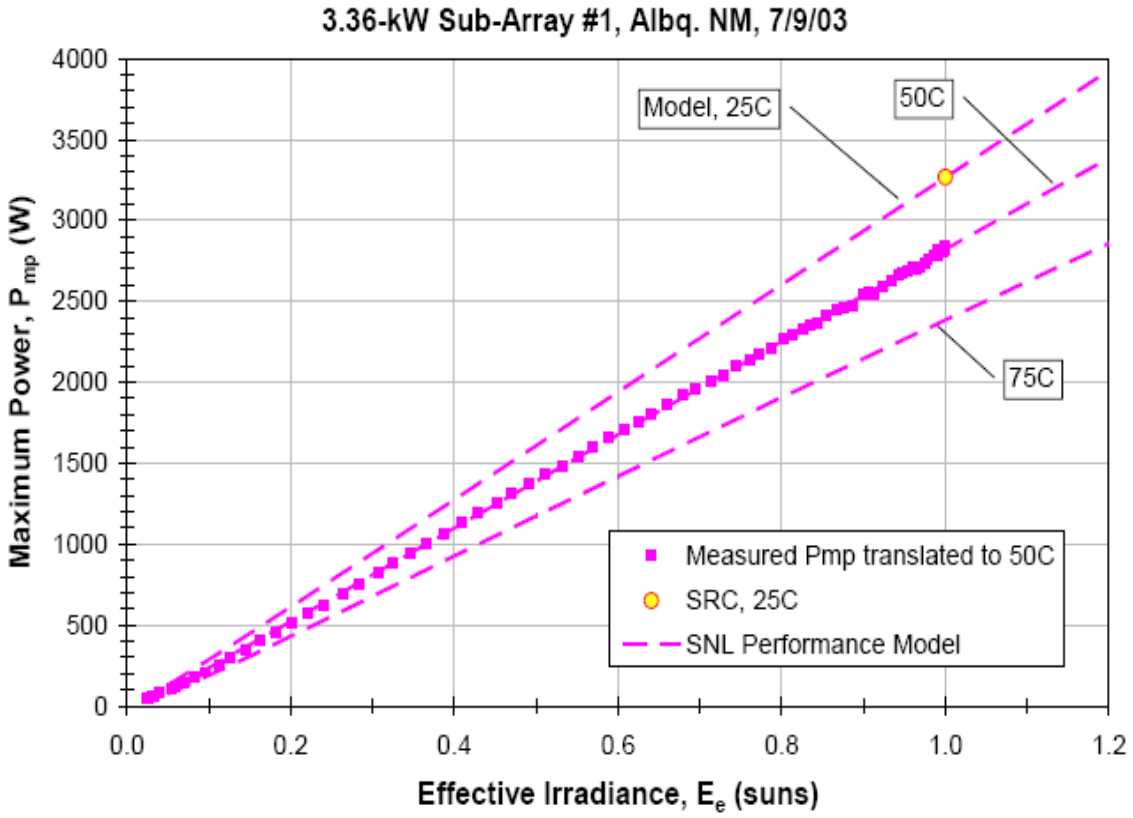

Fig. 4.9 PV power curve at three operating temperatures [34]

Fig. 4.9 shows that the PV generation power production is as a function of effective irradiance and cell operating temperature. From the solar cell characteristics, there is a temperature coefficient *k*. When the cell operating temperature increases by 1℃, the active power output decreases by *k*. *k* is provided by PV generation manufacturers. In a

practical PV solar module in [8], $k = -0.52\%$. Thus, the active power output of PV generation can be expressed as follows,

$$P = rA\eta(1 - k\Delta T) \tag{4-6}$$

where ΔT is the uncertainty of the PV cell temperature. Let $R = rA\eta$ and $T = 1 - k\Delta T$,

$$P = RT \tag{4-7}$$

In the PPF algorithm, the PDF of PV generation output where the probability is approximated as a function needs be defined. Let the PDFs of R and T be $f_R(R)$ and $f_T(T)$. The parameters *R* and *T* are independent, so that their joint PDF is as follows,

$$f_{R,T}(R,T) = f_R(R)\, f_T(T) \tag{4-8}$$

According to the following theorem of the PDF of a product (proved in Appendix C), the PDF of PV generation output can be obtained. If *X* and *Y* have joint PDF *f* (*u*), show that the PDF of *U* = *XY* is given by,

$$f_U(u) = \int_{-\infty}^{\infty} f(x, u/x)\left|x\right|^{-1} dx \tag{4-9}$$

Thus, the PDF of *P* is as follows,

$$f_P(P) = \int_{-\infty}^{\infty} f_R(R)\, f_T(\frac{P}{R})\left|R\right|^{-1} dR \tag{4-10}$$

According to [32], [33], the solar irradiance *r* accounts for cloud cover and other insolation reducing phenomena during a certain interval. It has been shown that the solar irradiance *r* can be described reasonably well by a beta distribution,

$$f_r(r) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{r}{r_{\max}}\right)^{a-1}\left(1 - \frac{r}{r_{\max}}\right)^{b-1} \tag{4-11}$$

where $r_{\max}$ is the maximum solar irradiance; *a* and *b* are the shape parameters of the distribution. $a = \mu\left[\mu(1-\mu)/\sigma^2 - 1\right]$, $b = (1-\mu)\left[\mu(1-\mu)/\sigma^2 - 1\right]$.

Let $R = rA\eta$, the PDF of R is deduced as follows,

$$f_R(R) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{R}{R_{\max}}\right)^{a-1}\left(1-\frac{R}{R_{\max}}\right)^{b-1} \tag{4-12}$$

where $R_{\max}$ is the maximum value of *R*.

ΔT is the forecast error of the PV cell temperature. Its PDF can be obtained from historical data. The forecast error can be used to reflect the random variable. In the absence of the PDF of the forecast error, the normal distribution is used often to express the forecast error, since it has traditionally been considered adequate for representations of certain forms of random error [15]. Therefore, in this dissertation, ΔT is assumed to follow the normal distribution, $\Delta T \sim N(0, \sigma_{\Delta T}^2)$.

$$f_{\Delta T}(\Delta T) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta T}} e^{-\frac{\Delta T^2}{2\sigma_{\Delta T}^2}} \tag{4-13}$$

Let $T = 1 - k\Delta T$, $T \sim N(1, (k\sigma_{\Delta T})^2)$. Reference [34] proposed the expected PV cell operating temperature with an accuracy of about $\pm 5^{\circ}\mathrm{C}$. So this dissertation assumes that $\mu_{\Delta T} = 0$, $\sigma_{\Delta T} = 5$. $\Delta T \sim N(0, 25)$. $T \sim N(1, 25k^2)$.

Then, the PDF of *P* can be expressed as follows,

$$f_P(P) = \int_{-\infty}^{\infty} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{R}{R_{\max}}\right)^{a-1}\left(1-\frac{R}{R_{\max}}\right)^{b-1} 1/\sqrt{2\pi}k\sigma_{\Delta T} \exp\left[-(P/R-1)^2/2k^2\sigma_{\Delta T}^2\right]|R|^{-1}\, dR \tag{4-14}$$

Fig. 4.10 shows the proposed probabilistic model curve in comparison with the PDF of the actual data excluding the periodic data discussed in Section 4.1.1.

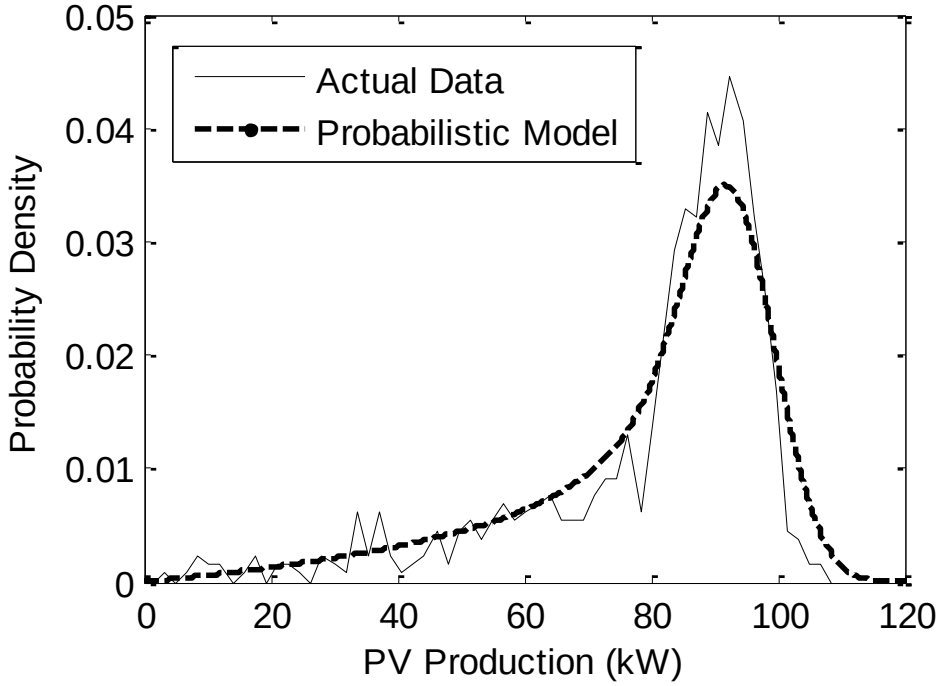

Fig. 4.10 PDF curve of PV generation for both probabilistic model and actual data

To quantify the accuracy of the probabilistic model, the ARMS error is computed. Regarding the actual data as the reference, the ARMS error in Fig. 4.10 is 0.377%. It is observed that the novel probabilistic model gives an accurate approximation.

4.2 Probabilistic Model of Loads

Loads are randomly distributed variables, because there are many exogenous factors that could influence loads, such as the variation in customer's behaviors, the variation in environmental conditions, and the variation in electrical appliances and installations. The probabilistic models of loads may vary significantly for different customers and conditions. Therefore, the probabilistic model for the load needs to be established by

observing the historic data of specific loads. There are 46 available models for different customer classes in [48]. The most usual assumption for electric loads is normal distribution. In [49], the load model is described as a beta probability distribution. In the CAISO load modeling approach, the obtained load is established as a hyperbolic distribution [50].

In this study, the actual data of loads is observed. The load data were obtained from the Electric Reliability Council of Texas (ERCOT) [54]. The same steps as for the PV generation model are applied to load data. The periodic components are removed from the actual load data by simple subtraction of the assumed form. The PDF curve of the load uncertainty is given in Fig. 4.11. The coefficient of variation (CV) of the load is 10.60%. The ARMS error of the load model in Fig. 4.11 is 0.0139%. According to the depicted curve, it is reasonable to assume that this is a rough approximation of a normal probability density.

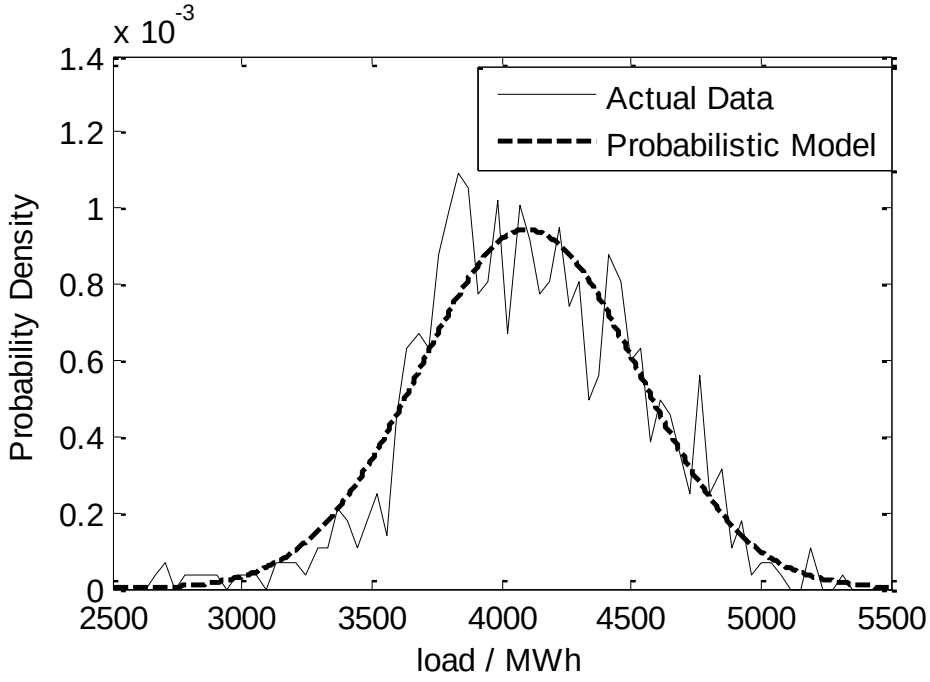

Fig. 4.11 PDF curve of load for both probabilistic model and actual data

Chapter 5

PROBABILISTIC POWER FLOW CONSIDERING GENERATION DISPATCHING OPERATION

In the previous analysis, the behavior of conventional generation operation is not considered in the probabilistic model of power systems, and it assumed that the generators' outputs are fixed and all the uncertainties of PV generations and loads are compensated by the slack bus alone, which is not realistic. This assumption could be acceptable when the variances of PV generations have a small value or are located near the slack bus.

