POWER SYSTEM MODELLING AND SIMULATION USING COLLOCATION METHODS

by

Hang Yin

B.Eng (Electrical Engineering and Automation),

Nanjing University of Aeronautics and Astronautics, China 2005

M.Eng. (Power Electronics and Power Drives),

Nanjing University of Aeronautics and Astronautics, China 2008

M.Eng. (Electrical and Electronic Engineering),

The University of Adelaide, Australia, 2015

Thesis submitted for the degree of

Doctor of Philosophy

in

School of Electrical and Electronic Engineering

The University of Adelaide, Australia

January 2021

THE UNIVERSITY of ADELAIDE

CONTENTS

ABSTRACT

Due to diminishing of natural resources and greenhouse effect, renewable energy is wildly used through the world. In South Australia (SA), the total wind power generation provides up to 40% of power supply to the consumer in 2019, and the rooftop solar PV generation installed capacity is steadily increase. Whereas, due to the uncertain and variability features of these renewable generation, their impact on power system planning and operation is significant and needs to be considered. This thesis introduces a novel power system planning tools based on the "so called" collocation method to handle uncertainties and variability. The proposed approach includes three major parts: power system uncertainties modelling; probabilistic power flow (PPF) computation; and guidelines for practical applications in power system planning.

First part of the thesis presents methods for power system uncertainties modelling. As an example, uncertainties of SA power network are considered. They include system demand, PV generation, wind power generation, and operation of interconnectors to Victoria (VIC). Historical data is used to construct probabilistic models to represent uncertainties. For system demand, available data are directly used to construct a probabilistic model. Typical type of PV generation in SA is roof top which are evenly spread in each region. Variability patterns of PV generation are highly related to weather and seasons. In addition, those patterns are correlated with the system demand as well. Hence, in our study, the PV generation data are combined with the system demand data. In modelling wind power generation, wind speed data acquired from nearby weather stations are used. Mapping of wind speed data to corresponding wind power generation data is proposed by applying collocation method. Uncertainty in operation of the interconnectors between SA and VIC is accounted for by considering the tie-line as system demand.

The second part of the thesis details the PPF computation by using collocation method. The traditional deterministic power flow (DPF) computation method does not consider the probabilistic nature of power system uncertainties, therefore, the calculation results from DPF computation may unrealistically assess the power system performance. Hence, it is imperative to change the DPF computation method to include the impact of those system uncertainties. The commonly used PPF simulation method, the Monte Carlo

Simulation (MCS), can calculate accurate simulation results but with expense of high computational burden. To overcome this limitation, collocation method is used to improve the computational efficiency. In this thesis, two collocation methods are proposed to account for different circumstance of PPF analysis. Historical data of SA power system is used to evaluate the practicability and feasibility of those methods. First of those two proposed methods is Probabilistic Collocation Method (PCM) which is based on the orthogonal polynomials and the Gaussian quadrature integration. This mathematical method is based on multivariate interpolation on a regular grid. Its computational efficiency is highly affected by the system dimension. When a system with a small number of uncertainties is considered, the PCM computation is faster than the MCS. However, for high number of system uncertainties the computation time is large and comparable to the MCS method. To overcome this limitation, another collocation method is introduced to solve high dimension PPF analysis, named the Sparse Grid Interpolation (SGI). In this interpolation method only a small subset of regular grid collocation points is used. This approach vastly improves the computation efficiency when handling high dimension PPF computation.

In the last part of thesis, the demonstration and guidelines for practical applications of collocation method in power system planning are presented. Comparing with the MCS method, the significant advancement in computation efficiency in solving PPF model of SA makes this method makes this method favourable for practicing engineers. In this part, not just the historical power system data of SA are used, but also the forecasted data. Objective is to predict the impact of varying and increasing demand and generation with uncertainties to the transmission network operation. Computational effectiveness and practicability of the proposed novel methodology is evaluated via comparison with the well-established MCS method.

ORIGINALITY DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed:_____ _____

Date:20/01/2021

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my principal supervisor, Dr Rastko Živanović, for his constant guidance and encouragement throughout this study. Without his advice and support my research work would not have been possible. I would like to express my deepest appreciation to my co-supervisor, Dr Said Al-Sarawi for his valuable suggestions and constructive advice. I also thank Assoc. Prof. Wen Soong, the Head of School, Dr Withawat Withayachumnankul, the Postgraduate Coordinator, for their kindly help and advice.

I would also like to thank the professional staff in the School of Electrical and Electronic Engineering, Mr Stephen Guest, Mr David Bowler, Mr Pavel Simcik, Ms Deborah Koch, Mrs Rose-Marie Descalzi, Ms Daphne Zammit, Mrs Laura McNamara, Mr Mark Innes, Mr Ian Linke and Mr Ryan King, for their kind help and support throughout my PhD study. I am really grateful to them. My thanks and appreciations also go to my colleagues for their selfless help.

Last but not least, I deepest gratitude to my parents for the love and encouragement provided to me through my entire life. I also would like to thank my wife, Ms Lina Yuan, for her continuous support, inspiration during my PhD study.

THESIS CONVENTIONS

Typesetting

This thesis is typeset using the Microsoft Office software. Word 2016 was used as text editor. Visio 2013 was used to produce drawings.

Spelling

The Australian English spelling is adopted in this thesis.

Referencing

Institute of Electrical and Electronics Engineers (IEEE) Transaction style is used for referencing and citation in this thesis.

Units

International System of Units (SI units) are used in this thesis.

PUBLICATIONS

Journal

[1] H. Yin and R. Zivanovic, "Using probabilistic collocation method for neighboring wind farms modelling and power flow computation of South Australia grid," *IET Generation, Transmission & Distribution*, vol. 11, no. 14, pp. 3568-3575, Oct. 2017.

[2] H. Yin and R. Zivanovic, "Practical application of collocation method in power flow study of South Australia grid," *International Journal of Electrical Power & Energy Systems*, vol. 94, pp. 160-170, Jan. 2018.

[3] H. Yin and R. Zivanovic, "Probabilistic power flow computation using collocation method and including correlation modeling," *International Transaction on Electrical Energy System*, vol. 29, no. 4, e2796, Apr. 2019.

Conference

[1] H. Yin and R. Zivanovic, "An application of probabilistic collocation method in wind farms modelling and power system simulation," *2016 IEEE Innovative Smart Grid Technologies – Asia,* Melbourne, Australia, Dec. 2016, pp. 681-686.

[2] M. Crnkovic, B. Parker, R. Korte, H. Yin and R. Zivanovic, "Managing uncertainty in the power flow studies of South Australian transmission network," *CIGRE 2018*, Paris, France, Aug. 2018, C1-202.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

SA	South Australia
PV	Photovoltaic
DPF	Deterministic power flow
PPF	Probabilistic power flow
MCS	Monte Carlo simulation
SRS	Simple random sampling
LHS	Latin Hypercube sampling
PDF	Probabilistic density function
LSS	Latin Supercube sampling
FFT	Fast Fourier transform
PEM	Point estimation method
PCM	Probabilistic collocation method
SGI	Sparse grid interpolation
DASGI	Dimension adaptive sparse grid interpolation
CDF	Cumulative Distribution Function
GMM	Gaussian Mixture Model
FGI	Full grid interpolation
EP	Eyre Peninsula
YP	York Peninsula
MN	Mid North

MET	Metropolitan
RL	Riverland
SE	South East
EUSE	Expected Unserved Energy
VCR	Value of Customer Reliability
VIC	Victoria

Chapter 1

INTRODUCTION

THIS chapter is intended to provide a brief introduction on power system modelling, simulation and planning. The motivation for the research conducted in this doctorate thesis is presented. The original contribution is presented and substantiated with the summary at the end of this chapter.

1.1 Introduction and motivation

Due to the decreasing of natural resources and greenhouse effect, the renewable energy is in the spotlight around the world and becomes a significant part in power generation [1]. In South Australia (SA), since 2000, the total install capacity of wind generation is up to 2142MW, and about 1078 MW of rooftop solar photovoltaic (PV) generation is spread through the major regions of state since 2009, In recent years, new solar farms have been proposed [2]. Those installed renewable generation provided up to 49% of power supply to the consumer in 2018-19 [3]. However, because of the fluctuating and intermittent features of those renewable resources, due to wind speed and solar irradiation, they cannot be scheduled and dispatched as with traditional synchronous generators, therefore this will affect planning and operating a power system [4]. In addition, traditional power system time-varying, such as system demand, are deeply associated with human living habits, and they are varying daily and seasonally [5]. In this view, as a conventional power flow analysis method, the deterministic power flow (DPF) analysis lacks modelling of the probabilistic nature of power system uncertainties. Therefore, to resolve this issue, probabilistic power flow (PPF) analysis was firstly introduced in 1974 [6], [7]. Although PPF analysis is still computationally similar to conventional power flow calculation as DPF, it does not just calculate the values of system variables such as bus voltages, power flows, frequency, etc., but also has the ability to quantify the probability of the impact of uncertain inputs to the power system variables [6], [7].

Currently, a variety of methodologies, each with pros and cons, are proposed to solve PPF problem. These methods can be classified into three groups: numerical methods, analytical methods and approximate methods.

The commonly used numerical method is Monte Carlo simulation (MCS) method. This is a conceptually simple and straightforward method which can reveal the stochastic behaviour of system random variables. The MCS method consists of repeatedly running the same model with different system inputs (i.e. samples) that follow known probability density functions (PDFs). One common way to select the MCS samples is simple random sampling (SRS) [8]. By randomly selecting the samples from the population, the MCS-SRS can easily depict the accurate stochastic behaviours of the desired outputs. However, the computation cost of this method is quite heavy due to large amount of simulation points involved to achieve high accuracy simulation results.

To further improve the computation efficiency, Latin hypercube sampling (LHS) is applied to MCS. This method is adopted for PPF analysis with success [9]-[11]. Comparing with SRS, with the same amount of sampling points, the LHS can cover more of the sampling space of inputs, therefore, the PDFs of those inputs are better represented, thus higher simulation accuracy can be achieved. On the other hand, to achieve the same accuracy of the simulation results as MCS-SRS, less sampling points are involved into the simulation by using MCS-LHS, hence the computation efficiency is improved. For high-dimensional system simulation, or higher accuracy results, the Latin supercube sampling (LSS) is introduced which can more precisely capture the statistical information of desired output variables [12], [13].

The second group of PPF analysis approaches, named the analytical methods, have been also extensively applied to power system probabilistic analysis. Those methods mainly comprise of: the fast Fourier transform (FFT) method [14]; the multilinear algorithm based method [15]; the cumulant method combined with 3 types of expansions which are Gram-Charlier expansion [16]-[18], Cornish-Fisher expansion [17], [18] and Edgeworth expansion [18]. Compared to MCS method, instead of spending large amount of computation time on DPF model, the analytical method uses mathematical expressions to solve the PPF computation, therefore, it is faster than numerical method in solving the PPF cases, the main concerns in this approach are the complexity of building the mathematical expressions and the accuracy limited by assumptions and approximations [7].

The last type of PPF analysis method is named the approximation method, and it uses a similar numerical mechanism as MCS. A typical approximate method, the point estimation method (PEM) was first introduced to deal with probabilistic analysis in general [19], [20], and then it was applied to handle PPF computation [21]-[24]. The PEM uses deterministic sampling technique to solve probabilistic problems, with fewer samples required. It can achieve similar accuracy as using MCS method, but less computational effort is involved. The PEM normally just provides the statistical moments (e.g. mean and standard deviation) of desired outputs, instead of complete statistical features (i.e. PDF). Computation cost depends on the number of system uncertainties. Besides PEM, another approximate method, named collocation method, is introduced to handle power system probabilistic analysis [25]-[37]. Two typical collocation methods are probabilistic collocation method (PCM) and sparse grid

interpolation (SGI) method. The PCM was mainly adopted to solve various power system probabilistic problems [25]-[31], such as evaluating the impact of uncertainties in power system dynamic simulation [25]-[27]; small disturbance analysis [28], [29]; handling power system damping and voltage collapse analysis [30]; uncertainty quantification in state estimation [31]; and it starts to be used in handling the PPF analysis [32], [33]. Similar as the PEM, the PCM can provide the statistical moments, moreover, the complete random samples of desired outputs can also be delivered. Because the computation cost of the PCM is highly affected by system dimension [31], and it is normally used to deal with low dimension system. To overcome this limitation, the SGI method was introduced which can provide the complete statistical features of selected system outputs and it is computationally efficient in solving low to moderate dimensional problems [31], [34]-[40]. Furthermore, based on the SGI, the dimension adaptive sparse grid interpolation (DASGI) method is developed to handle high dimensional PPF computation [37], [40]. In [37], the IEEE test system was used to show that the DASGI has similar accuracy as MCS but striking time savings. Furthermore in [40], the DASGI is implemented for PPF of South Australian transmission system. This is the first practical demonstration of DASGI where effectiveness and feasibility of solving realistic power system PPF problems were evaluated.

Another significant factor that affects the accuracy of PPF analysis is the inherent dependent relationship between system uncertainties. In most of presented research [6], [7], [9]-[11], [13]-[18], [21]-[24], the dependence between the uncertainties is neglected; or only linear dependence is assumed. In reality, the dependences between system uncertainties are nonlinear [37], [40]-[43], hence, ignoring the dependence or using linear approximation may cause misleading results. To address this issue, the Copula theory is applied [37], [41]-[44]. The Copula describes the inherent nonlinear dependence of random variables based on their uniformed marginal probability distribution. Instead of using the PDFs, the cumulative distribution functions (CDFs) of the variables are used to generate marginal univariate distributions, and then according to a specified Copula function, the nonlinear dependences are created. Combined with DASGI [37] or previously used MCS method [41]-[44], the Copula method is a proper and efficient way to build nonlinear dependences between power system uncertainties. In [33] and [40], the newly developed PCM based system uncertainties correlation modelling method can not only mathematically describe the interdependent

relationship between uncertainties but also has the ability to reduce the number of system uncertainties, hence further improving the computation efficiency by using the collocation method to conducting the PPF analysis.