When the PV penetration increases, the stochastic characteristic of PV generation might exert detrimental impacts on the active power balance in bulk systems. These variations of PV generation sources should be compensated by the generation dispatching operation with automatic generation control (AGC) or the day-ahead generation scheduling settlement, and several generation sources may need to be adjusted. Therefore, it is more realistic to consider the generation dispatch in the PPF algorithm.

As the foregoing discussion illustrates, the representation of the utility company operating policy in the event of generation surplus or deficiency forms an important aspect of the PPF computation [43], so that the correct allocation of conventional generation is also considered in the PPF algorithm. Under actual conditions, the generators should be scheduled to accommodate the uncertainties in power systems.

5.1 Probabilistic Model of Generation Dispatching Operation

In actual applications, realistic dispatching laws are complicated and could cause significant computational problems. Therefore these dispatching laws are complicated for

the PPF algorithm to handle. To find a meaningful approach to represent the dispatching laws, a linear model is adopted [43]. The linear model of generation dispatching operation is given by,

$$\Delta P_{gen} = T\Delta P_{PV} + \Delta P_{gen,F} \tag{5-1}$$

where ΔP_{gen} and ΔP_{PV} are the uncertainty variable vectors of P_{gen} and P_{PV}. T is the sensitivity matrix of conventional generation. T_{ij} represents the change in the i^{th} generation for a unity change in the j^{th} PV generation. $\Delta P_{gen,F}$ is the redistribution of the available generation for some generation outage. Under normal conditions (no generator outages), $\Delta P_{gen,F} = 0$.

$$\Delta P_{gen} = T\Delta P_{PV} \tag{5-2}$$

In this research, the outage conditions are excluded. References [11] and [51] considered the outage conditions. This dissertation only considers the dispatching strategy of balancing the PV generation variation.

Equation (5-2) shows how the generation sources balance the changes of PV generation. This model also indicates that the variations of conventional generation sources depend on the variations of the PV generation and various dispatching strategies. The conventional generation production variations are not independent variables and are functions of PV generation uncertainties. It is expected that systems with very high penetration of PV resources will require attention to the high ramp rate capabilities of conventional generation resources to balance the PV generation. For the generators with low ramp rate, the associated element value in the T matrix in (5-2) is set to zero.

The T matrix may be selected as follows [43],

- by deciding how a PV generation change is to be shared among the neighboring generators;
- by fitting a linear model over a reasonable set of conditions as obtained using an optimal power flow program;
- by using optimal sensitivity equations, which adjust the generations to minimize the cost [55].

Since the probabilistic model of PV generation is known, the probabilistic model of each conventional generation production can be obtained based on the established dispatching model in (5-2).

5.2 Probabilistic Model of Power System Considering the Generation Dispatching Operation

In Section 3, the linear models for bus voltages and line flows have been formulated without considering the generation dispatching operation in (3-4). When the linear model of the generation dispatching operation is established, the linear model of the power system is revised. Expanding the first equation in (3-4) (the linear model of power injection),

$$\begin{pmatrix} \Delta\delta \\ \Delta|V| \end{pmatrix} = K \begin{pmatrix} \Delta P \\ \Delta Q \end{pmatrix} = \left(K_{P_L}, K_{P_{PV}}, K_{P_{gen}}, K_Q \right) \begin{pmatrix} \Delta P_L \\ \Delta P_{PV} \\ \Delta P_{gen} \\ \Delta Q \end{pmatrix} \tag{5-3}$$

Then, (5-3) can be modified as follows,

$$\begin{pmatrix} \Delta\delta \\ \Delta|V| \end{pmatrix} = \left(K_{P_L}, K_{P_{PV}}, K_Q\right) \begin{pmatrix} \Delta P_L \\ \Delta P_{PV} \\ \Delta Q \end{pmatrix} + K_{P_{gen}} \Delta P_{gen} \tag{5-4}$$

Substituting equation (5-2) in equation (5-4),

$$\begin{pmatrix} \Delta\delta \\ \Delta|V| \end{pmatrix} = \left(K_{P_L}, K_{P_{PV}}, K_Q\right) \begin{pmatrix} \Delta P_L \\ \Delta P_{PV} \\ \Delta Q \end{pmatrix} + K_{P_{gen}} T \Delta P_{PV} \tag{5-5}$$

Then, equation (5-5) is rewritten as,

$$\begin{pmatrix} \Delta\delta \\ \Delta|V| \end{pmatrix} = \underbrace{\left[K_{P_L}, \left(K_{P_{PV}} + K_{P_{gen}} T\right), K_Q\right]}_{K_{amend}} \begin{pmatrix} \Delta P_L \\ \Delta P_{PV} \\ \Delta Q \end{pmatrix} \tag{5-6}$$

By adding the generation dispatching operation model, the *K* matrix is revised to K_{amend} matrix. The *L* matrix should be also revised as $L_{amend} = HK_{amend}$. The linear model of the power system in (3-4) is now changed to be,

$$\begin{cases} \Delta x = K_{amend} \Delta y \\ \Delta z = HK_{amend} \Delta y = L_{amend} \Delta y \end{cases} \tag{5-7}$$

Since the generation dispatching operation is used to compensate the uncertainties of PV generation resources, the uncertainty influence on power systems will be changed by using the probabilistic model in (5-7).

5.3 Probabilistic Optimal Power Dispatching Strategy

As mentioned in the above sections, the high level of uncertainty of PV generation may seriously impact power system performance. However, the generation dispatch can compensate the change of PV generation, limit the influence of the PV, and improve

power system reliability. Based on the PPF algorithm considering generation dispatch, a dispatching strategy is proposed considering both the economic cost and the uncertainty influence of the PV. The objective of the dispatch is to minimize the total cost of generations.

Since the PDF of PV generation is known (see Section 4), based on the established linear model of generation dispatching operation shown in (5-2), the moments and cumulants of the generation output can be computed by using the cumulant method. The PDF of the generation output can be approximated by using Gram-Charlier Type A expansion,

$$f_{P_{gen,i}}(P_{gen,i}) = \frac{1}{\sigma_{gen,i}} \sum_{j=0}^{n} \frac{c_j}{j!} \varphi^{(j)}\left(\frac{P_{gen,i} - \mu_{gen,i}}{\sigma_{gen,i}}\right) \tag{5-8}$$

The cost function for each generator is given as $C_i(P_{gen,i})$, the expected value of the generation cost is,

$$E\left[C_i(P_{gen,i})\right] = \int C_i(P_{gen,i}) f_{P_{gen,i}}(P_{gen,i}) dP_{gen,i} \tag{5-9}$$

In the actual system operation, the operators prevent the overload of each system branch. Since the CDF $F_k(x)$ of the line flow k can be approximated from the proposed PPF algorithm using the Gram-Charlier expansion, the overload probability (OLP) of the line flow k is given by,

$$\mathrm{OLP}_k = 1 - \sum_{j=0}^{n} \frac{c_j}{j!} \phi^{(j)}\left(\frac{x_{\mathrm{limit},k} - \mu_{\mathrm{branch},k}}{\sigma_{\mathrm{branch},k}}\right) \tag{5-10}$$

where $x_{\mathrm{limit},k}$ is the thermal limit of the line flow k.

Then, the probabilistic optimal dispatching strategy is established as follows,

- The objective of the dispatch is to minimize the expected value of total generation cost based on (5-9);
- The OLP of each line flow, which is calculated based on the proposed PPF algorithm considering the generation dispatch according to (5-10), is constrained to a certain limit;
- The changes in PV generation levels are balanced by the dispatching strategy based on (5-2), and the sum of the power increments of PV generations is compensated by the sum of the increments of conventional generators.

$$\begin{aligned} &\min \ \sum_{i}^{N_G} E\left[C_i(P_{gen,i})\right] \\ &\text{Subject to} \\ &\quad \mathrm{OLP}_k \le \xi \\ &\quad \Delta P_{gen} = T\Delta P_{PV} \\ &\quad \sum_{i}^{N_G} \Delta P_{gen,i} = -\sum_{i}^{N_{PV}} \Delta P_{PV,i} \\ &\quad -1 \le T_{ij} \le 0 \end{aligned} \tag{5-11}$$

where ξ is the overload probability limit. N_G and N_{PV} are the number of generations and the number of PV generation resources, respectively. OLP_k represents the overload probability of line k; and T represents the matrix of the dispatching law. For the generators that cannot perform fast rescheduling, the associated element value in the T matrix can be set to zero. This would mean that only the generators with high ramp rates are considered in the optimal dispatching strategy.

By solving this optimization problem, the optimal dispatching matrix T can be obtained. This optimal dispatching strategy considering uncertainty factors may lead to a

less expensive and better dispatch in the sense that uncertainty is modeled rather than ignored.

Chapter 6

CASE STUDY

This chapter presents an example of the use and performance of the proposed PPF algorithm on a realistic power system. The results and analysis of the study conducted are presented.

6.1 Test System Description

The proposed PPF algorithm and probabilistic optimal power dispatching strategy have been implemented using MATLAB, and tested on the Arizona area of the WECC transmission system consisting of 2497 buses and 2971 lines. The test system has 1070 loads, 174 conventional generators and 179 PV generation resources. A full description of this system is found in [53], and a representation of the area is shown in Fig. 6.1.

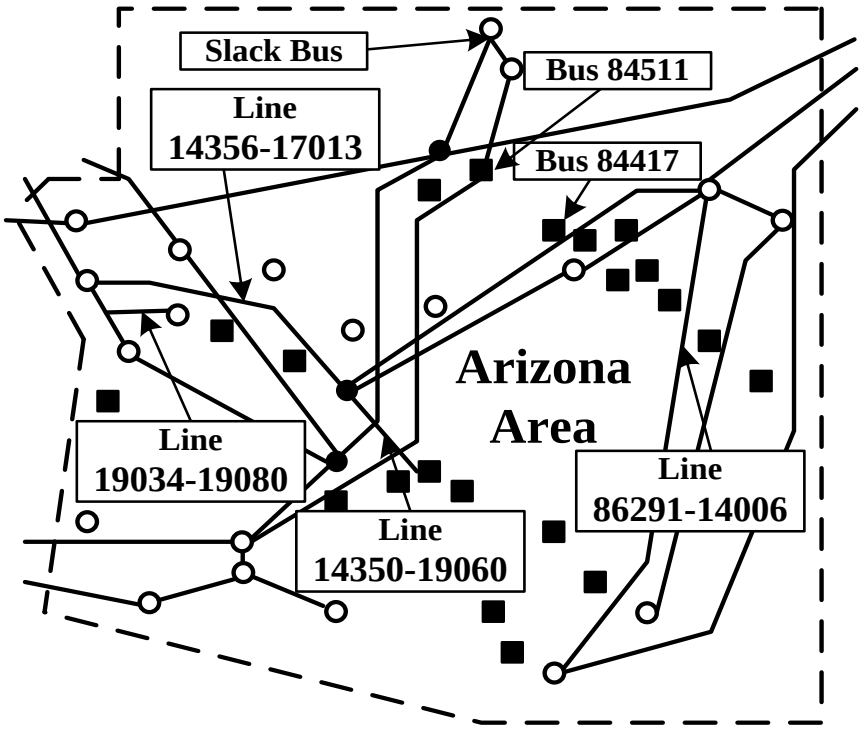

Fig. 6.1 Simplified portion of the WECC in Arizona (■– PV generation locations; ○ – conventional generator locations)

In the case study, the probabilistic models of PV generation and load discussed in Section IV are applied with the same CVs. The PV generation penetration is varied up to 20%.

6.2 PPF Results for Various Types of Expansions

In this section, the three expansions (Gram-Charlier expansion, Edgeworth expansion, and Cornish-Fisher expansion) mentioned in Section 2.3 are applied to approximate the CDF of both system voltages and line power flows.

The PV generation penetration is set to be 20% of the load. The *correlation coefficient* among the PV resources is set to be 0.5 for illustration purposes. The case study takes into account the first six self cumulants. The second order joint cumulants among PV resources is considered, and it is assumed that the joint cumulants of order higher than two are zero. In order to evaluate the accuracy of the result, MCS with 10,000 samples is utilized as a comparative reference, and the parameter β (shown in (3-18)) is set as less than 1% for all the result variables. The MCS uses the nonlinear Newton-Raphson approach. The ARMS error is computed as the accuracy index of the approximation expansions. The CDF curves from MCS and the three different expansions are analyzed for the bus voltages and line flows of the test system. In the case study, the observed buses and branches are near the PV generation locations and are subjected to more uncertainty influence.

Fig. 6.2 and Fig. 6.3 show the CDF curves of the voltage magnitude at bus 84417 and the line flow through line 14356-17013. Table 6.1 and Table 6.2 compare the mean value, the standard deviation, the ARMS error, the confidence level (10% and 90%),

steady-state overvoltage (or overload conditions) probability and computing time. Fig. 6.4 and Table 6.3 show the probabilistic results of the slack bus output.