To summarise: the MCS based PPF analysis methods discussed above can achieve high accuracy of desired outputs with complete statistical features revealed, however, the computation cost is very high. Therefore, analytical methods and early approximation methods are introduced to overcome this limitation. Further improvements are the collocation methods, belonging to the class of approximation methods. The two types of collocation method are: the PCM (normally used to provide the statistical moments when dealing with power system probabilistic analysis) and the SGI method that improves the PPF computation efficiency.

The collocation methods are a unique, fast methods which are both computation efficiency and accurate. Therefore, the motivation of this PhD research work is to choose the collocation method as the basic PPF analysis method to do probabilistic analysis of a realistic power transmission network. Based on the characteristics of the two typical collocation methods, the PCM is used to conduct low dimensional PPF computation, while the SGI is applied to high dimension case. Meanwhile, the PCM is also developed to build the system uncertainties model, such as the dependent model of uncertainties related to wind power generation. This new proposed dependence modelling method has the ability to reduce the system dimension, hence, the computation cost is furtherly decreased.

1.2 Statement of original contributions

This thesis presents several original contributions within two themes: the modelling of correlated system uncertainties and the PPF computation. The contributions are:

- Based on the PCM, a stochastic correlation model of the wind speeds in neighbouring wind farms is built. This model has the ability to reduce the number of system uncertainties. By constructing a polynomial mapping equation, the wind speeds in a wind farm can be represented by the wind speeds of the neighbouring wind farm, hence, the number of system uncertainties is reduced, therefore, lower PPF computation cost can be achieved by using collocation

method. The details of the proposed modelling method are presented in a journal publication [33].

- The PCM is applied in running the PPF computation. This method has similar simulation patterns as MCS. However, a new simulation model, named as a polynomial mapping equation, is constructed to replace the DPF model to handle the PPF simulation. This surrogate model has extremely low computation cost which results in significant improvement in the PPF simulation speed. The comparison between simulation results obtained using the PCM and those based on the MCS is presented in a conference [32] and journal publication [33].
- By using wind farms historical data, the realistic model of wind power generation is constructed via application of the PCM. The PCM directly generates a polynomial mapping from sampled wind speed data to wind generation power data. The factors affecting the actual wind power generation, such as wind turbine parameters, atmospheric temperature, etc., are not needed to be considered separately. The details of wind power generation modelling are presented in a journal publication [38].
- The SGI method is adopted to solve realistic power system PPF computation. Based on the SA power system historical data and aggregated transmission network model, the SGI is compared with MCS method in both accuracy and computation efficiency. The simulation results demonstrate the incredible time savings when using SGI method with similar accuracy achieved. Furthermore, the SGI method is applied to test probabilistic planning approach in transmission service provider ElectraNet, and the results are presented at the CIGRE Session 2018 in Paris [39]. With this method, a network planner can, for each contingency, calculate the probability of constraints and load shedding occurrences and predict the impact of load shedding.
- Based on the PCM, multivariate PDFs method is introduced to construct more general system uncertainties correlation model. Without any additional complex analysis step, this method is more practical to handle large number of interdependent uncertainties in realistic power system case. Furthermore, in combination with the DASGI method, the computation cost of PPF calculation is significantly reduced, therefore, making the method more practical and competent to handle high dimensional PPF analysis with dependences among

input variables. The proposed correlation modelling method is presented in a journal publication [40].

1.3 Thesis structure

The thesis structure is shown in Figure 1.1. The thesis consists of seven chapters, and summaries of each chapter are provided below.

Chapter 1 provides a state of art introduction on probabilistic power system modelling, simulation and planning. The motivation of the research presented in this thesis is elaborated. The structure of the thesis and the summarises of the original contributions are discussed.

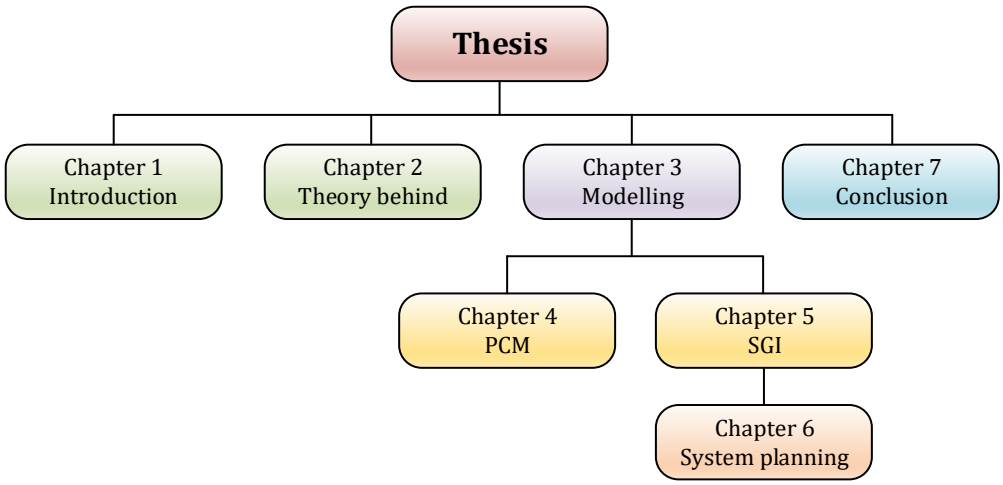

Figure 1.1. The tree diagram of the thesis.

Chapter 2 describes the basic theories behind the proposed PPF analysis methods which include PCM and SGI. Based on the orthogonal polynomials and the Gaussian quadrature integration, the PCM creates the polynomial mapping from uncertain inputs to desired outputs. This arbitrary mapping can easily reveal the statistical characteristics, and furthermore, the complete random samples of desired outputs or their PDFs can be delivered. It is shown that the PCM is more competent in handling low or medium system dimension cases compared to the MCS. The SGI, introduced by Sergey A. Smolyak [49], is discussed in the chapter. It is based on a special way of tensor product grid and can handle large number of dimensions. Similar as the PCM, the SGI-based method has the ability to provide the statistic features and entire PDFs of selected outputs.

Chapter 3 proposes the power system uncertainties modelling technique. Based on the SA power system historical data, the uncertainties considered here are divided into two

types: system demands and wind farms generation. In order to capture more statistic features of system uncertainties, the Gaussian Mixture Model (GMM) is applied to the historical data. The method has the ability to present an accurate approximation to the non-Gaussian distributions. By using the PCM, the realistic relationship between wind speed and wind power are constructed. Moreover, the interdependences between system uncertainties are considered, and two uncertainty correlation modelling methods are proposed which are all based on the PCM. Those proposed uncertainty correlation modelling methods can reduce the number of system uncertainties. Therefore, they make the collocation method be more competent to handle higher dimensional PPF analysis.

Chapter 4 details the PPF computation based on the PCM. The SA regional aggregated transmission network model is used to verify the proposed PPF analysis methods. The PCM generates using a small number of samples (i.e. collocation points) a polynomial mapping equation to substitute the DPF model. Now, rather than calling the DPF model for every simulation sample, a surrogate model is used where the computation time is dramatically reduced. However, the collocation points selected for the PCM will exponentially increase when increasing the number of system uncertainties (i.e. problem dimension). To make the PCM be more practical and competent to handle PPF analysis, the PCM based uncertainty correlation model is combined with the PCM. This approach has the ability to reduce the number of system uncertainties, hence, further improving the computation efficiency of the PCM to solve the PPF computation.

Chapter 5 shows another PPF computation method based on the DASGI approach. To overcome the limitation of the PCM in handling high dimension in PPF computation, the DASGI method is applied to the SA regional aggregated transmission network model. Compared with the conventional SGI method, the DASGI further reduces the computation cost for handling the PPF analysis. Similar as the PCM, instead of using the DPF model, the DASGI method creates a surrogate model which has incredible computation efficiency. Meanwhile, combined with the PCM based correlation model, the computation burden is further reduced which makes the DASGI method more competent to solve high dimension PPF problem.

Chapter 6 presents a practical application of the SGI method in probabilistic power system planning. According to the high computation efficiency and accuracy, the SGI method is adequate to handle power system economic planning. Based on the SA

regional aggregated transmission network model, the case study shows how to properly justify a network investment and to calculate the expected unserved energy as a measure of the worth of the investment. It is shown that for this practical task the SGI method can achieve high computation efficiency as compared to the MCS method.

Chapter 7 concludes the thesis and presents the outlook of the directions for the future research.

Chapter 2

THEORY BEHIND THE COLLOCATION METHODS

THE collocation methods proposed to solve the PPF computation problem in this thesis are the Probabilistic Collocation Method (PCM) and the Sparse Grid Interpolation (SGI) method. The fundamental background mathematics of the PCM are the orthogonal polynomials and the Gaussian quadrature integration. Whereas, the SGI method is based on the Smolyak's construction for a sparse grid interpolation. The outcome of both methods is construction of a surrogate model which has the ability to improve efficiency in the PPF computation.

2.1 Probabilistic Collocation Method

The PCM was first developed to construct a global climate model [45]-[46] by creating a polynomial mapping between the uncertain inputs and the desired outputs. It can reduce the order of model with only few properly selected simulation points. The coefficients of this polynomial mapping equation are computed by using those collocation points. The basic mathematical methods used in the PCM are the orthogonal polynomials and the Gaussian quadrature integration [47].

2.1.1 Orthogonal Polynomials

Inner product in the space of polynomials is defined as

$$\langle \boldsymbol{g}(\boldsymbol{x}), \boldsymbol{h}(\boldsymbol{x}) \rangle = \int_{\boldsymbol{D}} \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{g}(\boldsymbol{x})\, \boldsymbol{h}(\boldsymbol{x}) \boldsymbol{dx}, \tag{2.1}$$

where $f(x)$ is a nonnegative weight function defined within the space D. According to the definition of inner product, a pair of polynomials is said to be orthogonal only if their inner product equals zero. Then a set of orthogonal polynomials H is defined as,

$$< \boldsymbol{h_i}, \boldsymbol{h_j} >= \begin{cases} \mathbf{1}, & \boldsymbol{i = j} \\ \mathbf{0}, & \boldsymbol{i \neq j}' \end{cases} \tag{2.2}$$

where h_i is a polynomial of order i. For each order i, h_i has exactly i roots within the space of D. These roots are the crucial part of constructing the PPF polynomial mapping equation.

2.1.2 Gaussian Quadrature Integration

Gaussian quadrature integration solves the integral by evaluating the output value of each properly selected input point, and multiplying with proper weight, then applying summation [47]. The calculation of the integral is shown as,

$$\int_{\boldsymbol{D}} \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{g}(\boldsymbol{x}) \boldsymbol{dx} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{f_i} \boldsymbol{g}(\boldsymbol{x_i}), \tag{2.3}$$

where weight function $f(x)$ determines the constant coefficient f_i, and $g(x_i)$ is the evaluated output value on point x_i in region D, where the x_i are the roots of the orthogonal polynomials set H. The Eq. (2.3) is exact for the $g(x)$ expressed by a polynomial with order less than or equal to $(2n-1)$. This polynomial of order $(2n-1)$ can be represented based on orthonormal polynomials h_i with constant coefficients a_i and b_i,

$$\boldsymbol{g(x) = h_n(x)\left(a_{n-1}h_{n-1}(x) + \cdots + a_0h_0(x)\right) + b_{n-1}h_{n-1}(x) + \cdots + b_0h_0(x),} \tag{2.4}$$

Normally $h_0(x)$ is a constant. So, if Eq. (2.4) is multiplied by $h_n(x)$ on both sides, the integral can be determined by the definition of orthogonal polynomials as,

$$\begin{gathered}\boldsymbol{\int_D f(x)h_n(x)g(x)\,dx = \int_D f(x)h_n(x)\Big(h_n(x)\big(a_{n-1}h_{n-1}(x) + \cdots + a_0h_0(x)\big) + b_{n-1}h_{n-1}(x) + \cdots + b_0h_0(x)\Big)dx =}\\ \boldsymbol{\int_D f(x)h_n(x)h_n(x)\big(a_{n-1}h_{n-1}(x) + \cdots + a_0h_0(x)\big)dx +}\\ \boldsymbol{\int_D f(x)h_n(x)b_{n-1}h_{n-1}(x)dx + \cdots + \int_D f(x)h_n(x)b_0h_0(x)dx =}\\ \boldsymbol{\int_D f(x)h_n(x)h_n(x)\big(a_{n-1}h_{n-1}(x) + \cdots + a_0h_0(x)\big)dx +}\\ \boldsymbol{\int_D f(x)h_n(x)b_0h_0(x)dx = \int_D f(x)h_n(x)\Big(h_n(x)\big(a_{n-1}h_{n-1}(x) + \cdots + a_0h_0(x)\big) + b_0h_0(x)\Big)dx,}\end{gathered} \tag{2.5}$$

and then if Eq. (2.5) is divided by $h_n(x)$ on both sides, the Eq. (2.3) can be rewritten as,

$$\begin{gathered}\boldsymbol{\int_D f(x)g(x)\,dx = \int_D f(x)h_n(x)a_{n-1}h_{n-1}(x)dx + \cdots +}\\ \boldsymbol{\int_D f(x)h_n(x)a_0h_0(x)dx + \int_D f(x)b_0h_0(x)dx = b_0\int_D f(x)h_0(x)dx.}\end{gathered} \tag{2.6}$$

Furthermore, if setting $h_0(x)$ to one, and then the integration result will equal to b_0,

$$\boldsymbol{\int_D f(x)g(x) = \sum_{i=1}^{n} f_i g(x_i) = b_0,} \tag{2.7}$$

thus, the integration can be easily solved.

2.1.3 The PCM – one uncertain parameter

Based on the theories of Gaussian quadrature integration and orthogonal polynomials, given an uncertain parameter x with known PDF $f(x)$, and the output of interest is denoted as $g(x)$, it intends to construct a polynomial mapping to illustrate the relationship between the input x and the output $g(x)$. In the context of the PCM, the approximate $g(x)$ can be approximated by

$$\boldsymbol{\hat{g}(x) = k_0' + k_1'x + \cdots + k_{n-1}'x^{n-1},} \tag{2.8}$$

where k_i' are constant coefficients. Instead of monomials in the Eq. (2.8), the $g(x)$ can be represented as the sum of orthogonal polynomials $h_i(x)$,

$$\widehat{\boldsymbol{g}}(\boldsymbol{x}) = \boldsymbol{k_0 h_0}(\boldsymbol{x}) + \boldsymbol{k_1 h_1}(\boldsymbol{x}) \cdots + \boldsymbol{k_{n-1} h_{n-1}}(\boldsymbol{x}), \tag{2.9}$$

where the coefficients k_i of $\hat{g}(x)$ are obtained by using the collocation points which are the roots of $h_n(x)$. The linear system of equations for n collocation points x_i is formed using Eq. (**2.9**),

$$\begin{bmatrix} \widehat{\boldsymbol{g}}(\boldsymbol{x_1}) \\ \vdots \\ \widehat{\boldsymbol{g}}(\boldsymbol{x_n}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{h_{n-1}}(\boldsymbol{x_1}) & \cdots & \boldsymbol{h_0}(\boldsymbol{x_1}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{h_{n-1}}(\boldsymbol{x_n}) & \cdots & \boldsymbol{h_0}(\boldsymbol{x_n}) \end{bmatrix} \begin{bmatrix} \boldsymbol{k_{n-1}} \\ \vdots \\ \boldsymbol{k_0} \end{bmatrix}. \tag{2.10}$$

If setting $h_0(x) = 1$, and $f(x)$ represents the PDF of the uncertain input, and then according to Eq. (2.7), the mean of $g(x)$ is

$$\boldsymbol{E_{g(x)}} = \boldsymbol{E_{\widehat{g}(x)}} = \boldsymbol{k_0}, \tag{2.11}$$

Furthermore, the variance of $g(x)$ can be calculated,

$$\boldsymbol{\sigma^2_{g(x)}} = \boldsymbol{\sigma^2_{\widehat{g}(x)}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n-1}} \boldsymbol{k_i}^{\boldsymbol{2}}. \tag{2.12}$$

Once this polynomial mapping equation is constructed, the complete PDF of desired output can also be delivered.

The general procedures for the one uncertain parameter PCM approach are summarized in 5 steps:

- Obtain the probability density function (PDF) of each uncertain input.
- Generate the collocation points (samples) based on corresponding orthogonal polynomials.
- Solve the deterministic model with those collocation points.
- Construct the polynomial mapping equation (i.e. interpolant).
- Obtain the stochastic features of desired outputs through numerical interpolation and integration.

2.1.4 The PCM – multi-parameter problem

When the uncertain parameters are more than one, there are two possible cases: independent multivariate PCM and correlated multivariate PCM [26], [48]. In our work, the correlation model of system uncertainties is first constructed, and then the rest of those uncertainties are treated as independent, hence, in our work, only independent

multivariate case of PCM is considered.

If there are m uncertain parameters x_1, x_2, …, x_m in the system, where the corresponding independent PDFs are denoted as $f_{X_1}(x_1)$, $f_{X_2}(x_2)$, …, $f_{X_m}(x_m)$ respectively, then the output of interest $g(x_1, x_2, \dots, x_m)$ can be approximated by [48],

$$\hat{\boldsymbol{g}}(\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_m}) = \sum_{\boldsymbol{k_1}=\boldsymbol{0}}^{\boldsymbol{n_1}-\boldsymbol{1}} \sum_{\boldsymbol{k_2}=\boldsymbol{0}}^{\boldsymbol{n_2}-\boldsymbol{1}} \cdots \sum_{\boldsymbol{k_m}=\boldsymbol{0}}^{\boldsymbol{n_m}-\boldsymbol{1}} \boldsymbol{B}_{\boldsymbol{k_1},\dots,\boldsymbol{k_m}} \prod_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{m}} \boldsymbol{x_i^{k_i}}, \tag{2.13}$$

where $B_{k_1,\dots,k_m}$ are constant coefficients.

For a single uncertain parameter case, the number of coefficients of $\hat{g}(x)$ is n $(k_0, k_1, \cdots k_{n-1})$, so at least n collocation points are needed to solve for those coefficients. For m uncertain parameters case, to solve for all of the coefficient $B_{k_1,\dots,k_m}$, $\prod_{i=1}^{m} n_i$ collocation points are required, and those collocation points are the roots of corresponding orthogonal polynomials $h_{n_i}^{i}(x_i)$.

The PCM of multi-parameter problem has the same procedures as the signal uncertain parameter case, whereas number of the required collocation points needed to solve the coefficients of the mapping equation are exponentially increasing. Therefore, more computational effort is needed to generate multi-parameter mapping model. The computation efficiency of the PCM is decreasing while increasing the number of system uncertain parameters. Hence, the PCM is suitable to deal with low dimensional system analysis.

2.2 Sparse Grid Interpolation

In order to overcome the limitation of the PCM, the Sparse Grid Interpolation method is introduced to solve high dimensional power system probabilistic problems. The SGI was first proposed by Sergey A. Smolyak in 1963 [49] and it is used to integrate or interpolate high dimensional functions. Similar as the PCM, the SGI method picks up proper collocation points to generate a surrogate model to replace the time-consuming DPF model, therefore, dramatic computation time saving is achieved.

2.2.1 Full Grid Interpolation

Most of the probabilistic power system analysis problems are multivariate. The basic approach in solving such problems is using the full grid sample input data (i.e. the tensor product grid) to the method extends the univariate interpolation to multivariate

interpolation, and it is formed as [50]

$$(\boldsymbol{U}^{\boldsymbol{i}_1}\otimes\cdots\otimes\boldsymbol{U}^{\boldsymbol{i}_d})(\boldsymbol{f})=\sum_{\boldsymbol{j_1=1}}^{\boldsymbol{n_1}}\cdots\sum_{\boldsymbol{j_d=1}}^{\boldsymbol{n_d}}\left(\boldsymbol{a}_{\boldsymbol{j_1}}^{\boldsymbol{i}_1}\otimes\cdots\otimes\boldsymbol{a}_{\boldsymbol{j_d}}^{\boldsymbol{i}_d}\right)\cdot\boldsymbol{f}\left(\boldsymbol{x}_{\boldsymbol{j_1}}^{\boldsymbol{i}_1}\otimes\cdots\otimes\boldsymbol{x}_{\boldsymbol{j_1}}^{\boldsymbol{i}_1}\right), \quad (2.14)$$

where multi-index $i=(i_1,\dots,i_d)$ determines the ith interpolation level and $j_1,\dots,j_d$ are interpolation point indices for each dimension, while for each dimension the number of interpolation points is represented by $n_1,\dots,n_d$, and $x_{j_1}^{i_1},\dots,x_{j_d}^{i_d}$ are $(j_1,\dots,j_d)$ th interpolation point on ith interpolation level of each dimension with corresponding interpolation basis polynomials represented by $a_{j_1}^{i_1},\dots,a_{j_d}^{i_d}$. The tensor product is denoted by $\otimes$, therefore, $x_{j_1}^{i_1}\otimes\cdots\otimes x_{j_1}^{i_1}$ is a tensor product grid of input data. However, this general full grid tensor product interpolation has very high computational burden since the total number of interpolation points will exponentially increase when increasing the number of input variables, the so-called curse of dimensionality [36].

2.2.2 Sparse Grid Interpolation

To break the so-called curse of dimensionality, the SGI is developed introduced as the Smolyak's construction [49]. Instead of applying the general full grid tensor product, the Smolyak's algorithm extends the univariate interpolation formulas to the multivariate case with a special way of tensor product. The Sparse Grid interpolant $A(q,d)$ is constructed with a relatively small number of collocation points, and it is shown as

$$\boldsymbol{A}(\boldsymbol{q},\boldsymbol{d})=\sum_{|\mathbf{i}|\leq\boldsymbol{q}}(\Delta^{\boldsymbol{i}_1}\otimes\cdots\otimes\Delta^{\boldsymbol{i}_d}), \quad (2.15)$$

where q is determined by the sum of interpolation level l and the dimension d, i is a multi-index defined as in Eq. (2.14), and $|\mathbf{i}|$ equals to $i_1+\cdots+i_d$, and incremental interpolant Δ^i is given by

$$\Delta^{\boldsymbol{i}}=\boldsymbol{U}^{\boldsymbol{i}}-\boldsymbol{U}^{\boldsymbol{i-1}}, \quad (2.16)$$

with initial condition $\Delta^1=U^1$ and $U^0=0$. For $q\geq d$, $A(q,d)$ can be expressed by

$$\boldsymbol{A}(\boldsymbol{q},\boldsymbol{d})=\sum_{\boldsymbol{q-d+1}\leq|\mathbf{i}|\leq\boldsymbol{q}}(-\mathbf{1})^{\boldsymbol{q}-|\mathbf{i}|}\cdot\binom{\boldsymbol{d-1}}{\boldsymbol{q}-|\mathbf{i}|}\cdot(\boldsymbol{U}^{\boldsymbol{i}_1}\otimes\cdots\otimes\boldsymbol{U}^{\boldsymbol{i}_d}). \quad (2.17)$$

So $A(q,d)$ is a portion and a linear combination of tensor product operators where only a small number of collocation points are used to construct it, therefore the computation cost is strikingly reduced. Hence, comparing with full grid interpolation, the SGI is more capable to solve high dimension cases, where this is only possible and give advantage for

smooth functions [36], [51]-[52]. In the PPF problem, the smoothness of the solution with random input data is expected, so this is not limited by this problem.

2.2.3 Dimension Adaptive Sparse Grid Interpolation

Based on the standard SGI, Dimension Adaptive Sparse Grid Interpolation (DASGI) is developed to further reduce the computation effort in dealing with high dimension problems. Instead of treating all dimensions equally, the DASGI has the ability to automatically detect the importance of each dimension, and then putting more effort on the more important ones [34]-[37]. The core concept of DASGI is transforming the nodal basis to multivariate hierarchical basis [35], and the hierarchical multivariate interpolation formula is formed as [37],

$$\Delta \boldsymbol{A}(\boldsymbol{q},\boldsymbol{d}) = \sum_{|\mathbf{i}|=\boldsymbol{q}} \sum_{\mathbf{j}} \boldsymbol{a}_{\mathbf{j}}^{\mathbf{i}} \cdot \boldsymbol{\omega}_{\mathbf{j}}^{\mathbf{i}}, \tag{2.18}$$

where $a_{\mathbf{j}}^{\mathbf{i}}$ is a hierarchical basis functions, multi-index $i = (i_1, \dots, i_d)$ determines the ith interpolation level and $j = (j_1, \dots, j_d)$ is interpolation point indices for each dimension, $\omega_{\mathbf{j}}^{\mathbf{i}}$ is defined as hierarchical surpluses which are the difference of the function values between two consecutive interpolation levels. For continuous functions, the hierarchical surpluses decrease to zero when the interpolation level increases to infinity. Hence, the hierarchical surplus is fundamental ideal on implementation of adaptive processing of DASGI.

The general procedures for the DASGI approach are summarized in 4 steps [36]:

a. Set the Smolyak construction level $l = 0$.

b. For d-dimensional problem, build the corresponding sparse grid on level l, defined as $A_{d+l,d}$:

   - Calculate the function value of initial collocation points.

   - Add new generated neighboring points into active index.

   - Make $l = l + 1$.

c. Check the operation condition: $l \leq l_{max}$ and that active index is not empty:

   - Save the active index into old index set.

- Calculate the corresponding hierarchical surplus of each collocation point in the old index set.

- Add the old index set to $A_{d+l-1,d}$, update to new adaptive sparse grid $A_{d+l,d}$.

- If each collocation point in the old index set, the corresponding $\omega_{\mathbf{j}}^{\mathbf{i}} \geq \varepsilon$ (accuracy requirement), make $l = l + 1$, repeat step c again, otherwise, skip to step d.

d. Obtain the stochastic features (i.e. PDF) of desired outputs using the system uncertain inputs data.

For high dimension cases, sparse grid interpolation reduced the required number of collocation points in constructing the surrogate model, hence, much higher computation efficiency is achieved, even more, DASGI is developed to further improve the ability of handling high dimension problems. Therefore, DASGI is the most competent method to solve high-dimensional power system parametric problems.

2.3 Conclusions

This chapter describes the basic mathematical theories behind the methods adopted in our work: these are the PCM and the SGI. Although these two methods are based on different theories, they all intend to construct an efficient surrogate model to replace traditional DPF model, therefore the PPF computation cost will be reduced. Each method has its pros and cons: based on the orthogonal polynomials and the Gaussian quadrature integration, the PCM is a simple method with less computation effort involving in generating the surrogate model, whereas it is highly affected by the system dimension. Hence the PCM is more competent in solving low or medium dimension PPF analysis. The SGI developed according to the Smolyak's algorithm breaks the so-called curse of dimensionality and is adequate to solve the PPF problem with large number of uncertain parameters.