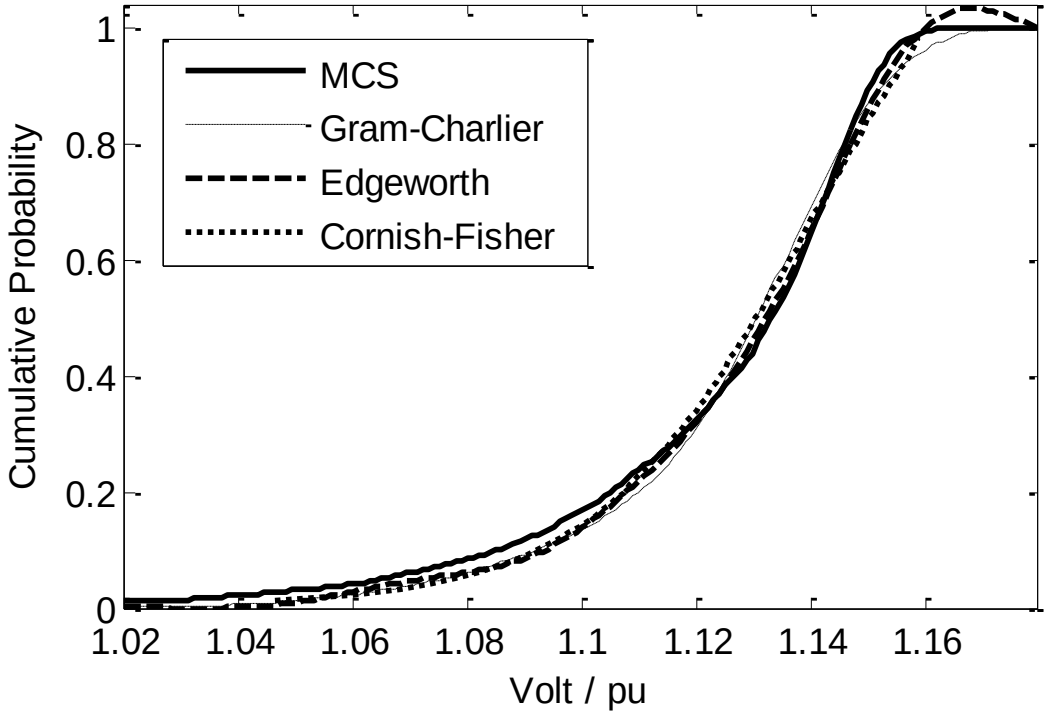

Fig. 6.2 CDF curves of the voltage magnitude at bus 84417 for three different types of expansions

Table 6.1 Comparison of the results for the voltage magnitude at bus 84417 for three different types of expansions*

	MCS	GC	EW	CF
mean / pu	1.12406	1.12722	1.12722	1.12722
stand deviation / pu	0.02913	0.02488	0.02488	0.02488
ARMS	0	0.24%	0.21%	0.17%
10% confidence level / pu	1.08550	1.09135	1.09008	1.08992
90% confidence level / pu	1.15063	1.15179	1.15024	1.15069
OVP (>1.1 pu)	83.23%	85.36%	84.70%	84.03%
Time / s	2147.30	452.14	452.15	452.14

*OVP – steady state overvoltage probability; GC – Gram-Charlier; EW – Edgeworth; CF – Cornish-Fisher.

It can be observed from Fig. 6.2 and Table 6.1 that the Cornish-Fisher expansion provides a better curve fit for the CDF curve of the voltage magnitude at bus 84417, at least over a given range of $|V|$. Additionally, cumulant methods can save computational time compared with MCS.

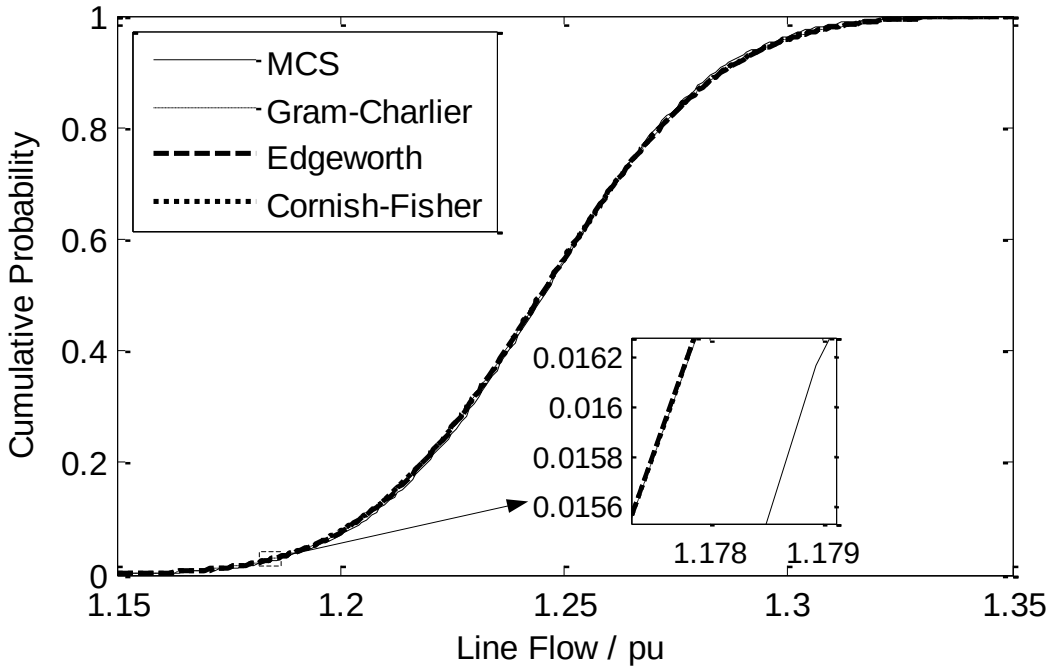

Fig. 6.3 CDF curves of the line flow through line 14356-17013 in three different types of expansions

Table 6.2 Comparison of the results for the line flow through line 14356-17013 for three different types of expansions*

	MCS	GC	EW	CF
mean / pu	1.24516	1.24500	1.24500	1.24500
stand deviation / pu	0.03192	0.03217	0.03217	0.03217
ARMS	0	0.01813%	0.01809%	0.01806%
10% confidence level / pu	1.20423	1.20408	1.20409	1.20409
90% confidence level / pu	1.28575	1.28641	1.28642	1.28641
OLP (>1.2 pu)	92.06%	92.07%	92.07%	92.07%

*OLP – overload probability.

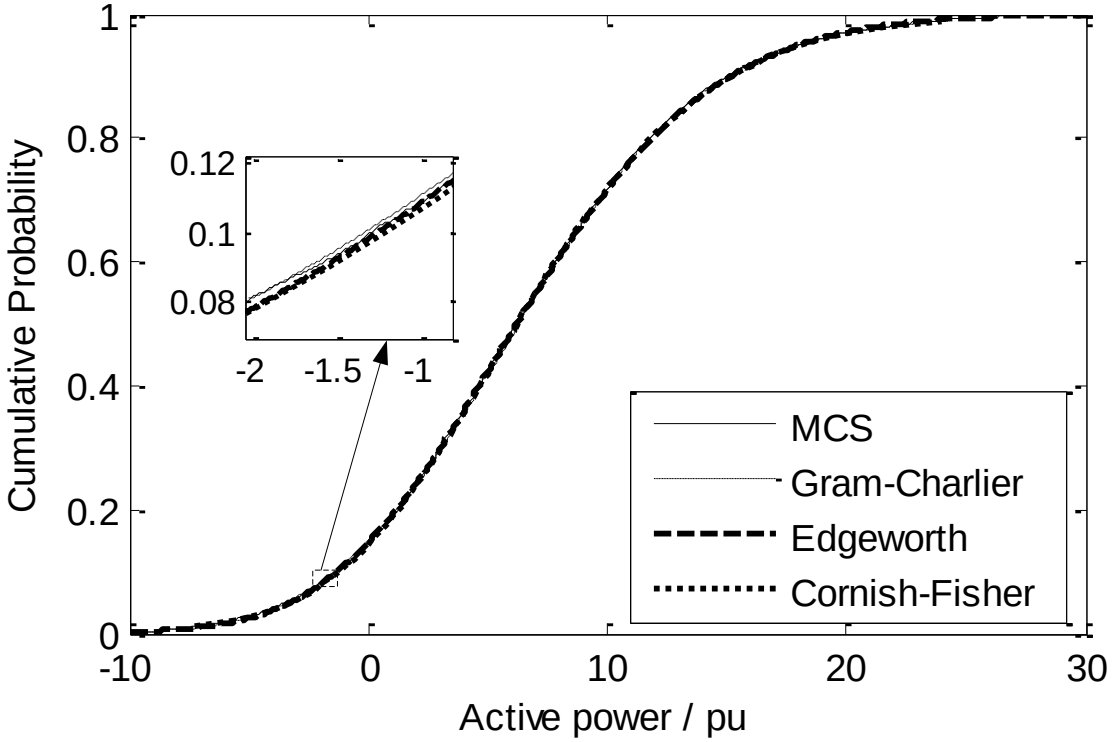

Fig. 6.4 CDF curves of the slack bus active power for three different types of expansions

Table 6.3 Comparison of the results for the slack bus active power for three different types of expansions

	MCS	GC	EW	CF
mean / pu	6.63984	6.66000	6.66000	6.66000
stand deviation / pu	6.51580	6.53959	6.53959	6.53959
ARMS	0	2.56E-05	2.38E-05	2.56E-05
10% confidence level / pu	-1.28500	-1.32951	-1.25194	-1.21708
90% confidence level / pu	15.08250	15.07947	15.22006	15.18475

The results for the slack bus active power production indicate that the variance of the slack bus production is large (the CV of the slack bus output is 98.13%) due to the assumption that all the power uncertainty is balanced by the slack bus generation alone.

According to the results of the line flow through line 14356-17013 and the slack bus active power, all the three expansions give good approximations because the indicated line flow probability distribution is qualitatively close to normally distributed.

For some non-normal distributions, the CDF curve cannot be fit well with some of the expansions. For example, the CDF curves of the voltage magnitude at bus 84511 are shown in Fig. 6.5 for the three exemplar expansions. It can be observed that the Cornish-Fisher and Edgeworth expansions perform worse at the probability close to 1, and the Gram-Charlier expansion gave a better approximation. Fig. 6.5 also indicates that the Cornish-Fisher and Edgeworth expansions are not always more accurate than the Gram-Charlier expansion, since they may have "unreliable tail behavior" in some cases. Depending on the ultimate application of the CDF curves, these issues may be problematic.

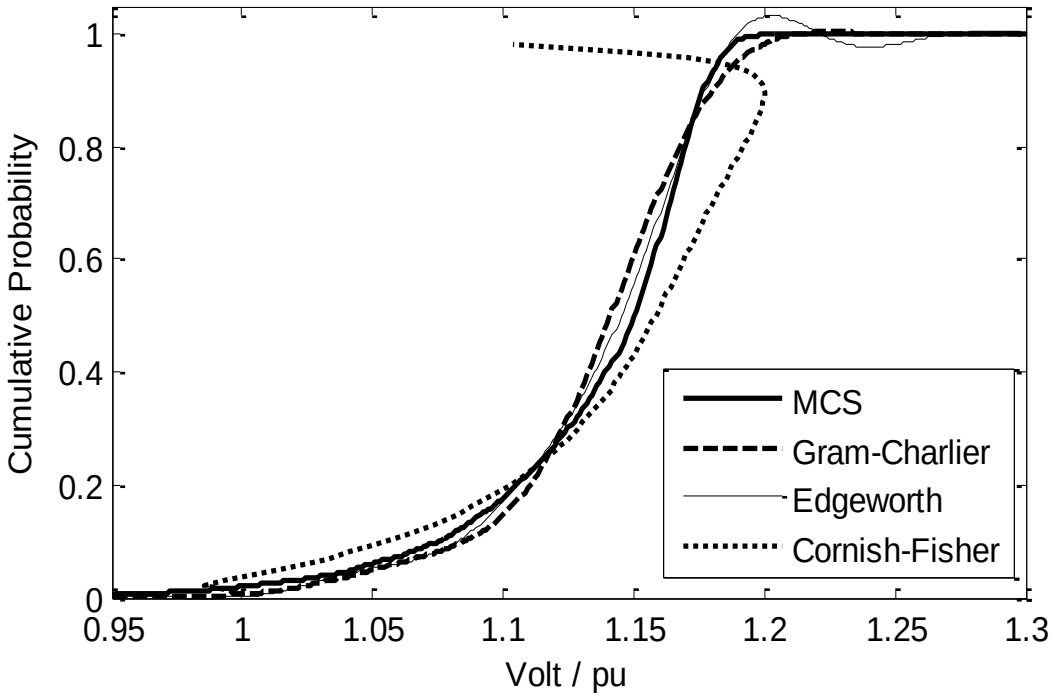

Fig. 6.5 CDF curves of the voltage magnitude at bus 84511 for three different types of expansions

6.3 PPF Results for Different PV Generation Correlations

As stated in the foregoing paragraphs, the neighboring PV resources may be highly correlated. To evaluate the effect of PV generation correlation on the transmission system uncertainty, the results for three different representative correlation levels are compared. In these simulations, the correlation coefficient ρ among PV resources is set to be 0.0, 0.5 and 1.0, and all the CDF curves are approximated by the Gram-Charlier expansion. If $\rho = 0.0$, all the PV resources are statistically independent. In contrast, all the PV resources are totally positively dependent in the case of $\rho = 1.0$. Since the neighboring PV generations are typically positively correlated, the negative correlation cases ($\rho < 0.0$) are not considered.