Chapter 3

POWER SYSTEM UNCERTAINTIES MODELLING

THIS chapter describes how to model uncertainties in power systems. As an example, the SA power system historical data are used to construct the system uncertainties' model. These uncertainties are mainly divided into two types which are system demands and wind farms generation. Normally the system demand is following normal distribution and the wind speed can be model by Weibull distribution. However, in practical applications these uncertainties are not strictly following those known PDFs. Therefore, the Gaussian Mixture Model (GMM) method which has the ability to approximate any non-Gaussian distribution, is introduced. Moreover, the dependences between system uncertainties should be considered. To address this issue, two correlation modelling methods are used and described in this chapter. These two proposed correlation modelling methods can reduce the number of system uncertainties which makes the collocation method more computationally competent in handling higher dimensional PPF analysis.

3.1 Introduction

To obtain the PDF of desired power system state variable, the PPF requires the PDF of the system uncertainties to be known, so accurate modelling of the system uncertainties is a critical part of PPF analysis. The main power system uncertainties considered in our work are system demand and wind farms generation. The system demand is typically following normal distribution [53]. Based on measured wind speed data, wind speed can be modelled using two-parameter Weibull distribution [54]-[59]. However, in some practical cases the PDF of system demand and wind speeds are not following any known distribution, so Gaussian Mixture Model (GMM) is used which gives an accurate approximation to the non-Gaussian distributions [38], [40], [57] - [59]. The commonly used methods to transfer wind speed to wind power are:

- methods based on wind-power curve: in most cases a simplified linear, quadratic or power-law curves are used [54];
- methods based on theoretical wind-power formula [55]-[56]: those methods are based on empirical analysis and may not be suitable or accurate for a real system.

Overcoming the limitation of commonly used wind power model, the PCM is proposed to obtain realistic relationship between wind speed and wind generation power [32], [38]. The PCM was first used in the field of global climate modelling [46], [60]-[61] which aims to map the uncertain parameters to desired outcomes with a lower order polynomial by properly choosing simulation points. The simulation results indicate that the PCM has high accuracy and low computation cost in handling non-linearities. Based on this feature, the PCM is applied to construct the wind power generation model.

In practical applications, the inherent dependences between system uncertainties are needed to be considered to obtain more realistic computation results. Hence, constructing the correlation model is another challenging task of a PPF analysis. In our work, the correlations between system demands and between neighbouring wind farms are considered. The copula theory is first applied [37], [41]-[44] to model the nonlinear dependences. Based on the uniform marginal probability distribution of the dependent uncertainties, the Copula has the ability to construct these nonlinear dependent relationships between those correlated system uncertainties. This method is verified in [37], [41]-[44] and present as a proper and efficient way to build nonlinear dependences

between power system uncertainties. In our work, according to different scenarios, two correlation modelling methods are introduced:

- small number correlated uncertainties modelled using PCM with Fuzzy logical optimization [32]-[33];
- large number correlated uncertainties modelled using a special multivariate PDFs modelling method based on PCM [40]; this approachis applied to our example SA power system to build the correlation models.

Besides building the nonlinear dependent relationship between the uncertainties, those newly developed methods can also reduce the number of system uncertainties, hence further improving the computation efficiency by using the collocation method to conduct the PPF analysis.

The remainder of this chapter is organised as follows, Section 3.2 describes the power system uncertainties modelling, in which the modelling method and types of system uncertainties are detailed. Interdependence of the power system uncertainties is considered and modelled in Section 3.3.

3.2 Power system uncertainties modelling

There are 6 regions in SA transmission system [2]. They are Eyre Peninsula (EP), Yorke Peninsula (YP), Mid North (MN), Metropolitan (MET), Riverland (RL) and South East (SE). There is an interconnector between SA and Victoria. The uncertainties of SA power network considered in our work include system demand, interconnector power flow, PV generation and wind farms generation. 5 years historical data with half hourly resolution are used to construct the stochastic model. Currently, common type of PV installation in SA is roof top panels which are approximately evenly spread across each region. Hence in our study, the PV generation is combined with corresponding regional demand data to simplify the system model.

3.2.1 Gaussian Mixture Model

The Gaussian Mixture Model (GMM) is a weighted sum of finite Gaussian distribution components. The Gaussian distribution, so called normal distribution, is defined as [62],

$$\boldsymbol{f_{Gauss(\mu,\sigma)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-((x-\mu)^2/2\sigma^2)}}, \quad (3.1)$$

where, μ and σ are the mean value and standard deviation, respectively. Then, for a one-dimensional random variable X, the PDF of $f_X(x)$ in the form of GMM is defined as [63],

$$\boldsymbol{f_X(x) = \sum_{k=1}^{n} a_k f_{Gauss(\mu_k,\sigma_k)}(x),} \tag{3.2}$$

where a_k, μ_k and σ_k are, respectively, the weight function, mean and standard deviation of the kth Gaussian components. Since the PDF is non-negative and its integral over the sample space equals to unity, the weight function has the condition of $\sum_{k=1}^{n} a_k$=1. Furthermore, the mean and the variance of the variable X can be expressed as [64],

$$\boldsymbol{\mu_X = \sum_{k=1}^{n} a_k \mu_k,} \tag{3.3}$$

and

$$\boldsymbol{\sigma_X^2 = \sum_{k=1}^{n} a_k (\sigma_k^2 + (\mu_k - \mu_X)^2).} \tag{3.4}$$

Figure 3.1. shows an example of a random variable X modelled by a GMM with five components. The constructed GMM is sum of the individual weighted Gaussian components. Apparently, although each component is following Gaussian distribution, the sum of these components is non-Gaussian distribution.

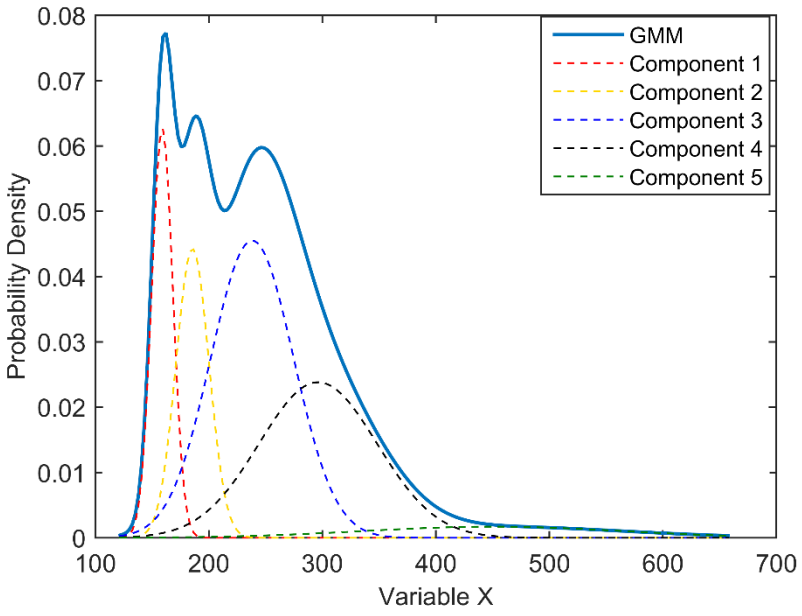

Figure 3.1. GMM approximation with 5 components.

The common procedure to fit a GMM model has the following steps:

a. Obtain the frequency histogram of the data which counts the number of data dropping into each range represented with *Bins*.
b. Compute the discrete probability P_i of each *Bins* as $P_i = f_i/f_{total}$, where f_i is the

number of data located in the ith Bin, f_{total} is the total number of data.

c. Set a minimum number of Gaussian components m, solve each parameter of those Gaussian components by using expectation maximization (EM) algorithm. This is an iterative method to find the maximum likelihood estimates of the parameters of the Gaussian components for given data [65].
d. Compute the coefficient of determination R^2 which indicates how well a statistical model fits data [66], in regression, the R^2 measures the goodness of fit of the regression predictions approximating the sampled data point. And then increase the number of Gaussian components to $m+1$, repeat step c with new number of Gaussian components, compute R^2 again and compare with previous one, repeat this step until the difference value $R^2_{m+1} - R^2_m$ is below the setting threshold value Δ.

The coefficient of determination R^2 is formed as,

$$\boldsymbol{R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{3.5}$$

where y_i is the probability of sampled data, f_i is the corresponding probability of estimated curve, and $\bar{y}$ is the mean value of y_i. The value of R^2 is close to 1 indicates that the regression line is more perfectly fitting the data. There is no unified specification on this goodness of fit metric, so by considering the computation efficient and the complexity of the model, the threshold value is set to $\Delta = 0.002$, just for demonstrating the procedures of constructing the GMM. Considering the computation cost, the maximum number of Gaussian components m is set up to 7.

3.2.2 System demand

For demonstration purpose, the region MET is divided into two aggregated sub regions, MET1 and MET2, respectively. And the demand data of sub region MET1 of MET is used to demonstrate the GMM processing.

First, the frequency histogram of demand data is generated and then the corresponding discrete PDF is shown in Figure 3.2. Based on the discrete PDF, following the GMM steps above, when the number of Gaussian components $m = 5$, the difference of two consecutive R^2 is below the threshold value $\Delta = 0.002$.

The Figure 3.3. shows the estimated PDF with blue solid line which is matching well the discrete PDF shown in Figure 3.2. with the $R^2 = 0.9999$. It is composed of 5 Gaussian components shown in Figure 3.3. using dashed lines. The same steps are applied to other regional demand data, the values of R^2 for each m and the difference of two consecutive R^2 are shown in Table 3.1.

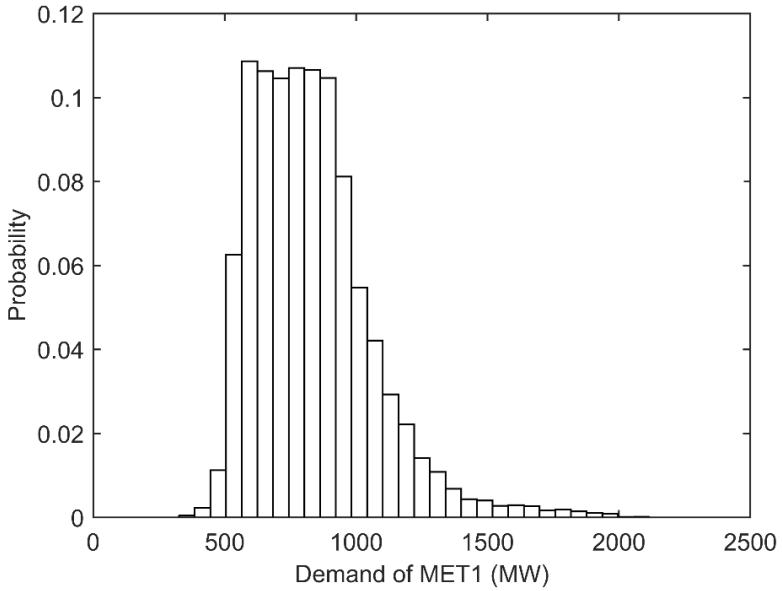

Figure 3.2. Discrete PDF of demand in MET1.

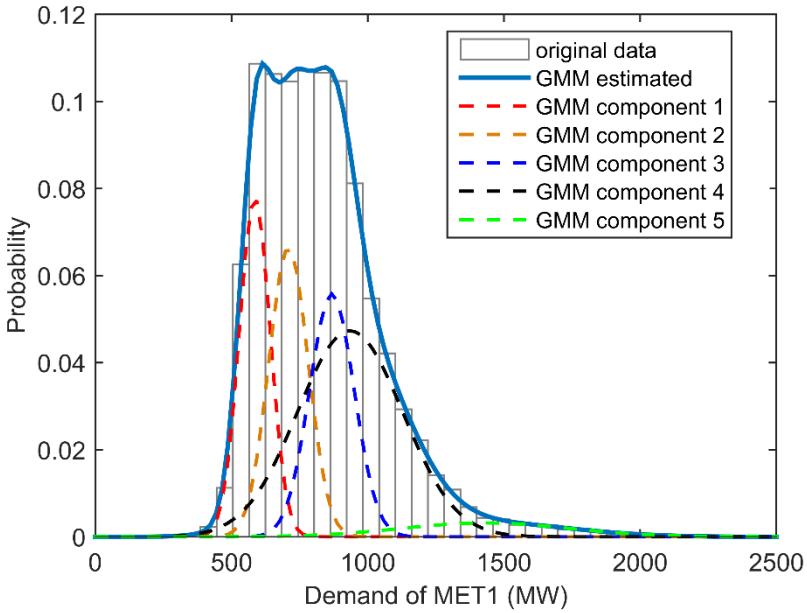

Figure 3.3. GMM estimated PDF of demand data of MET1.

In Table 3.1, all the values of R^2 close to 1 are shaded. This indicates that the estimated results based on the GMM can match well the discrete PDFs generated by sampled data. Except the region EP, all the differences between two consecutive R^2 are below the threshold value of 0.002. In the region EP, the values of R^2 at $m = 2$ and $m = 3$ are identical which indicates that further increase of the number of Gaussian components will not improve modelling accuracy. Hence, $m = 2$ is chosen for the region EP model to reduce the complexity.

Table 3.1 The R^2 and the difference with previous R^2 at each m of regional demand data by GMM.