Fig. 6.6 shows the CDF curves of the steady state voltage magnitude at bus 84511 (see Fig. 6.1) for different PV generation correlation conditions. Table 6.4 compares the mean value, the standard deviation, the confidence level (10% and 90%), steady state overvoltage probability and computing time for each of the correlation cases. As a comparison, the CDF curves and simulation results of the steady state voltage magnitude at bus 84511 by using MCS are shown in Fig. 6.7 and Table 6.5, respectively.

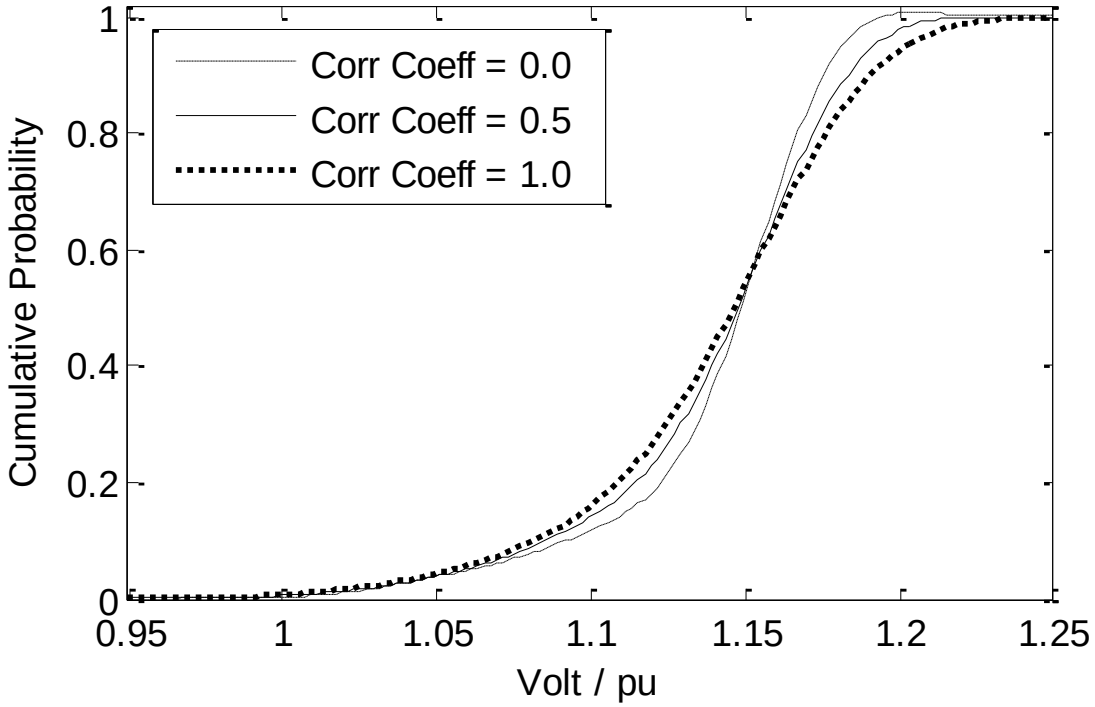

Fig. 6.6 CDF curves of the steady state voltage magnitude at bus 84511 for different PV generation correlations (Corr Coeff –correlation coefficient)

Table 6.4 Comparison of the results for the steady state voltage magnitude at bus 84511 for different PV generation correlations

	Corr Coeff = 0.0	Corr Coeff = 0.5	Corr Coeff = 1.0
mean / pu	1.14027	1.14027	1.14027
stand deviation / pu	0.03514	0.03998	0.04436
10% confidence level / pu	1.08726	1.08537	1.08136
90% confidence level / pu	1.17397	1.18301	1.19052
OVP (>1.1 pu)	86.49%	85.91%	83.99%
Time / s	135.69	452.14	453.34

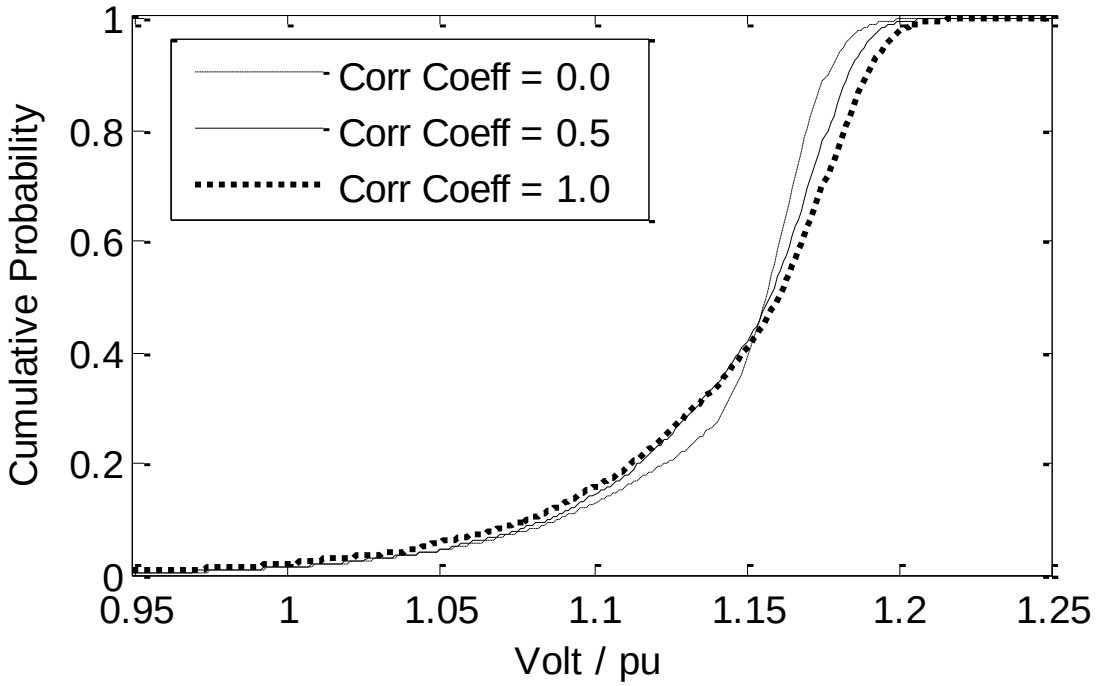

Fig. 6.7 CDF curves of the steady state voltage magnitude at bus 84511 by using MCS for different PV generation correlations

Table 6.5 Comparison of the MCS results for the steady state voltage magnitude at bus 84511 for different PV generation correlations

	Corr Coeff = 0.0	Corr Coeff = 0.5	Corr Coeff = 1.0
mean / pu	1.14318	1.14327	1.14430
stand deviation / pu	0.03611	0.04066	0.04580
10% confidence level / pu	1.08883	1.08561	1.08060
90% confidence level / pu	1.17588	1.18275	1.18799
OVP (>1.1 pu)	87.21%	85.65%	84.38%

Fig. 6.8 and Table 6.6 show the CDF curves and simulation results of the line flow through line 14356-17013 for different PV generation correlation conditions.

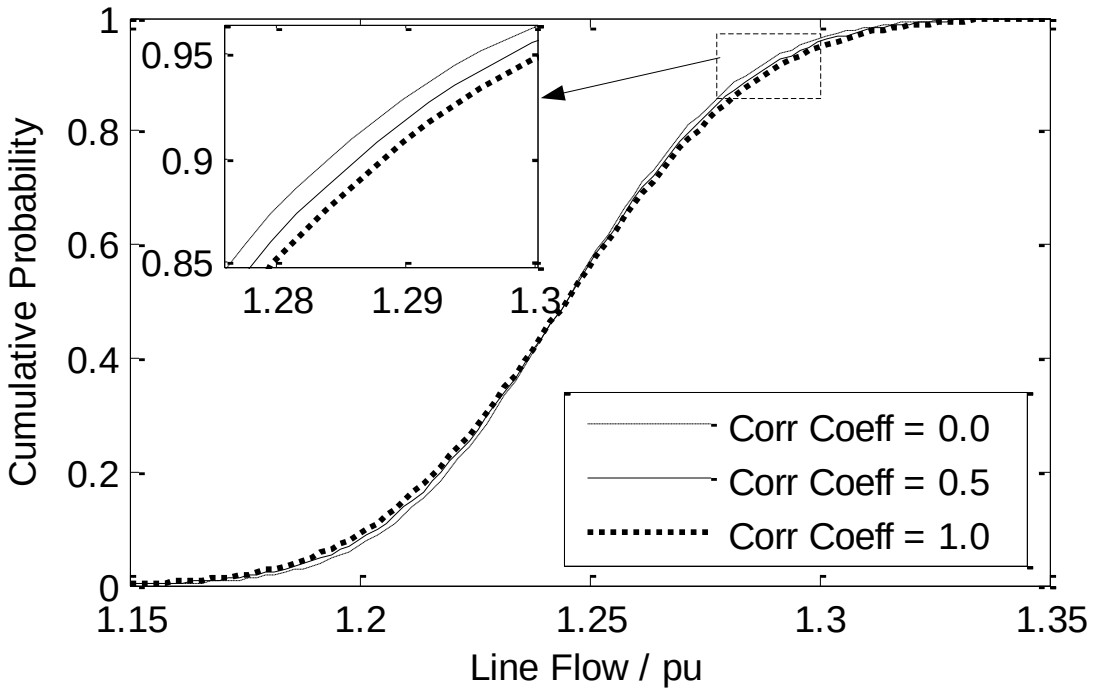

Fig. 6.8 CDF curves of the line flow through line 14356-17013 for different PV generation correlations

Table 6.6 Comparison of the results for the line flow through line 14356-17013 for different PV generation correlations

	Corr Coeff = 0.0	Corr Coeff = 0.5	Corr Coeff = 1.0
mean / pu	1.24500	1.24500	1.24500
stand deviation / pu	0.03050	0.03217	0.03376
10% confidence level / pu	1.20463	1.20408	1.20187
90% confidence level / pu	1.28556	1.28641	1.28830
OLP (>1.2 pu)	92.36%	92.07%	90.95%

Based on the results obtained by increasing the correlation coefficient among PV resources, it is observed that the mean values of both bus voltage magnitude and line flow are constant, but the standard deviations increase.

According to the results, the totally dependent condition has the maximum uncertainties of bus voltages and line flows. The minimum uncertainties of bus voltages and line flows occur in the totally independent condition. The uncertainties of the partially dependent condition are in the middle. The results indicate that the more positive correlation the PV generations have, the more uncertainties of bus voltages and line flows occur. The uncertainty problems become more serious since neighboring PV generation outputs change with a similar trend (i.e., positive correlation).

On the other hand, the comparison of the computation time shown in Table 6.4 indicates that the consideration of correlation among the PV resources increases the computational burden, because a large number of joint cumulants and joint moments need to be calculated.

6.4 PPF Results for Different PV Generation Penetrations

To evaluate the influence of PV generation uncertainty, different PV generation penetration conditions of 0%, 5%, 10%, 15% and 20% of total load are observed. The *correlation coefficient* among PV resources is fixed at 0.5. The Gram-Charlier expansion is again utilized to approximate the CDF curves. The CDF curves and simulation results of the steady state voltage magnitude at bus 84511 for different PV generation penetrations are shown in Fig. 6.9 and Table 6.7, respectively.

Fig. 6.9 and Table 6.7 show that the steady state voltage violation problem and uncertainty problem are more serious when PV generation penetration increases, as expected, since PV generation is not allowed to provide voltage control.

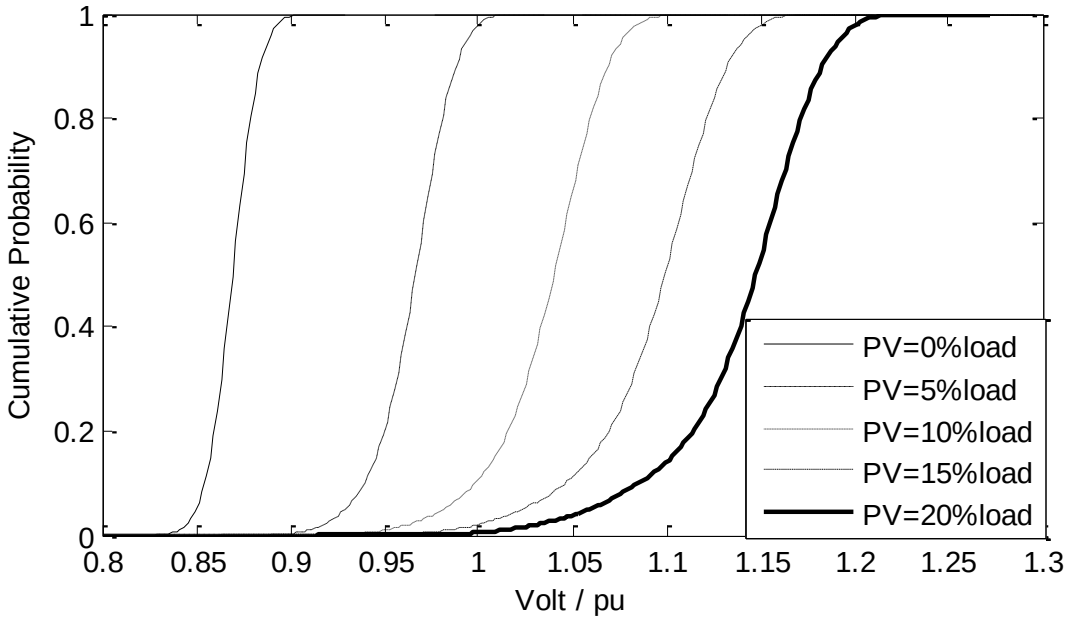

Fig. 6.9 CDF curves of the steady state voltage magnitude at bus 84511 for different PV generation penetrations

Table 6.7 Comparison of the results for the steady state voltage magnitude at bus 84511 for different PV generation penetrations

	PV generation penetration				
Index	0%	5%	10%	15%	20%
mean / pu	0.86940	0.96489	1.03640	1.09360	1.14027
stand deviation / pu	0.01169	0.01955	0.02871	0.03533	0.03998
10% confidence level / pu	0.85449	0.93929	0.99830	1.04600	1.08537
90% confidence level / pu	0.88432	0.98808	1.06899	1.13253	1.18301
OVP (>1.1 pu)	0.00%	5.2E-8%	0.25%	48.76%	85.91%

The CDF curves and the simulation results of the line flow through line 14350-19060 for different PV generation penetrations are displayed in Fig. 6.10 and Table 6.8. According to the results, the overload probability is relieved with the increase of PV

generation penetration, since PV generation usually supplies the load locally and reduces the line flows on most transmission lines.

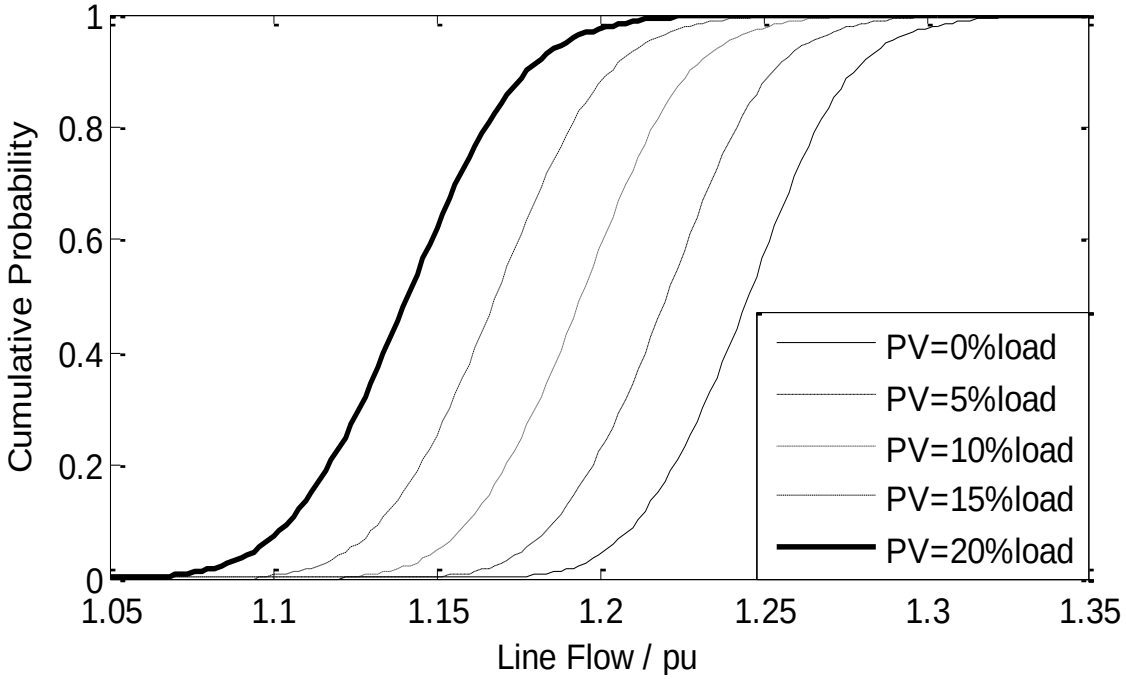

Fig. 6.10 CDF curves of the line flow through line 14350-19060 for different PV generation penetrations

Table 6.8 Comparison of the results for the line flow through line 14350-19060 for different PV generation penetrations

	PV generation penetration				
Index	0%	5%	10%	15%	20%
mean / pu	1.24500	1.22000	1.19400	1.16800	1.14100
stand deviation / pu	0.02661	0.02676	0.02712	0.02791	0.02898
10% confidence level / pu	1.20457	1.18558	1.15921	1.13233	1.10412
90% confidence level / pu	1.28564	1.25302	1.22769	1.20301	1.17766
OLP (>1.2 pu)	92.36%	77.22%	41.15%	11.96%	2.47%

6.5 PPF Results considering Generation Dispatching Operation

The proposed PPF algorithm considering generation dispatching operation is also tested in the Arizona area of the WECC transmission system.

This case study uses the economic dispatch (the optimal sensitivity equations) to obtain the *T* matrix. To examine the advantages of the algorithm, the PPF algorithm without considering the generation dispatch is also applied, and it assumes that the slack bus generation entirely balances the system uncertainty. To evaluate the result accuracy and efficiency, MCS for the two cases with 10000 samples (the coefficient of variation β is set as less than 1% for all the result variables) is applied as a reference. Considering that the neighboring PV generation productions are typically positively correlated, the correlation coefficient among PV resources is chosen to be +0.3 for testing the proposed algorithm.

Fig. 6.11-Fig. 6.14 show the CDF curves of the voltage phase angle and voltage magnitude at bus 84417 and the line flow active power and reactive power through line 86291-14006, respectively. Table 6.9-Table 6.12 list the ARMS error, mean value, standard deviation and the confidence level of 10% and 90%. These data are compared versus MCS results. Table 6.13 compares the computational time of different algorithms.

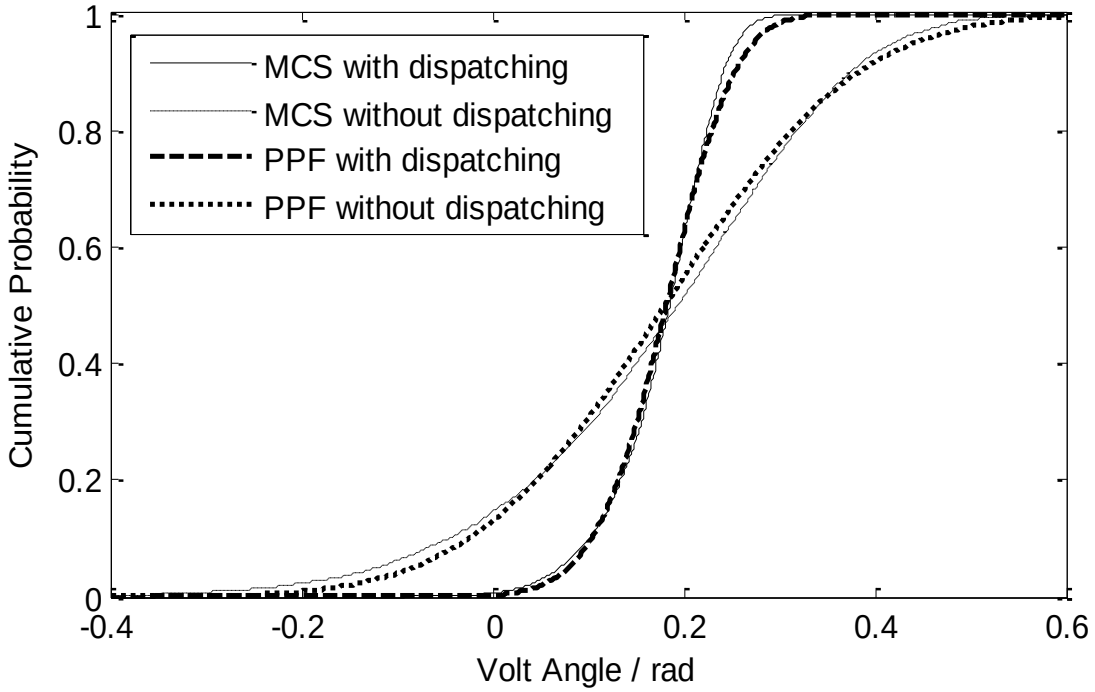

Fig. 6.11 CDF curves of the voltage angle at bus 84417 considering generation dispatching operation

Table 6.9 Comparison of the results for the voltage angle at bus 84417 considering generation dispatching operation

	With dispatching		Without dispatching	
Index	MCS	PPF	MCS	PPF
ARMS / %	0	0.03854%	0	0.05321%
mean / radians	0.17742	0.17977	0.17720	0.17977
stand deviation / radians	0.05618	0.05872	0.16605	0.15893
10% confidence level / radians	0.10043	0.10361	-0.04514	-0.02415
90% confidence level / radians	0.24108	0.25326	0.37469	0.38313

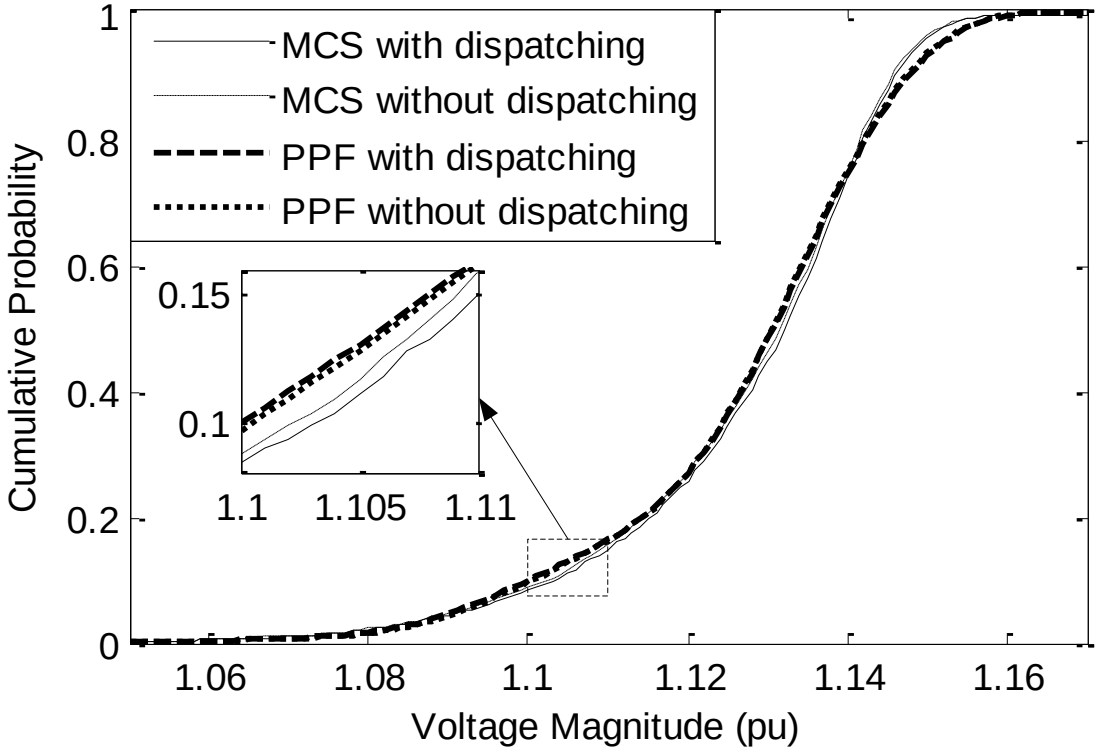

Fig. 6.12 CDF curves of the voltage magnitude at bus 84417 considering generation dispatching operation

Table 6.10 Comparison of the results for the voltage magnitude at bus 84417 considering generation dispatching operation

	With dispatching		Without dispatching	
Index	MCS	PPF	MCS	PPF
ARMS / %	0	0.04960%	0	0.03692%
mean / pu	1.12773	1.12722	1.1271	1.12722
stand deviation / pu	0.01794	0.01837	0.0181	0.01805
10% confidence level / pu	1.10328	1.10003	1.10237	1.10055
90% confidence level / pu	1.14579	1.1475	1.14547	1.14724