Region	R^2 (difference with previous R^2)						
	m=1	m=2	m=3	m=4	m=5	m=6	m=7
EP	0.9922 (null)	0.9993 (0.0071)	0.9993 (0)				
YP	0.9512 (null)	0.9926 (0.0414)	0.9963 (0.0037)	0.9984 (0.0021)	0.9997 (0.0013)		
MN	0.9744 (null)	0.9859 (0.0115)	0.9992 (0.0133)	0.9995 (0.0003)			
MET1	0.9535 (null)	0.9899 (0.0364)	0.9971 (0.0081)	0.9994 (0.0023)	0.9999 (0.0005)		
MET2	0.9842 (null)	0.9956 (0.0114)	0.9980 (0.0024)	0.9997 (0.0017)			
RL	0.9270 (null)	0.9912 (0.0642)	0.9961 (0.0052)	0.9983 (0.0021)	0.9992 (0.0009)		
SE	0.9977 (null)	0.9988 (0.0011)					

3.2.3 Interconnector

There are two interconnectors between SA and VIC which are Murraylink (located between the region RL of SA and Red Cliffs of VIC) and Heywood interconnector (located between the region SE of SA and Heywood of VIC), and they provide exchange of electrical power between two states. In our project, the Murraylink interconnector data has been combined and modelled with the demand data of region RL of SA, so only the Heywood interconnector is considered here.

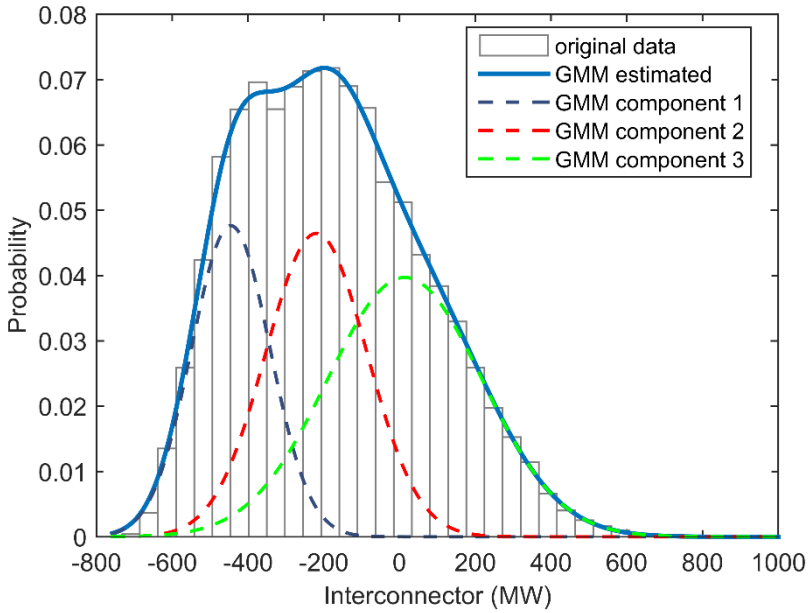

Figure 3.4. GMM estimated PDF of Interconnector data.

The Heywood interconnector power is represented as a system load of SA. When exporting power from SA, the interconnector power is a positive load to SA, otherwise, it acts as a negative load to SA. The estimated PDF by GMM compared with the sampled discrete PDF is shown in Figure 3.4. The GMM of the Interconnector is composed of 3 Gaussian components with different coloured dashed lines; the $R^2 = 0.9998$, and the values of R^2 at each value of m are shown in Table 3.2.

Table 3.2 R^2 and difference with previous R^2 at each m of Interconnector data by GMM.

Region	R^2 (difference with previous R^2)						
	m=1	m=2	m=3	m=4	m=5	m=6	m=7
MN	0.9649 (null)	0.9941 (0.0292)	0.9998 (0.0057)	0.9998 (0)			

In Table 3.2, similar as for the region EP shown in Table 3.1, the values of R^2 at $m = 3$ and $m = 4$ are identical which indicates that further increase of the number m does not increase accuracy. Hence, $m = 3$ is selected.

3.2.4 Wind farms generation

The wind power generation is extremely depended on wind speed, and the wind power is not typically following any known distributions, hence, the common procedures to model wind power generation include wind speed modelling and wind speed to wind power converting. The wind speed normally follows two-parameter Weibull distribution. In our work, instead of using Weibull distribution, the already presented GMM method is applied to fit wind speed data. The PCM is applied to construct the realistic relationship between wind speed and wind power. In this approach, the wind speed data and wind power data of YP region are used to explain the procedures of modelling.

3.2.4.1 Wind speed modelling

The similar GMM procedure used to model the system demand and interconnector power are applied to wind speed data of YP region. The estimated PDF modelled by GMM is compared with the sampled discrete PDF and shown in Figure 3.5. To meet the selected threshold value $\Delta = 0.002$, 5 Gaussian components are used, and the final $R^2 = 0.9991$, which is very close to 1, is obtained. This value of R^2 indicates that the estimated PDF is well matched the wind speed data. The estimated PDF, shown by blue solid line in

Figure 3.5. is compared with Weibull distribution shown by red solid line with the parameters to ($\lambda = 5.603, k = 2.329$). The comparison shows that, to some degree, it has similar trend and shape of Weibull distribution, whereas, it pays more attentions to the details and gives more accurate description of the actual wind speed data. The values of R^2 corresponding to each value of m are shown in Table 3.3. Similar as discussed above, when the $m = 5$, the value of Δ reduces below the set threshold value before reaching the maximum number of Gaussian components.

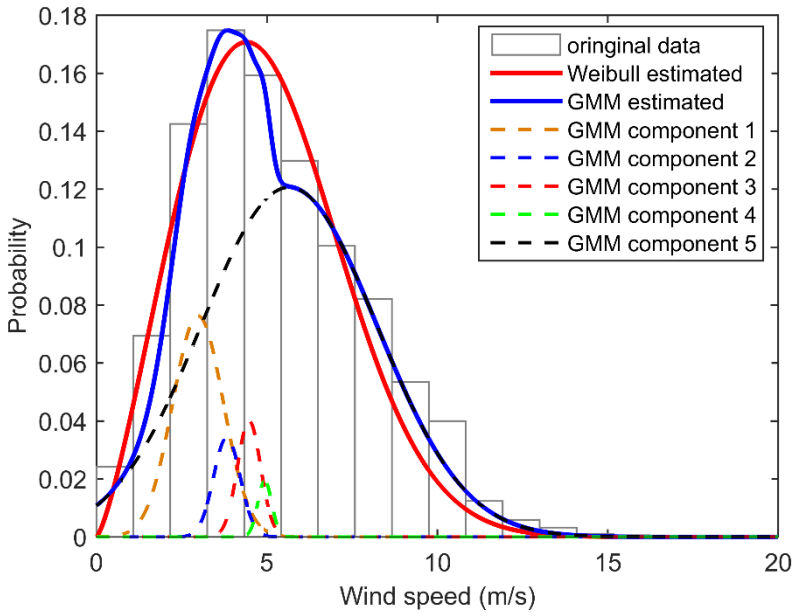

Figure 3.5. GMM estimated PDF of wind speed data of YP region.

Table 3.3 R^2 and difference with previous R^2 at each m of wind speed data of YP region fitted by GMM.

Region	R^2 (difference with previous R^2)						
	m=1	m=2	m=3	m=4	m=5	m=6	m=7
MN	0.9649 (null)	0.9941 (0.0292)	0.9968 (0.0027)	0.9989 (0.0021)	0.9991 (0.0002)		

3.2.4.2 Wind power modelling

Based on the PCM theory, details of modelling the relationship between wind speed and wind power are:

a. Obtaining the orthogonal polynomials set of wind speed. The orthogonal polynomials set can be generated by using Gram-Schmidt processing [47] with 2 steps: first generating the monic orthogonal polynomials as shown in Eq. (3.6), where $(p_0, p_1 \dots, p_n)$ is the monic orthogonal polynomials set, and the PDF of x is

chosen as the weight function of the inner product; second generating the orthogonal polynomials as shown in Eq. (3.7).

$$\begin{cases} \boldsymbol{p_0(x) = 1} \\ \boldsymbol{p_1(x) = x - \dfrac{\langle x, p_0(x)\rangle}{\langle p_0(x), p_0(x)\rangle} p_0(x)} \\ \boldsymbol{p_2(x) = x^2 - \dfrac{\langle x^2, p_0(x)\rangle}{\langle p_0(x), p_0(x)\rangle} p_0(x) - \dfrac{\langle x^2, p_1(x)\rangle}{\langle p_1(x), p_1(x)\rangle} p_1(x)} \\ \vdots \end{cases}, \tag{3.6}$$

$$\begin{cases} \boldsymbol{h_0(x) = p_0(x)} \\ \boldsymbol{h_1(x) = \dfrac{p_1(x)}{\langle p_1(x), p_1(x)\rangle}} \\ \boldsymbol{h_2(x) = \dfrac{p_2(x)}{\langle p_2(x), p_2(x)\rangle}} \\ \vdots \end{cases}. \tag{3.7}$$

Based on the Eq. (3.6) and (3.7), with the estimated PDF of wind speed in YP region $f(v)$, the orthogonal polynomials set can be generated,

$$\begin{cases} \boldsymbol{h_0(v) = 1} \\ \boldsymbol{h_1(v) = 0.4104v - 2.0066} \\ \boldsymbol{h_2(v) = 0.1253v^2 - 1.3966v + 3.0894} \\ \boldsymbol{h_3(v) = 0.03222v^3 - 0.5989v^2 + 3.1209v - 4.2203} \\ \boldsymbol{h_4(v) = 0.007360v^4 - 0.1990v^3 + 1.7572v^2 - 5.7458v + 5.4001} \\ \vdots \end{cases}, \tag{3.8}$$

where $h_0, \dots, h_4$ are orthogonal polynomials set, v is the wind speed of YP region. The roots of each orthogonal polynomial are the collocation points used to construct the mapping polynomial.

b. Constructing the mapping equation between wind speed and wind power. From Eq. (2.8), in our study, the input x is the wind speed v, and the output $\hat{g}(x)$ is the wind power $g(v)$. For n^{th} order PCM, $n+1$ coefficients $k'_{i=(0,1,\dots n)}$ are required which can be computed from the roots of $h_{n+1}(x)$, Eq. (3.8). Because the actual wind power generation is not only just affected by wind speed but also some other factors, such as wind turbine parameters, atmospheric temperature, etc. Therefore, for same wind speed at different sampling time point, the acquired wind power data may different. Based on wind speed and wind power data of YP region, each root is generated from orthogonal polynomial, the paired wind power data of those roots are actually in a range instead of a constant value as shown in Table 3.4. For instance, constructing

the 1st order PCM: 2 coefficients are computed by 2 wind speed roots (3.1m/s and 8.2m/s) obtained from $h_2(v)$; the corresponding acquired wind power data at each wind speed root is varying in a range.

Table 3.4 Wind speed and paired wind power in YP region.

YP	Wind speed roots (m/s)	Paired wind power range (MW)
PCM-linear	3.1	0 to 40
	8.2	32 to 85
PCM-quadratic	2.1	0 to 10
	5.7	5 to 70
	10.3	56 to 86
⋮		

First, the mean values of those paired wind power data at each wind speed are used to solve for the coefficients of PCM mapping equation, the first few of polynomial equations are computed as,

$$\begin{cases} \boldsymbol{g_1(v) = 9.8361v - 14.5738} \\ \boldsymbol{g_2(v) = 0.1981v^2 + 6.1072v - 6.2466}, \\ \vdots \end{cases} \tag{3.9}$$

where $g_n(v)$ is n^{th} order PCM mapping equation.

c. Comparing the PCM approximation results with sampled data. The stochastic characters of generated wind power data from PCM can be calculated and compared with sampled wind power data, shown in Table 3.5.

Table 3.5 Expected value, variance and standard deviation of wind power in YP region.

YP	E	σ2	σ
Sampled data	31.2072	685.1636	26.1756
PCM-linear	33.5867	566.8610	23.8088
PCM-quadratic	29.1681	435.6809	20.8730
PCM-3rd order	31.1973	478.4241	21.8729
PCM-4th order	31.2153	490.1828	22.1401
PCM-5th order	31.2162	489.4592	22.1237
PCM-6th order	31.2002	486.7785	22.0631
PCM-7th order	31.2161	492.9828	22.2032

From Table 3.5, when the order of PCM is higher than 3, the mean values are very close to the actual value. This shows that the higher orders of polynomial model will capture more nonlinear features of the realistic relationship between wind speed and wind power. Relative error of those mean values compared with the mean of sampled data are shown in Figure 3.6.

Although the mean values of the higher order PCM are very close to the actual one,

the variance and standard deviation are significant different comparing with the actual value. This is caused by applying the mean value of paired wind power data at each wind speed root to compute the coefficients of the PCM mapping equation in step b. Hence, optimization needs to be implemented upon PCM results.

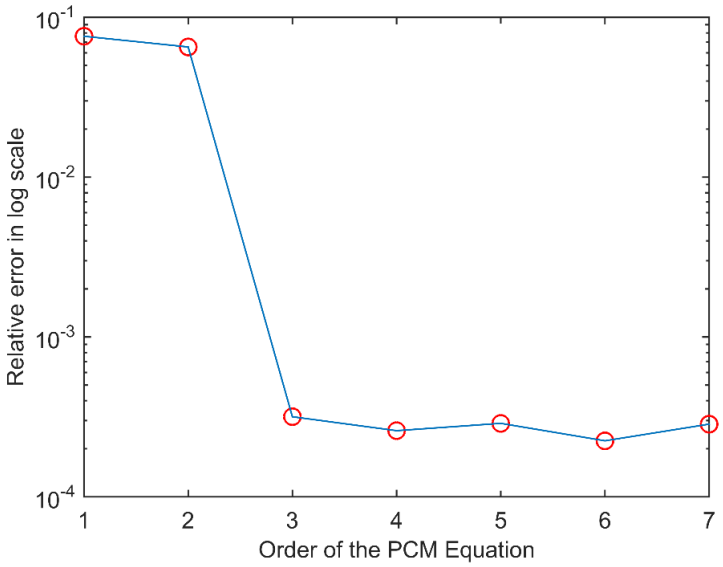

Figure 3.6. Relative error of estimated mean values.

d. Optimization. The Figure 3.6 shows that, when the order of PCM mapping equation is higher than 3, by further increasing the order of mapping equation will not significantly give better estimated results. Therefore, with considering the computation cost the 3rd PCM approximation results are used to illustrate the optimization method.