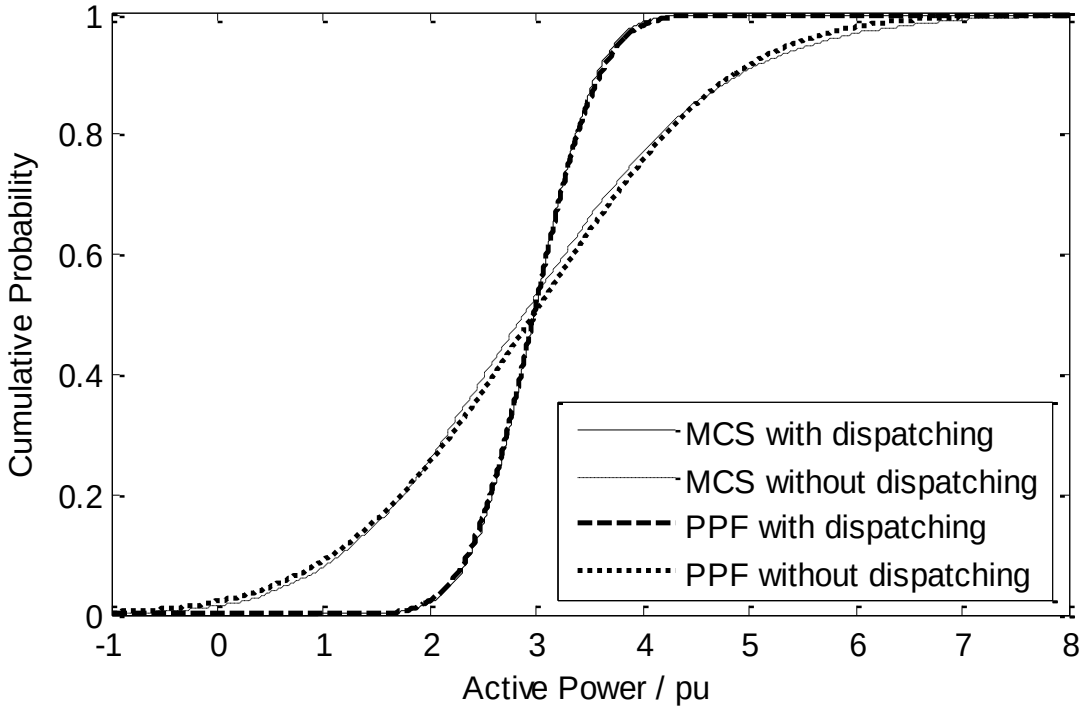

Fig. 6.13 CDF curves of the line flow active power through line 86291-14006 considering generation dispatching operation

Table 6.11 Comparison of the results for the line flow active power through line 86291-14006 considering generation dispatching operation

	With dispatching		Without dispatching	
Index	MCS	PPF	MCS	PPF
ARMS / %	0	0.0035%	0	0.0095%
mean / pu	2.97683	2.97500	2.97683	2.97500
stand deviation / pu	0.46272	0.48561	1.48475	1.47075
10% confidence level / pu	2.37775	2.35294	1.17150	1.09162
90% confidence level / pu	3.55900	3.59756	4.91375	4.86092

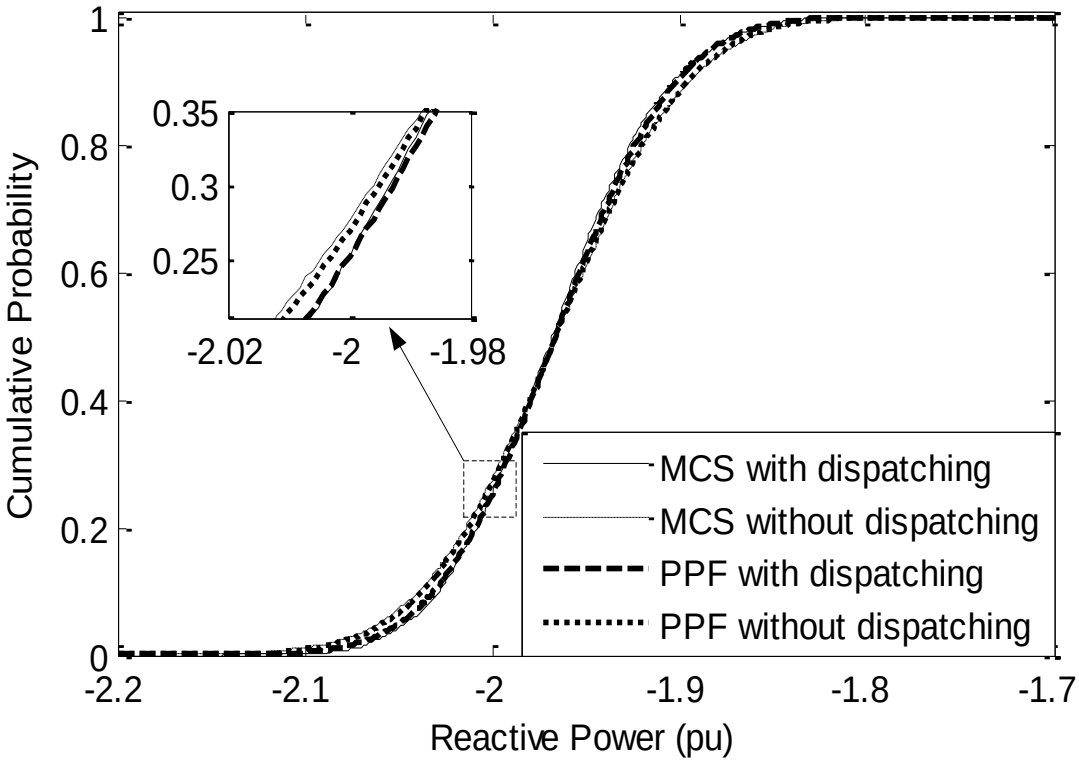

Fig. 6.14 CDF curves of the line flow reactive power through line 86291-14006 considering generation dispatching operation

Table 6.12 Comparison of the results for the line flow reactive power through line 86291-14006 considering generation dispatching operation

	With dispatching		Without dispatching	
Index	MCS	PPF	MCS	PPF
ARMS / %	0	0.01450%	0	0.00694%
mean / pu	-1.96665	-1.95600	-1.96629	-1.95600
stand deviation / pu	0.04902	0.05067	0.05686	0.05546
10% confidence level / pu	-2.02840	-2.03089	-2.03889	-2.03719
90% confidence level / pu	-1.90307	-1.90104	-1.89388	-1.89511

Table 6.13 Comparison of the computational time in different algorithms

Algorithm	computational time / s
MCS with dispatching model	2303.12
MCS without dispatching model	2270.10
PPF with dispatching model	476.43
PPF without dispatching model	458.92

Compared to the MCS result, the proposed PPF method can provide accurate approximate CDF curves requiring much less computational time. According to the results, the generation dispatching model has a significant influence on the uncertainties of bus voltage angles and line flow active powers but little effect on the uncertainties of bus voltage magnitudes and line flow reactive powers. The reason is that the generation dispatch only deals with the active power variation of PV generation, which has a strong coupling with the bus voltage angles and line flow active powers.

On the other hand, the results also indicate that the generation dispatch decreases the uncertainties of both bus voltage angles and line flow active powers in most cases. This occurs because the change of conventional generation always balances the variation of PV generation based on the generation dispatching law. However, the uncertainties increase in some cases, where the buses and branches are generally close to the conventional generators. An example is shown in Fig. 6.15 and Table 6.14. Line 19034-19080 is near a generator. The result indicates that the uncertainty in line flow active power increases when the generator dispatch is considered.

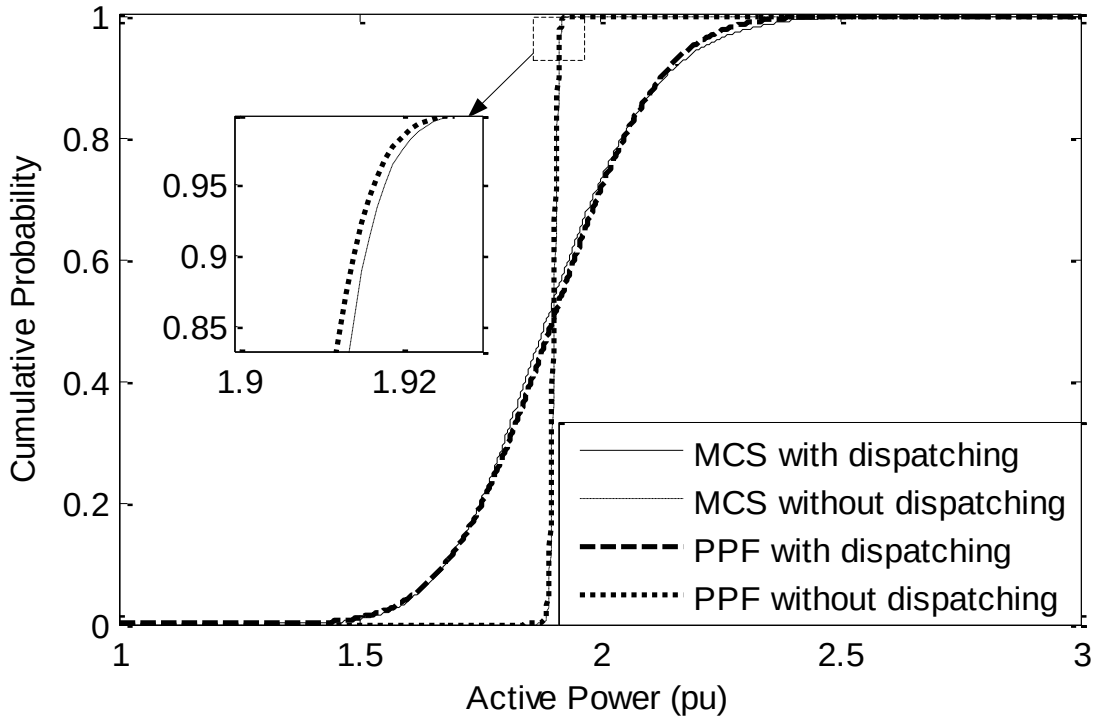

Fig. 6.15 CDF curves of the line flow through line 19034-19080 considering generation dispatching operation

Table 6.14 Comparison of the results for the line flow through line 19034-19080 considering generation dispatching operation

	With dispatching		Without dispatching	
Index	MCS	PPF	MCS	PPF
ARMS / %	0	0.04332%	0	0.03114%
mean / pu	1.90253	1.90500	1.90652	1.90500
stand deviation / pu	0.18075	0.17556	0.00705	0.00702
10% confidence level / pu	1.68267	1.68101	1.89738	1.89600
90% confidence level / pu	2.13831	2.13053	1.91546	1.91401

Slack bus is used to balance all the power in the system, and the uncertainty of the slack bus influences the system operation and generation cost. The CDF curves and uncertainty results of the slack bus active power are given in Fig. 6.16 and Table 6.15.

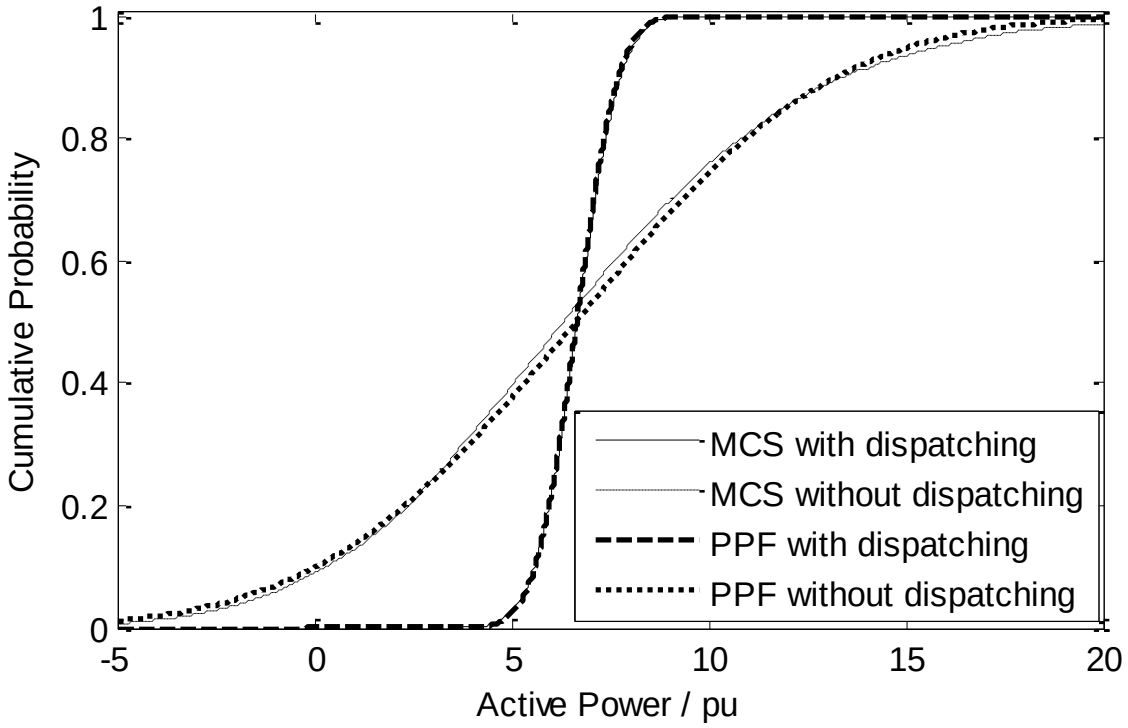

Fig. 6.16 CDF curves of the slack bus active power considering generation dispatching operation

Table 6.15 Comparison of the results for the slack bus active power considering generation dispatching operation

	With dispatching		Without dispatching	
Index	MCS	PPF	MCS	PPF
ARMS / %	0	0.0223%	0	0.0827%
mean / pu	6.68253	6.66000	6.67305	6.66000
stand deviation / pu	0.86451	0.82700	5.21365	5.15182
10% confidence level / pu	5.58115	5.60103	0.33500	0.06406
90% confidence level / pu	7.77826	7.72040	13.42900	13.26683

According to the results shown in Fig. 6.16 and Table 6.15, the standard deviation of the slack bus active power output with the generation dispatching model is 0.82700 pu; this generation level is much smaller than the power level without the generation dispatching model, namely 5.15182 pu. The slack bus output uncertainty is reduced since other conventional generators share in balancing the PV generation variation.