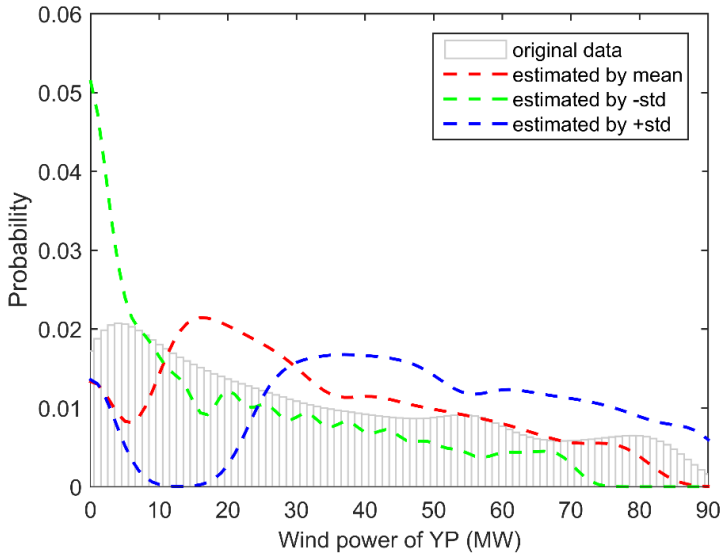

Figure 3.7. Histogram of sampled wind power and estimated wind power.

In addition to choosing the mean of wind power data for the calculation as detailed in step b, the ±σ (standard deviation) of corresponding wind power data are applied to the calculation as well. The histogram of generated wind power using mean, +σ and -σ are compared with the actual data are shown in Figure 3.7.

The grey bar plot is the sampled wind power data shown in Figure 3.7. When using mean value of wind power at each collocation point, the estimated value shown in red dashed line is mainly located in the middle range (comparing with sampled data). Whereas, applying ±σ (standard deviation) of corresponding wind power data to the calculation, shown as green and blue dashed line, make up for the lost features in the lower and higher wind power range.

Based on fuzzy logic theory, a set of membership functions are generated to model the true values of each of the input variables [67]. In our application those membership functions are applied to the 3 estimated wind power data (estimated by mean value and ±σ) to model the optimized values. First, the wind power range is divided into small sections, each with an interval of 5 MW. And then the ratio of the probability of each estimated value to actual value at each section is calculated, as follows,

$$\sum_{i=1}^{n} \boldsymbol{p_i WP_i} = \sum_{i=1}^{n} \boldsymbol{C_{-\sigma_i} p_{-\sigma\_i} WP'_{-\sigma\_i}} + \sum_{i=1}^{n} \boldsymbol{C_{m_i} p_{m\_i} WP'_{m\_i}} + \sum_{i=1}^{n} \boldsymbol{C_{+\sigma_i} p_{+\sigma\_i} WP'_{+\sigma\_i}}, \tag{3.10}$$

where n is the number of sections, p_i is the probability of actual wind power WP_i, $p_{-\sigma\_i}$ is the probability of estimated wind power $WP'_{-\sigma\_i}$ by using -σ value mapping, p_{m_i} is the probability of estimated wind power WP'_{m_i} by using mean value mapping, $p_{+\sigma\_i}$ is the probability of estimated wind power $WP'_{+\sigma\_i}$ by using +σ value mapping. Based on fuzzy logic theory, $C_{-\sigma_i}$, C_{m_i} and $C_{+\sigma_i}$ are non-negative coefficients with constrain condition of $C_{-\sigma_i} + C_{m_i} + C_{+\sigma_i} = 1$.

To solve three coefficients in Eq. (3.10) for each section, three equations are needed, they are defined as: constrain condition of $C_{-\sigma_i} + C_{m_i} + C_{+\sigma_i} = 1$; for each interval, $p_i WP_i$, $p_{-\sigma\_i} WP'_{-\sigma\_i}$, $p_{m\_i} WP'_{m\_i}$ and $p_{+\sigma\_i} WP'_{+\sigma\_i}$ are available; assuming those coefficients have similar ratio as the 3 probabilities ($p_{-\sigma\_i}$, p_{m_i} and $p_{+\sigma\_i}$).

With fuzzy logic optimization, the Table 3.6 shows the estimated wind power by optimized PCM has similar stochastic characteristic as sampled data. The histogram of both estimated wind power and sampled data is shown in Figure 3.8.

Table 3.6 Expected value, variance and standard deviation of wind power after optimization in YP region.

Yorke Peninsular	E	σ^2	σ
Sampled data	31.2072	685.1636	26.1756
PCM-3rd order	31.2077	672.2942	25.9286

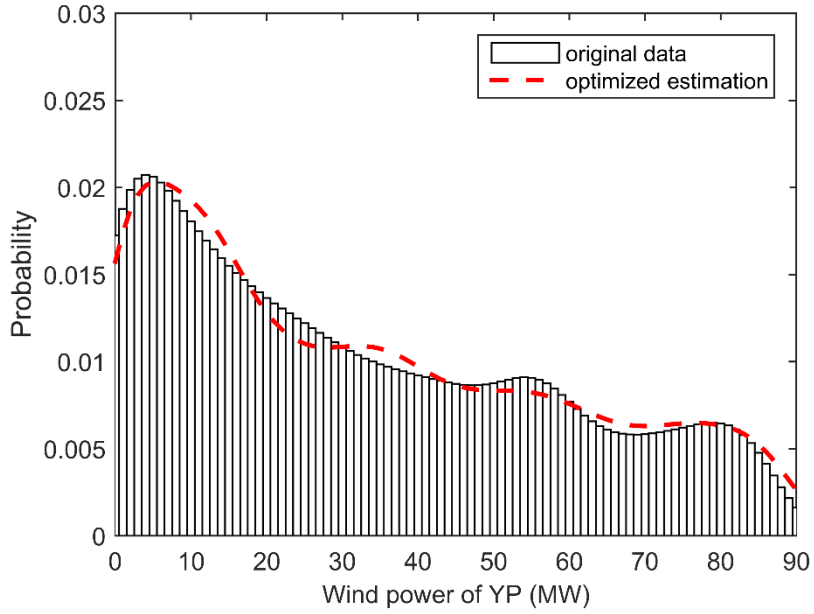

Figure 3.8. Histogram of sampled wind power and estimated wind power after optimization.

To compare the similarity of two histograms, the frequency histogram similarity index (*FHSI*) [37] is used to indicate how accurate is estimation compared to reference data. The *FHSI* is defined as,

$$\boldsymbol{FHSI} = \left(1 - \frac{1}{2}\sum_{i=1}^{N}\left[\left(\boldsymbol{Bin}_i^{ref} - \boldsymbol{Bin}_i^{est}\right)^2 + \left(\boldsymbol{P}_i^{ref} - \boldsymbol{P}_i^{est}\right)^2\right]^{\frac{1}{2}}\right) \times 100\% \quad (3.11)$$

where N is the number of *Bins*, Bin_i^{ref} and P_i^{ref} are values of reference histogram indicating the location of each *Bin* interval with corresponding probability of this *Bin*. Bin_i^{est} and P_i^{est} represent values of estimated histogram. This quantified value indicates the percentage of similarity of two histograms with higher value standing for higher accuracy of the estimated data. The *FHSI* of Figure 3.8 is 97.33% indicating the optimized PCM mapping results have excellent ability to represent the reference data. The same mapping procedures can be applied to other wind farms modelling. In this thesis, due to the confidential reason, most of the wind farms generation data are not

public, so the assumption made here is that all the wind turbines in other wind farms have the same characteristics as in the wind farms of YP region.

3.3 Power system uncertainties correlation modelling

In our work, based on the PCM, two types of system uncertainties correlation modelling methods are developed to suit different scenarios:

- for small number of correlated system uncertainties, an optimization is applied to each paired correlated uncertainties, then one of this paired data can be expressed by the another data with a polynomial mapping, hence reducing the number of the system uncertainties;
- for large number of correlated system uncertainties, the multivariate PDFs modelling method is introduced to present the dependent uncertainties, hence reducing the system dimensions.

Because the computation efficiency of collocation methods is highly affected by the system dimension, hence, both those correlation modelling methods aim to reduce the system dimensions, therefore improve the computation efficiency when using collocation method to solve PPF analysis.

3.3.1 Small amount of interdependent uncertainties modelling

The similar procedure as illustrated in section 3.2.4.2 (i.e. the PCM combined with Fuzzy logic optimization) is adopted to model the interdependent system uncertainties. Correlated wind farms data in SA is used to demonstrate this proposed method. In South Australia, most of the wind farms are installed in 3 regions, EP, MN and SE [2], and the location of those wind farms are shown in Figure 3.9.

Wind source data from weather stations of Snow Town and Clare of MN, which are close to the wind farms Snow Town (33.69°S, 138.13°E) and Waterloo (34.00°S, 138.90°E) shown in the red dashed circle in Figure 3.9 respectively, are used to illustrate the procedure of correlation modelling. The acquired data from Bureau of Meteorology Australia [68], with the resolution of 30 minutes, are used to generate the correlation model.

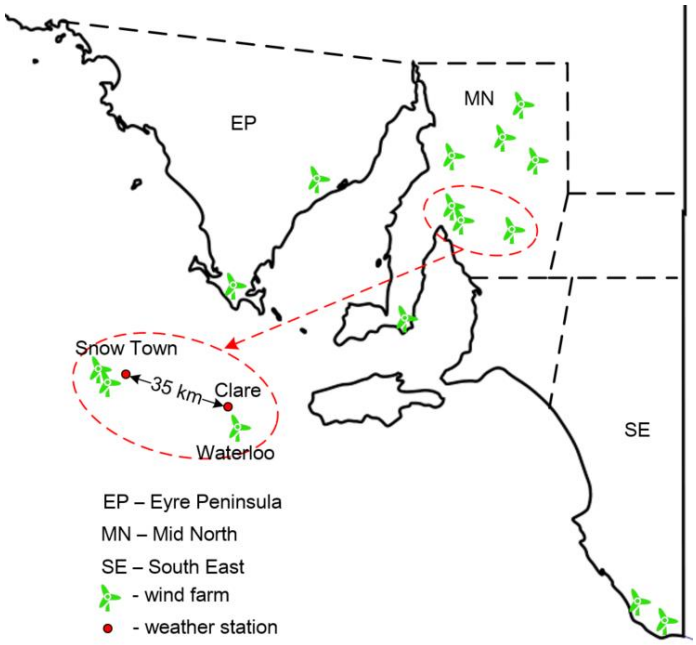

Figure 3.9. Locations of wind farms in SA.

To quantify the dependence between random variables, product moment correlation coefficient (also called linear or Pearson correlation) is first applied which can measure the linear interdependence between the uncertainties [69]. However, if the interdependent random variables include a nonlinear transformation, such as wind speed converted to wind power, then the product moment correlation coefficient method needs to be modified as discussed in [42]. To efficiently solve problems that involve nonlinearity, rank correlation method is introduced [69]. This method defines a product moment correlation of ranks of random variables. The Spearman's rank correlation coefficient (also called the Spearman's rho) is applied when evaluating the degree of monotonic relationship between the random variables [69],

$$\boldsymbol{\rho_{X,Y}} = \frac{\boldsymbol{cov}(\boldsymbol{r_X}, \boldsymbol{r_Y})}{\boldsymbol{\sigma_{r_X}\sigma_{r_Y}}}, \tag{3.12}$$

for random variables X and Y with the corresponding ranks variables r_X and r_Y (X and Y are in ordinal order) respectively, the rank correlation coefficient $\rho_{X,Y}$ equals to the covariance of the rank variables divided by the standard deviations of the rank variables. In here, for the wind speed data of two close weather station, Snow Town and Clare, the $\rho = 0.6031$ which confirms the correlation between the two wind speed data of close weather stations.

The wind source data of Clare are chosen as the input data to construct the correlation

model. Following the steps shown in section 3.2.4.2, before the optimization, for demonstration only, the stochastic characteristics of the estimated wind speed of Snow Town by applying the PCM-Linear and PCM-Quadratic are listed in Table 3.7. From Table 3.7, similar results are obtained compared with Table 3.5, the mean values of the estimated data are close to the sampled data, whereas, the variance is large. Therefore, the same optimization procedure is applied to the modelling. The PCM-Quadratic approximation is used, and the stochastic characteristics of optimized wind speed estimation are listed in Table 3.8.

Table 3.7 Expected value, variance and standard deviation of estimated wind speed in Snow Town.

Snow Town	E	σ^2	σ
Sampled data	5.7527	5.1616	2.2719
PCM-Linear	5.7017	1.5454	1.2431
PCM-Quadratic	5.7366	1.6895	1.2998

The adjusted results of both mean and variance values are almost exactly reproduced comparing with the sampled wind speed data of the Snow Town. Hence, with this polynomial mapping model, instead of representing both wind speed data in Clare and Snow Town as uncertain parameters, if wind speed in Clare is a uncertain parameter x, the wind speed in Snow Town is modelled as an output of the polynomial mapping $f(x)$. In this way, the number of system uncertain parameters are reduced. This method is also suitable for constructing correlation model of other correlated system parameters by repeating the procedure.

Table 3.8 Expected value, variance and standard deviation of estimated wind speed after optimization in Snow Town.

Snow Town	E	σ^2	σ
Sampled data	5.7527	5.1616	2.2719
PCM-Quadratic	5.7576	5.2508	2.2915

This PCM based correlation modelling method has the ability to directly reduce the system dimensions which makes the collocation method more computationally competent in PPF analysis. However, for this method the intermediate optimization is required. Modelling via optimization of each pair of interdependent uncertainties makes the computation burden to be quite heavy if a large number of interdependent uncertainties is considered. Therefore, it is more appropriate to handle small to medium number of interdependent uncertainties using this method.

3.3.2 Large number of interdependent uncertainties modelling

In order to overcome the limitation of the PCM combined with optimization modelling method, the multivariate PDFs method is introduced. To describe the details of this correlation modelling method, two correlated system load data, denoted by M_a and M_b, are chosen from the region MET. The corresponding rank correlation coefficient can be calculated by using (3.12) which equals to 0.6725. Based on the core concept of the PCM, the correlation model between two uncertainties is built. The PCM approach constructs the polynomial mapping. The reference data x_i is input to the mapping function and the output is estimate of correlated data as shown in Eq. (2.13). In this way, the number of system uncertainties is reduced. In the method presented here, M_a is set as reference in creating the correlation model. The details of the correlation modelling procedure are similar as in section 3.2.4.2 with few modifications to adapt the proposed multivariate PDFs method. The steps include:

a. The Box-Cox transformation is applied to the uncertainties data first [70]. This power transformation makes the data more normal distribution like, and it is formed as [70],

$$\boldsymbol{y^{(\lambda)}} = \begin{cases} \dfrac{\boldsymbol{y^{\lambda}-1}}{\boldsymbol{\lambda}}, & \boldsymbol{\lambda \neq 0} \\ \boldsymbol{\log{(y)}}, & \boldsymbol{\lambda = 0} \end{cases}. \tag{3.13}$$

   For uncertainty y, the λ is the transformation parameter. If $y < 0$, then Eq. (3.13) is modified to [70],

$$\boldsymbol{y^{(\lambda)}} = \begin{cases} \dfrac{\boldsymbol{(y+\lambda_2)^{\lambda_1}-1}}{\boldsymbol{\lambda}}, & \boldsymbol{\lambda_1 \neq 0} \\ \boldsymbol{\log(y+\lambda_2)}, & \boldsymbol{\lambda_1 = 0} \end{cases}, \tag{3.14}$$

   where $y + \lambda_2 > 0$. The suitable value of λ can be determined by plotting maximum likelihood function $L_{max}(\lambda)$ against λ. Details of this procedure can be found in [70]. The Box-Cox function of MATLAB [71] is the efficient way to find the proper value of λ. The PDFs of uncertainties M_a and M_b after application of the with Box-Cox transformation are shown in Figure 3.10.

b. Based on the PDF after the Box-Cox transformation of M_a, $f_{BC}(m_a)$, the first set of orthogonal polynomials can be generated by using Gram-Schmidt processing [47],

$$\begin{cases} \boldsymbol{h_0(m_a) = 1,} \\ \boldsymbol{h_1(m_a) = 2.93m_a - 6.24,} \\ \boldsymbol{h_2(m_a) = 5.92{m_a}^2 - 25.24m_a + 26.19,} \\ \boldsymbol{h_3(m_a) = 9.84{m_a}^3 - 62.79{m_a}^2 + 130.05m_a - 87.22,} \\ \boldsymbol{h_4(m_a) = 12.82{m_a}^4 - 105.21{m_a}^3 + 314.08{m_a}^2 - 403.14m_a + 187.08,} \\ \vdots \end{cases} \quad ,(3.15)$$

where the roots of each orthogonal polynomial, $h_0, h_1, \dots, h_{n+1}$, are the collocation points used to generate the mapping polynomial.

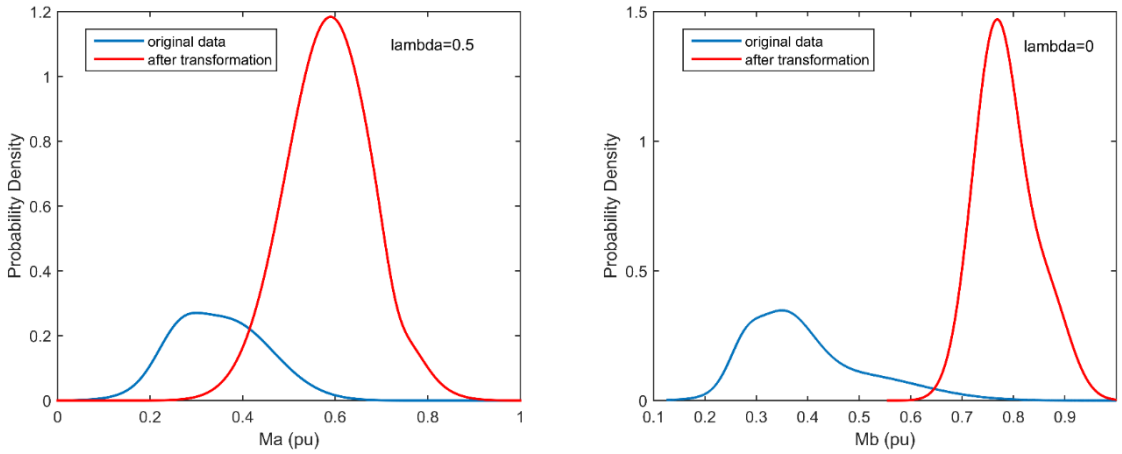

Figure 3.10. PDFs of M_a and M_b (normalized using maximum value of each system load) and with the Box-Cox transformation.

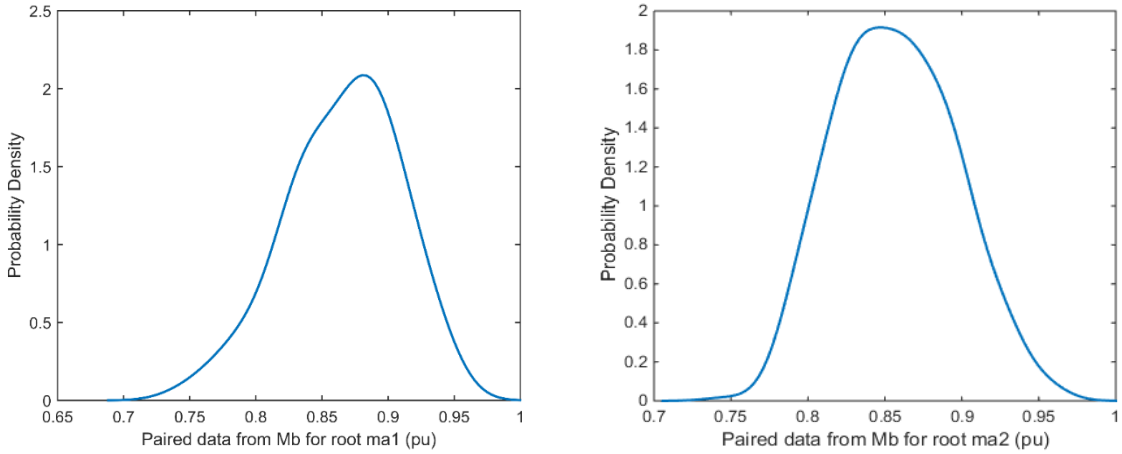

Figure 3.11. PDFs of paired data (normalized using maximum value of each system load).

c. The coefficients of n^{th} order PCM can be computed from the roots of $h_{n+1}(x)$. For each time point, the corresponding uncertainty data M_a and M_b are selected as a pair of input and output respectively. Whereas, for same values of M_a, the M_b is not a

constant value, and then, the calculated roots vary in a range which is similar to the case described in section 3.2.4.2. In here, as an example, details of constructing the linear mapping are demonstrated. For $n = 1$, the roots of $h_2 = (m_a)$ are used to calculate the coefficients of the linear mapping equation, denoted as m_{a1} and m_{a2}. The corresponding PDFs of paired data for each root are shown in Figure 3.11.

d. As shown in Figure 3.11 the paired outputs to the roots are not constant, hence, choosing any single value of those paired uncertainty data M_b as an output to construct the polynomial mapping will cause error. Instead of implementing the optimization step as described in section 3.2.4.2, a multivariate PDFs method is applied where the PDF of accurate polynomial mapping results is represented by the sum of the multivariate PDFs. Instead of selecting one point as an output, based on the PDFs of paired data, several points are chosen as outputs. To preferably capture all the stochastic features of the output, the selecting principles are:
   - The maximum probability density value or adjacent ones should be selected;
   - Based on the probability density, evenly select points through the whole range of the data in one direction (e.g., from lower value of data to higher value of data);
   - Keep the same selecting criterion for all paired data.

   In the procedure, the maximum probability density value is chosen as reference point, and evenly select other points on both sides of the reference point with the same declining ratio to the maximum probability density value. Those selecting criteria are applied to all other paired data. If m (m is odd number) points are selected, then the probability of generated output m_b' is formed as,

$$\boldsymbol{p_{m_b'}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \begin{cases} \dfrac{\boldsymbol{4i}}{\boldsymbol{(m+1)^2}}\boldsymbol{p_i}, & \boldsymbol{i \le \dfrac{m+1}{2}} \\ \dfrac{\boldsymbol{4(m-i+1)}}{\boldsymbol{(m+1)^2}}\boldsymbol{p_i}, & \boldsymbol{i > \dfrac{m+1}{2}} \end{cases} \tag{3.16}$$

   Therefore, the probability of calculated output $p_{m_b'}$ is the sum of m probabilities p_i $(i = 1,2, \dots, m)$ which are obtained by each selected point multiplied by their ratio. For example, if choosing $m = 5$, then the 5 points for the paired data from Figure 3.11 are shown in Figure 3.12. The points P_1, P_2, …, P_5, that are determined by their probabilities which have certain ratio to the maximum probability value, as defined in selection criterion.

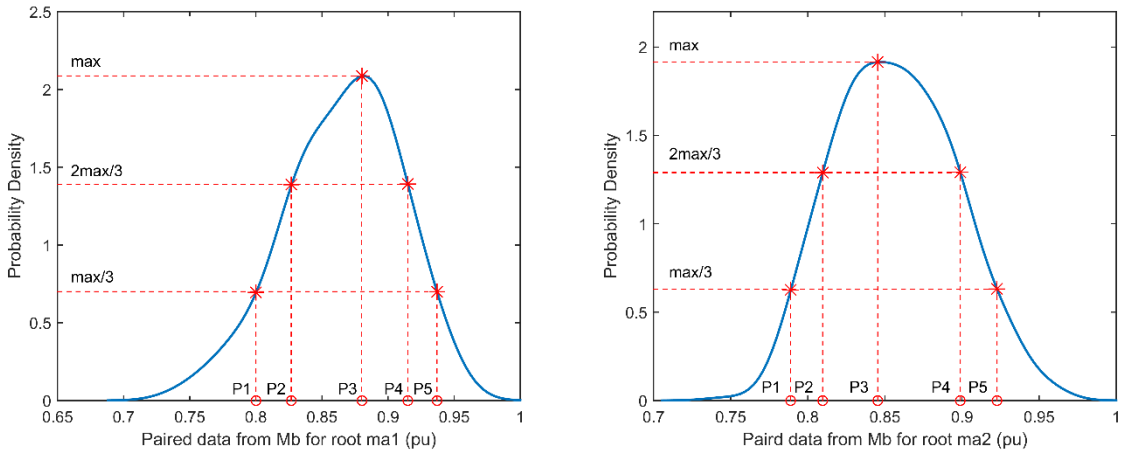

Figure 3.12. Selected point $m = 5$ case based on the PDFs of paired data.

e. For each paired point P_i $(i = 1,2, \dots ,5)$, the linear mapping equation can be generated, denoted as $m'_{b,i} = g_i(m_a)$, based on the historical data of M_a, the estimated $M'_{b,i}$ is calculated. Furthermore, the corresponding $p_i = p_{m'_{b,i}}$ is obtained, and then according to Eq. (3.16), the $p_{m'_b}$ is calculated. The estimated $p_{m'_b}$ after inverse Box-Cox transformation is compared with the original data probability p_{m_b} as shown in Figure 3.13.

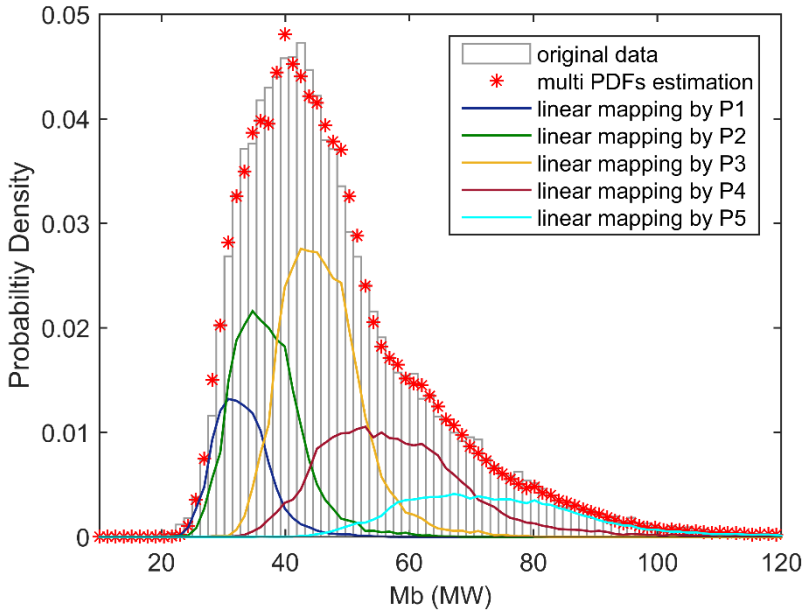

Figure 3.13. Frequency histogram of original data M_b and estimated data by linear mapping.

Based on the multivariate PDFs method, in Figure 3.13, the PDF of estimated data

(red asterisk) is sum of the 5 PDFs (coloured solid line) which are obtained by using the linear mapping equations. Each linear mapping equation is generated by using the corresponding selected points. According to Eq. (3.11), the *FHSI* of Figure 3.13 is 97.01% which shows the estimated data has higher similarity to the original data.