The results also illustrate that the generation dispatching behavior plays an important role in determining the impact of the system uncertainties and cannot be ignored.

6.6 PPF Results for the Probabilistic Optimal Power Dispatching Strategy

The proposed probabilistic optimal power dispatching strategy is also applied to the test system. As stated in Section 5.3, the objective is to minimize the expected value of the total cost of generation production under the overload probability limit. The control variable T matrix is obtained based on the proposed dispatching strategy. The proposed PPF algorithm is used as a test tool to evaluate the validity of the proposed strategy. In this case study, the overload probability limit ξ is set to be 5%. As a comparison, the case of economic dispatch (not considering the overload probability constraint) and the case without considering the generation dispatch (slack bus generator balances the variations of all the PV generations) are also applied.

Table 6.16 shows the total cost of various generation dispatching strategies. The result indicates that the case without dispatching costs most, since the slack bus generation compensates the system uncertainty alone. The proposed dispatching strategy costs a little more than the economic dispatch, since the overload probability constraint is considered.

The CDF curves and results of the line flow in line 86291-14006 are shown in Fig. 6.17 and Table 6.17. According to the PPF results, the proposed dispatching strategy performs better to limit the overload probability. Economic dispatch reduces operating cost but it cannot be utilized to control line overload problems.

Table 6.16 Comparison of the cost of different generation dispatching strategies

	Proposed dispatch	Economic dispatch	Without dispatching
Expected value of total cost ($)	78960.4	78949.3	79877.4

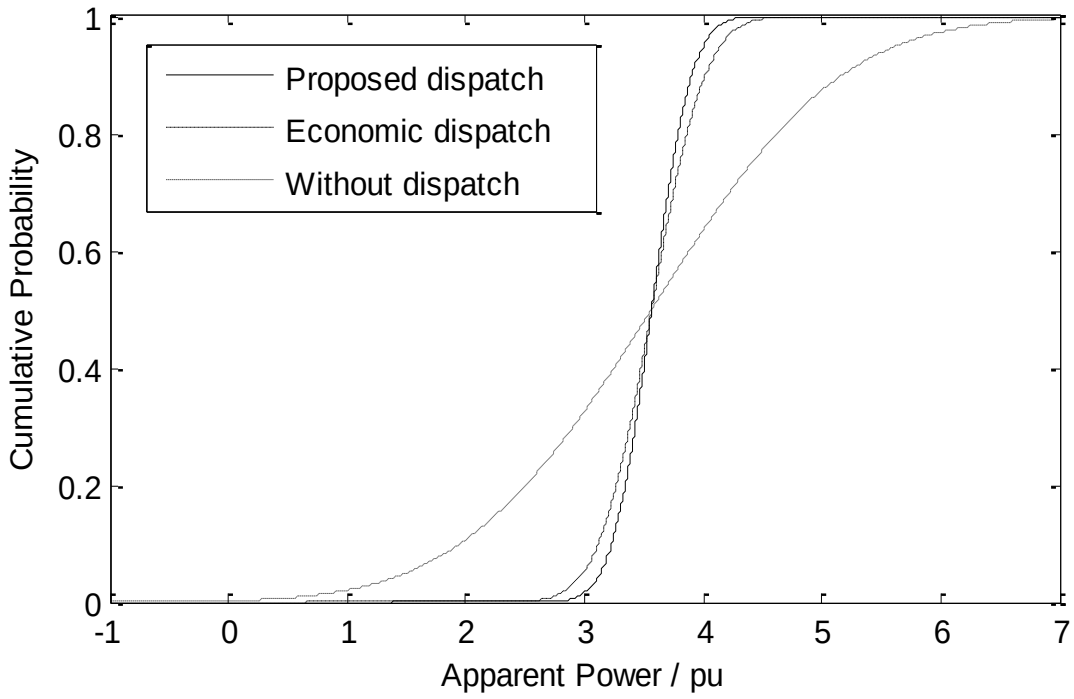

Fig. 6.17 CDF curves of the line flow through line 86291-14006 for different generation dispatching strategies

Table 6.17 Comparison of the results for the line flow through line 86291-14006 for different generation dispatching strategies

	Proposed dispatch	Economic dispatch	Without dispatching
OLP*	0.04995	0.10878	0.36419
mean / pu	3.56600	3.56600	3.56600
stand deviation / pu	0.26347	0.35185	1.25443
10% confidence level / pu	3.22857	3.11530	1.95968
90% confidence level / pu	3.90382	4.01710	5.17456

*The thermal limit of the line flow is 4.0 pu.

Chapter 7

CONCLUSIONS AND FUTURE WORK

7.1 Main Conclusions

In this dissertation, the PPF algorithm is used to evaluate the probabilistic characteristics of transmission systems with PV generation installed. Compared with the deterministic power flow method, the PPF algorithm has the ability to assess the system in a broader sense, including system reliability and performance. The cumulant method is applied to compute the CDF instead of using convolution calculations. The cumulant method is potentially suitable for large transmission systems.

This dissertation compares three types of approximation expansions based on cumulants, Gram-Charlier expansion, Edgeworth expansion, and Cornish-Fisher expansion. It can be concluded that all the three expansions are accurate methods when the objective distribution is close to a normal distribution. In many practical cases, the Cornish-Fisher and Edgeworth expansions can result in accurate models, but these two expansions may have the problems of bad tail behavior. The approximation used by cumulant methods must consider accuracy in the tail region since extreme behavior (e.g., far from the mean) is often of the greatest engineering interest. Because no established, straightforward approach exists for determining the type of expansion to use and the number of terms to use, care must be taken to examine the expansion on a use by case basis.

In this dissertation, the correlation among the input random variables is considered by using joint cumulants. This approach increases the computational burden.

Since different locations have their own geographical and environmental conditions, the probability distributions of PV generation are also various. The change of solar position has very limit influence on PV production variation. Clearness and brightness of the sky can have a significant positive impact on PV generation production. The varied meteorological conditions make a dominant impact on the PV generation output.

A novel probabilistic model of PV generation is developed based on the environmental conditions that impact PV behavior. Compared to the actual data of PV generation, the novel probabilistic model gives an accurate approximation.

In this dissertation, the proposed PPF algorithm based on the cumulant method takes into account conventional generation dispatch. The dispatch of conventional resources of high ramp rate compensates the variations of PV generation resources. The results show that the consideration of generation dispatch significantly influences the PPF results and should not be neglected in the PPF algorithm. The probabilistic optimal power dispatching strategy is proposed to consider the uncertainty problem in the generation cost optimization.

The proposed method is applied to the Arizona area of the WECC system for different levels of PV generation penetration. To evaluate the accuracy of the results, MCS with 10,000 samples is utilized as a comparative reference. Based on the probabilistic results and analysis, the following conclusions can be drawn:

1. The PPF algorithm offers an efficient method to calculate the impact of PV generation in terms of steady state bus voltages and line flows.

2. The positive correlation among the PV resources makes the uncertainty problem more serious. In the case study, the standard deviation of the steady state voltage

magnitude increases from 0.03514 pu to 0.04436 pu when the correlation coefficient among PV generations increases from 0.0 to 1.0.

3. PV generation uncertainty has an influence on the system uncertainty when PV generation penetration is high.

4. The steady state voltage magnitude violation problems are more serious in the cases with increased penetration of PV generation. For the system considered, steady state overvoltages do not occur with no PV installed, but at 20% PV penetration, overvoltages are expected 86.49% of the time.

5. Since most PV generation is close to loads, the power flow bottlenecks are relieved in some transmission lines. In the test system, an example considered shows that the overload probability reduces from 92.36% to 3.23%, when the PV penetration increases from 0% to 20%.

6. Although PV generation uncertainty has a strong impact on the system uncertainty problem, the generation dispatching behaviors reduce the uncertainty influence of PV generation.

7. The consideration of generation dispatch decreases the variances of bus voltage angles and line flow active powers in most cases, but has little effect on the bus voltage magnitudes and line flow reactive powers. The reason is that the generation dispatching operation only balances the active power of PV generation.

8. The consideration of generation dispatching operation decreases the variance of the slack bus output, since other conventional generations share the work of balancing the PV generation uncertainties.

9. Probabilistic optimal power dispatch can optimize the expected total generation operating cost constrained by line overload probability.

7.2 Future Work

Based on the conclusions obtained from the dissertation, the research work in this project could continue along the following paths.

- The PPF algorithm could also consider the systems with other renewable energy such as wind generation to evaluate the uncertainty impact.
- The impact of the uncertainty of network topology on the power system should be further considered in the PPF algorithm.
- The probabilistic analysis for contingencies in transmission systems with PV generation will be the next step of research.

REFERENCES

[1]. Z. Wang, and L. Alvarado, "Interval Arithmetic in Power Flow Analysis," *Proceedings of Power Industry Computer Application Conference*, pp. 156 – 162, 1991.

[2]. S. Rahman, and M. Bouzguenda, "A Model to Determine the Degree of Penetration and Energy Cost of Large Scale Utility Interactive Photovoltaic Systems," *Power Engineering Review, IEEE*, v. 14, No. 6, 1994.

[3]. K. H. Hussein, I. Muta, T. Hoshino, and M. Osakada, "Maximum Photovoltaic Power Tracking: an Algorithm for Rapidly Changing Atmospheric Conditions," *Proceedings of Generation, Transmission and Distribution Conference*, v. 142, No. 1, pp. 59 – 64, 1995.

[4]. J. Monedero, F. Dobon, A. Lugo, P. Valera, R. Osuna, L. Acosta, and G. Marichal, "Minimizing Energy Shadow Losses for Large PV Plants," *Photovoltaic Energy Conversion*, v. 2, pp. 2043 – 2045, 2003.

[5]. J. F. Dopazo, O. A. Klitin, and A. M. Sasson, "Stochastic Load Flows," *IEEE Transactions on Power Apparatus and Systems*, v. 94, No. 2, Match/April 1975, pp. 299 – 309.

[6]. B. Borkowska, "Probabilistic Load Flow," *IEEE Transactions on Power Apparatus and Systems*, v. 93, No. 3, pp. 752–759, 1974.

[7]. R. N. Allan, A. M. Leite da Silva, and R. C. Burchett, "Evaluation Methods and Accuracy in Probabilistic Load Flow Solutions," *IEEE Transactions on Power Apparatus and Systems*, v. 100, No. 5, pp. 2539 – 2546, July 2003.

[8]. Andrew S. Golder, "Photovoltaic Generator Modeling for Large Scale Distribution System Studies," Master of Science dissertation, Drexel University, Philadelphia, PA, 2006.

[9]. R. N. Allan, B. Borkowska, and C. H. Grigg, "Probabilistic Analysis of Power Flows," *Proceedings of the Institution of Electrical Engineers*, v. 121, No. 2, pp. 1551–1556, 1974.

[10]. R. N. Allan and M. R. G. Al-Shakarchi, "Probabilistic A. C. load flow," *Proceedings of the Institution of Electrical Engineers*, v. 123, No. 6, pp. 531–536, 1976.

[11]. R. N. Allan and M. R. G. Al-Shakarchi, "Probabilistic techniques in AC load flow analysis," *Proceedings of the Institution of Electrical Engineers*, v. 124, No. 2, pp. 154–160, 1976.

[12]. R. N. Allan, C. H. Grigg, and M. R. G. Al-Shakarchi, "Numerical Techniques in Probabilistic Load Flow Problems," *Int. J. Numerical Methods* Eng. 10, pp. 853–860, 1976.

[13]. G. T. Heydt and P. W. Sauer, “Stochastic Power Flow Study Methods,” presented at *the Allerton Conference on Circuits and Systems*, Champaign, IL, Oct. 1976.

[14]. M. Flam and A. Sasson, “Stochastic Load Flow Decoupled Implementation,” presented at *the IEEE Summer Meeting*, A77 515-0, Mexico City, Mexico, July 1977.

[15]. P. Sauer, “A Generalized Stochastic Power Flow Algorithm,” Ph. D. dissertation, Purdue University, W. LaFayette, IN, 1977.

[16]. A.P. Meliopoulos, A.G. Bakirtzis, R. Kovacs, “Power System Reliability Evaluation Using Stochastic Load Flows,” *IEEE Trans. on Power Apparatus and Systems*, v. PAS-103, No. 5, pp. 1084-1091, May 1984.

[17]. G. K. Stefopoulos, A. P. Meliopoulos and G. J. Cokkinides, “Probabilistic Power Flow with Non-Conforming Electric Loads,” *Proc. of the 8th International Conference on Probabilistic Methods Applied to Power Systems*, Iowa State University, Ames, Iowa, September 12-16, pp. 525 – 531, 2004.

[18]. A.M. Leite da Silva, R.A.G. Fernández, and C. Singh, “Generating Capacity Reliability Evaluation Based on Monte Carlo Simulation and Cross-Entropy Methods,” *IEEE Trans. on Power Systems*, vol. 25, no. 1, pp. 129-137, Feb. 2010.

[19]. Pei Zhang, Stephen T. Lee, “Probabilistic Load Flow Computation Using the Method of Combined Cumulants and Gram-Charlier Expansion,” *IEEE Trans. on power systems*, v. 19, No. 1, pp. 676-682, February 2004.

[20]. J. Usaola, "Probabilistic Load Flow in Systems with Wind Generation," *IET Generation, Transmission and Distribution*, v. 3, No. 12, pp. 1031-1041, 2009.

[21]. S. Blinnikov, R. Moessner, "Expansions for Nearly Gaussian Distributions," *Astron. Astrophys. Suppl. Ser.* 130, pp. 193-205, 1998.

[22]. H. Cramer, "Mathematical Methods of Statistics," Princeton: Princeton University Press, 1946.

[23]. A. Stuart, J. K. Ord, "Kendall's Advanced Theory of Statistics," 6th ed., vol. 1, E. Arnold, Ed. New York: Halsted Press, 1994, pp. 236-239.

[24]. S. R. Jaschke, "The Cornish-Fisher-expansion in the Context of Delta-Gamma-normal Approximations". Discussion Paper 54, Sonderforschungsbereich 373, Humboldt-Universitaet zu, Berlin, 2001. Available: http://www.jaschke-net.de/papers/CoFi.pdf.

[25]. R. N. Allan and M. R. G. Al-Shakarchi, “Linear Dependence Between Nodal Powers in Probabilistic AC Load Flow,” *Proc. IEE*, v. 124, pp. 529-534, 1977.

[26]. P.W. Sauer, G.T. Heydt, "A Generalized Stochastic Power Flow Algorithm," Paper A78-544-9, presented *at the 1978 IEEE/PES Summer Meeting*, July 1978.

[27]. A.M. Leite da Silva, V.L. Arienti, R.N. Allan, "Probabilistic Load Flow Considering Dependence between Input Nodal Powers," *IEEE Trans. on Power Apparatus and Systems*, v. PAS-103, No. 6, pp. 1524-1530, June 1984.

[28]. A.P.S. Meliopoulos, G.J. Cokkinides, X. Y. Chao, "A New Probabilistic Power Flow Analysis Method," *IEEE Transactions on Power Systems*, v. 5, No. 1, pp. 182–190, 1990.

[29]. R. N. Allan, C. H. Grigg, D. A. Newey and R. F. Simmons, "Probabilistic Power Flow Techniques Extended and Applied to Operational Decision Making," *Proceedings of the Institution of Electrical Engineers*, v. 123, No. 12, pp. 1317-1324, 1976.

[30]. McCullagh P. "Tensor Methods in Statistics." Chapman & Hall, London, 1987.

[31]. Chengshan Wang, Haifeng Zheng, Yinghua Xie, Kai Chen, "Probabilistic Power Flow containing Distributed Generation in Distribution System," *Automation of Electric power systems*, v. 29, No. 24, pp. 15-19, 2005.

[32]. S. H. Karaki, R. B. Chedid, R. Ramadan, "Probabilistic Performance Assessment of Autonomous Solar-wind Energy Conversion Systems," *IEEE Transactions on Energy Conversion*, v. 14, No. 3, pp. 766 – 772, 1999.

[33]. I. Abouzahr, R. Ramakumar, "Loss of Power Supply Probability of Stand-alone Photovoltaic Aystems: a Closed Form Solution Approach," *IEEE Transactions on Energy Conversion*, v. 6, No. 1, pp. 1 – 11, 1991.

[34]. D. L. King, W. E. Boyson, J. A. Kratochvil, "Photovoltaic Array Performance Model," Sandia National Laboratories, Albuquerque, New Mexico, 2004.

[35]. T. Van Cutsem, "An Approach to Corrective Control of Voltage Instability using Simulation and Sensitivity," *IEEE Transactions on Power Systems*, Vol. 10, No. 2, pp. 616-622, May 1995.

[36]. X. Wang, G. C. Ejebe, J. Tong, J. G. Waight, "Preventive/ Corrective Control for Voltage Stability Using Direct Interior Point Method," *IEEE Transactions on Power Systems*, v. 13, No. 3, pp. 878-883, August 1998.

[37]. Z. Feng, V. Ajjarapu, D. J. Maratukulam, "A Comprehensive Approach for Preventive and Corrective Control to Mitigate Voltage Collapse," *IEEE Transactions on Power Systems*, v. 15, No. 2, pp. 791-797, May 2000.

[38]. F. Capitanescu, T. V. Cutsem, "Preventive Control of Voltage Security Margins: A Multicontingency Sensitivity-Based Approach," *IEEE Transactions on Power Systems*, v. 17, No. 2, pp. 358-364, May 2002.

[39]. C. A. Canizares, "Calculating Optimal System Parameters to Maximize the Distance to Saddle-Node Bifurcations," *IEEE Transactions on Circuits and Systems-I*, v. 45, No. 3, pp. 225 - 237, March 1998.

[40]. R. Wang, R. H. Lasseter, "Redispatching Generation to Increase Power System Security Margin and Support Low Voltage Bus," *IEEE Transactions on Power Systems*, v. 15, No. 2, pp. 496 - 501, May 2000.

[41]. W. Rosehart, C. Canizares, V. Quintana, "Optimal Power Flow Incorporating Voltage Collapse Constraints," *Proceedings of the 1999 IEEE-PES Summer Meeting*, Edmonton, Alberta, pp. 820- 825, July 1999.

[42]. C. Y. Chung, L. Wang, F. Howell, P. Kundur, "Generation Rescheduling Methods to Improve Power Transfer Capability Constrained by Small Signal Stability," *IEEE Transactions on Power Systems*, v. 19, No. 1, pp. 524-530, February 2004.

[43]. G. J. Anders, "Probabilistic Concepts in Electric Power System," New York: Wiley, pp. 455–507, 1990.

[44]. Papoulis A., Pillai S.U. "Probability, Random Variables and Stochastic Processes," 4th Ed. McGraw-Hill, Boston 2002.

[45]. P. Sauer, G. Heydt, "A Convenient Multivariate Gram-Charlier Type A Series." *IEEE Transactions on Communications*, v. 27, No. 1, pp. 247-248, January 1979.

[46]. O. A. Oke, D. W. P. Thomas, G. M. Asher, L. R. A. X. de Menezes, "Probabilistic Load Flow for Distribution Systems with Wind Production Using Unscented Transform Method," *IEEE PES Innovative Smart Grid Technologies (ISGT)*, pp. 1-7, 2011.

[47]. Lei Dong, Weidong Cheng, Hai Bao, Yihan Yang, "Probabilistic Load Flow Analysis for Power System Containing Wind Farms," *Power and Energy Engineering Conference (APPEEC)*, pp. 1-4, 2010.

[48]. Anssi Seppala, "Load Research and Load Estimation in Electricity Distribution," Doctor of Technology dissertation, Helsinki University of Technology, Espoo, Finland, 1996.

[49]. S.W. Heunis, R. Herman, "A Probabilistic Model for Residential Consumer Loads," *IEEE Transactions on Power Systems*, v.17, No. 3, pp. 621-625, 2002.

[50]. Rafal Weron, B. Kozlowska, J. Nowicka-Zagrajek, "Modeling electricity loads in California: ARMA models with hyperbolic noise," *Proceedings of the NATO ARW on Application of Physics in Economic Modelling*, Prague, Feb. 8-10, 2001.

[51]. Z. Hu, Xifan Wang, "A Probabilistic Load Flow Method Considering Branch Outages," *IEEE Transactions on Power Systems*, v. 21, No. 2, pp. 507-514, 2006.

[52]. A. Schellenberg, W. Rosehart, J. Aguado, "Cumulant based stochastic optimal power flow (S-OPF) for variance optimization," *IEEE Power Engineering Society General Meeting*, San Francisco, CA, 12-16 June, 2005.

[53]. B. Sapkota, "Voltage stability assessment and enhancement of a large power system using static and dynamic approaches," Ph. D. dissertation, Arizona State University, Tempe AZ, 2009.

[54]. ERCOT, "Hourly Load Data Archives," available at: http://www.ercot.com/gridinfo/load/load_hist.

[55]. G. L. Viviani,; G. T. Heydt, "Stochastic Optimal Energy Dispatch," *IEEE Trans. on Power Apparatus and Systems*, v. PAS-100, No. 7, pp. 3221-3228, June 1981.

[56]. C. Kandilli, and K. Ulgen, "Solar Illumination and Estimating Daylight Availability of Global Solar Irradiance," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, v. 30, No. 12, 2008.

APPENDIX A

CAUSE OF THE VOLTAGE VIOLATION PROBLEM

This section discusses the reason of the voltage violation problem. As shown in Fig. A.1, in the two-bus system, the generation is at a PV bus, and the load is at a PQ bus. If the current in the system is predominantly inductive or capacitive, the voltage drop *jIX* primarily changes the voltage magnitude of U_2 (as shown in Fig. A.2 and Fig. A.3). On the other hand, if the current in the system is more in phase with the voltage, the voltage drop *jIX* primarily changes the voltage angle of U_2, and the voltage magnitude is nearly unchanged (shown in Fig. A.4). Therefore, the overvoltage problem occurs in the case where the leading reactive power is in the dominant position and there is a huge amount of reactive power in the system (shown in Fig. A.3). Thus, the voltage violation can be relieved by decreasing the system reactive power.

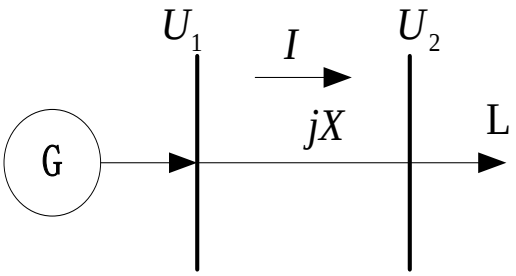

Fig. A.1 Two-bus system

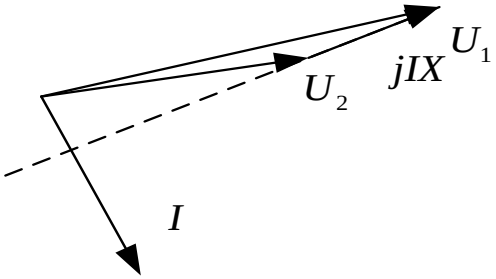

Fig. A.2 Predominantly inductive current

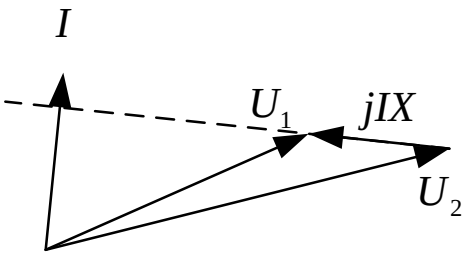

Fig. A.3 Predominantly capacitive current

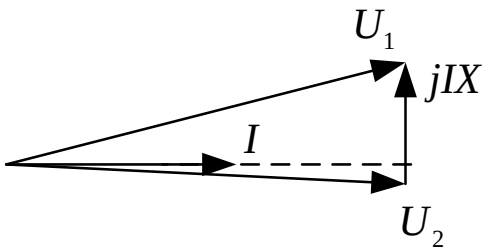

Fig. A.4 Current more in phase with the voltage

Since the voltage violation occurs when the PV generation penetration is increasing, this research focuses on the high PV generation penetration situation and determines approaches to reduce the bus voltage magnitude and sustain the voltage within an acceptable limit.

APPENDIX B

INTRODUCTION OF CDF AND PDF

In this section, cumulative distribution function (CDF) and probability density function (PDF) are introduces as follows [22][44].

The cumulative distribution function (CDF) $F(x)$ of the random variable X is at a value less than or equal to x and is defined as,

$$F(x) = P(X \le x) \tag{B-1}$$

The general properties of the CDF are listed as follows,

- F(x) is a non-decreasing function;
- $0 \le F(x) \le 1$, $F(-\infty)=0$, $F(+\infty)=1$;
- $P(x > x_0) = 1 - F(x_0)$;
- $P(a < x \le b) = F(b) - F(a)$;
- $P(x = x_0) = F(x_0) - F(x_0^-)$.

For a continuous random variable x, the probability density function (PDF) $f(x)$ is defined as,

$$\begin{aligned} f(x) &= \frac{dF(x)}{dx} \\ F(x) &= \int_{-\infty}^{x} f(u)du \end{aligned} \tag{B-2}$$

The PDF describes the relative likelihood of this random variable x to take on a given value.

APPENDIX C

PDF OF PRODUCT

If $\eta = \xi_1 \xi_2$, and (ξ_1, ξ_2) joint PDF is $f(x_1, x_2)$.

$$\begin{aligned} F(x) &= P\{\eta < x\} = P\{\xi_1 \xi_2 < x\} = \iint_{x_1 x_2 < x} f(x_1, x_2) dx_1 dx_2 \\ &= \int_0^\infty \left[\int_{-\infty}^{x/y} f(y, z) dz \right] dy + \int_{-\infty}^0 \left[\int_{x/y}^\infty f(y, z) dz \right] dy \end{aligned} \tag{C-1}$$

Thus, the PDF of *x* is as follows,

$$\begin{aligned} f(x) &= \int_0^\infty f\left(y, \frac{x}{y}\right) \frac{1}{y} dy - \int_{-\infty}^0 f\left(y, \frac{x}{y}\right) \frac{1}{y} dy \\ &= \int_{-\infty}^\infty f\left(y, \frac{x}{y}\right) \frac{1}{|y|} dy \end{aligned} \tag{C-2}$$