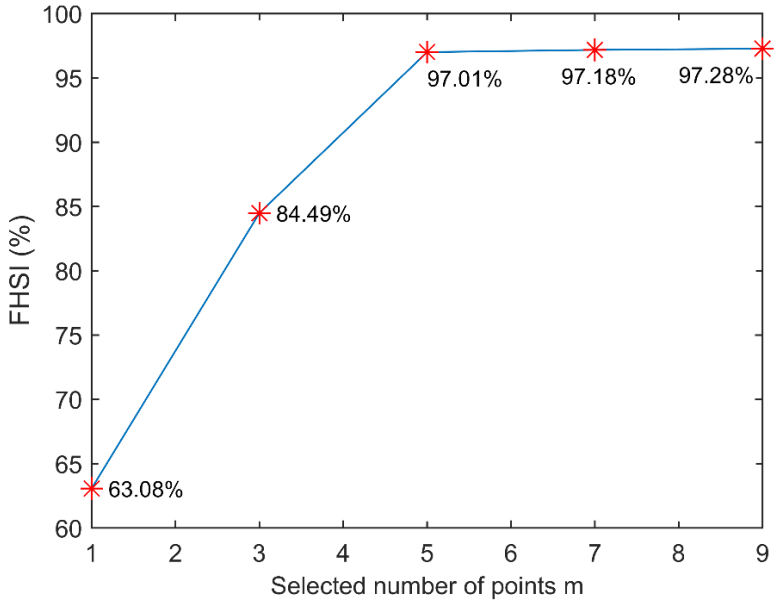

Figure 3.14. The *FHSI* of different number of selected points *m* for linear mapping.

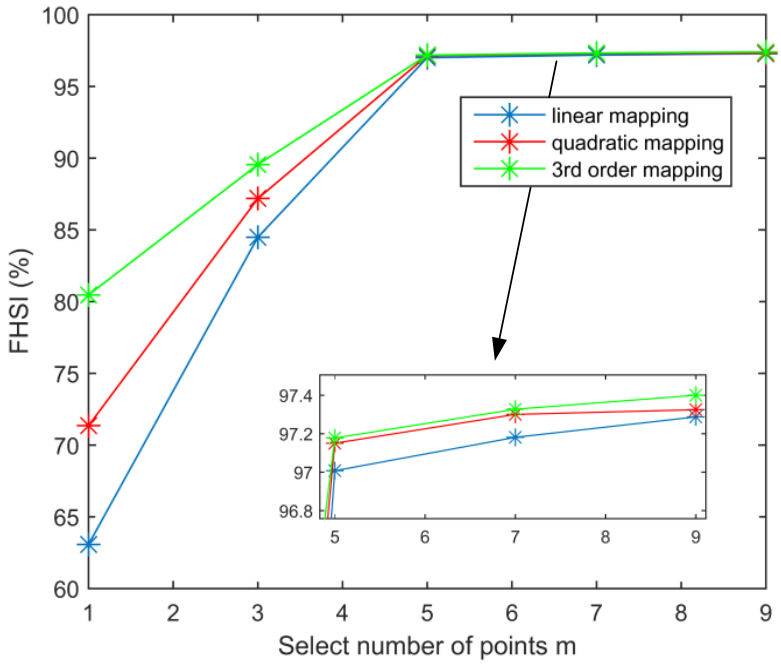

Figure 3.15. The *FHSI* of different number of selected points *m* for higher order mapping.

Repeating steps d and e, the *FHSI* with different numbers m can be calculated. The Figure 3.14 shows the *FHSI* of $m = 1, 3, \dots, 9$ where the *FHSI* is increasing with higher number m. However, when $m \geq 5$, for further increase of m, change in of *FHSI* is negligible.

For higher order mapping, steps c to e are repeated. The values of *FHSI* of $m = 1, 3, \dots, 9$ for each order are shown in Figure 3.15. It shows that by increasing the order of mapping higher values of *FHSI* can be achieved. This means higher accuracy of estimated results can be achieved. This improvement is more effective when dealing with lower value of m, whereas, when the number of points $m \geq 5$, further increasing the number of points will not significantly give higher values of *FHSI*. Therefore, with considering the computation cost, the 2nd order mapping with $m = 7$ is applied to other correlated data and the corresponding *FHSI* is calculated.

3.4 Conclusions

This chapter described the methodology used to construct the power system uncertainties models as well as the uncertainties correlation models. The PPF analysis is based on accurate probabilistic models of system uncertainties, and in realistic practical cases, most of the system uncertainties sampled data are not following any known probability distribution. Hence the GMM is applied to the sampled data which is able to construct the probabilistic model with high accuracy. Meanwhile, instead of using typical theoretical wind power generation model, the PCM is introduced to model the realistic relationship between wind speed and wind power. Moreover, combined with fuzzy logic optimisation, the estimated wind power by the PCM has high accuracy comparing with the sampled data.

The interdependent relationships between uncertainties are considered in this chapter. In order to adequately handle the correlated uncertainties, two correlation modelling methods are introduced to model the dependences of the correlated uncertainties. They are: the PCM-based method (including optimization) for modelling of small number of interdependent uncertainties; the multivariate PDFs method for modelling large number of interdependent uncertainties. Both of these modelling methods aim to reduce dimension of the problem, hence, to improve the computation capability of collocation method in dealing with PPF analysis.

Chapter 4

PROBABILISTIC POWER FLOW COMPUTATION USING PROBABILISTIC COLLOCATION METHOD

THIS chapter describes how to use the Probabilistic Collocation Method to solve power system PPF problem. Instead of using deterministic power flow model to solve for large number of simulation points as in Monte Carlo approach, the PCM has the ability to properly select small number of collocation points to generate a polynomial mapping model which is computationally efficient compared to non-linear power flow computation. When this surrogate polynomial model is constructed, all randomly generated simulation points, as in Monte Carlo method, are applied to this model. Simplified SA power system network model and corresponding historical data are used in demonstrating the proposed novel approach. The proposed technique incorporates correlation modelling method to improve accuracy of the PCM in PPF computation. Both single output and multiple output cases are tested and results were presented in this chapter. The results show the efficiency and the accuracy of the PCM in handling the PPF analysis.

4.1 Introduction

In SA, the renewable generation provided up to 49% of power supply to the consumer [3]. Due to the highly uncertain and variable features of those renewable resource, the PPF analysis method, such as MCS method, is developed to obtain more realistic results of the power flow computation. However, the computation burden of MCS is quite high, hence, to overcome this limitation, the PCM is proposed in our study to handle the PPF computation.

According to [25]-[33], the PCM obviously shows the ability of not just solving probabilistic problems but also handling PPF computation. The PCM can provide the statistical moments or complete random samples of desired outputs with high accuracy and incredible computation efficiency. Whereas, as described in section 2.1.4, when using the PCM to handle multi uncertain parameters case, the computation cost is exponentially increasing when the number of system uncertain parameters increases. Hence, in our study, the aim is to reduce the number of the system uncertainties, therefore it makes the PCM more competent in solving the PPF problem.

The approach of improving the capability of the PCM to handle the PPF computation is achieved by the correlation modelling method detailed in section 3.3.1. This method is based on the PCM theory as well. Base on this method, a polynomial mapping equation is constructed to express the relationship between those uncertainties and PPF problem solutions. One of the correlated uncertainties can be expressed by the other using a mapping equation, therefore, the number of the total uncertainties is reduced. Base on this approach, the case study demonstrates the advantage of the PCM as far as possible. The PCM combined with the proposed correlation modelling method to handle uncertainties in PPF analysis is illustrated in this chapter.

The remainder of this chapter is organised as follows, first the information of the testing system is described in Section 4.2. Based on this testing system, Section 4.3 details simulation method applied in the PPF computation. Then, the simulation results are obtained and discussed in Section 4.4. The summary of our proposed PPF simulation method is presented in Section 4.5.

4.2 Aggregated SA power system model

As described in section 3.2, the SA power system network can be divided into 6 regions.

In this chapter, based on the acquired historical regional data, an aggregated load flow model of SA is constructed to illustrate the computation efficiency and accuracy of using the PCM in PPF computation. The aggregated 6 buses power system model is shown in Figure 4.1,

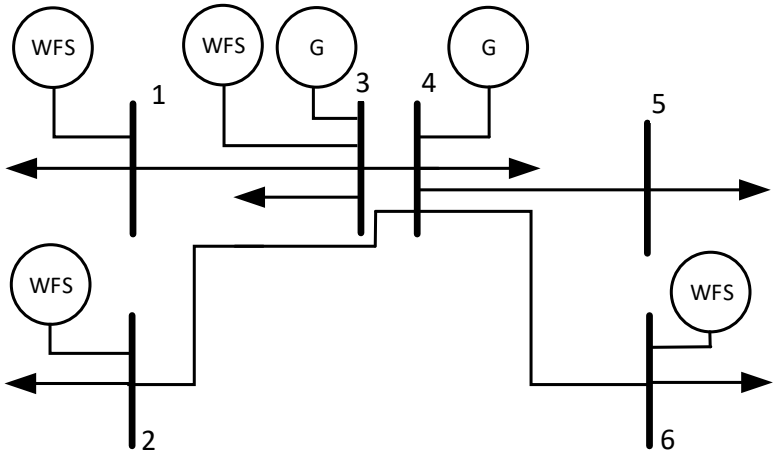

Figure 4.1. The aggregated SA power system.

Each bus in Figure 4.1 represents one region of SA, they are: bus 1 - YP, bus 2 - EP, bus 3 - MN, bus 4 - MET, bus 5 - RL and bus 6 - SE. Those 6 buses are connected with 270 kV transmission lines. In here, for demonstration purpose, 7 uncertain parameters are assumed in this aggregated system to compare the PCM with traditional MCS method. The uncertainties for each bus are summarized in Table 4.1.

Table 4.1 Type and number of uncertainties of 6 buses SA aggregated power network.

Regions	Bus number	Demand	Wind farms	Total
YP	Bus 1		1	1
EP	Bus 2		1	1
MN	Bus 3		2	2
MET	Bus 4	1		1
RL	Bus 5	1		1
SE	Bus 6		1	1
Total number of uncertainties				7

All the uncertainties shown in Table 4.1 are modelled by the method proposed in section 3.2. The wind speed correlation in region MN, and correlation between the region MN and YP, are considered. With a small number of interdependent uncertainties case, the correlation modelling method developed in section 3.3.1 is applied to model wind speeds from neighbouring wind farms in region MN, and the correlated wind speeds between two adjacent regions. Meanwhile, because only the wind power data of region

YP is considered, so during the simulation, the bus 1 and bus 3 are combined into one bus. The SA power network parameters can be found in [3].

4.3 Simulation methods

The MATLAB-based DPF simulation tool MATPOWER [72] is called to solve power flow problem for each sample of uncertainties data. The simulation includes MCS and PCM, the details of each simulation are:

a. MCS with Copula

The MCS is a straightforward PPF simulation method with high accuracy and normally used as benchmark to verify the effectiveness of other methods. Rather than using whole set of the original data, based on the PDFs of the system uncertainties, the MCS generates moderate number of simulation points to obtain high accurate estimation of results. The commonly used Copula method is adopted to formalize dependences between correlated uncertainties.

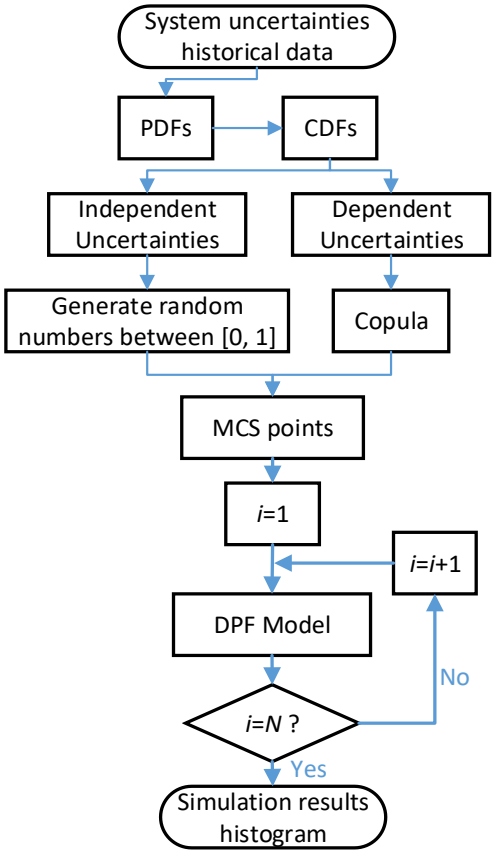

Figure 4.2. Block diagram of the MCS with Copula.

The processing block diagram is shown in Figure 4.2. By using the system historical

data, the PDFs of the system uncertainties are first obtained, and then transferred to the CDFs which are in the range between 0 and 1. Further processing steps are split into two branches:

1) for independent uncertainties, by randomly generating numbers between 0 and 1, the MCS points are created through solving CDFs^{-1} with those random numbers;

2) for dependent uncertainties, through the Copula function, the MCS points are generated.

Moderate number of simulation points $N = 40000$ (sampled from system uncertainty space 40000 times) are generated and used as input to DPF model. The DPF model is called to calculate the desired outputs until all simulation points have been applied. Whereupon the histograms of the selected outputs are generated.

b. PCM with proposed correlation modelling method

Instead of calling DPF model for every simulation point, the PCM technique with properly select small number of collocation points is used to generate polynomial mapping model. Then all the simulation points are applied to the polynomial mapping model. Since computation time of this model compared to DPF is negligible, incredible time saving in the simulation is achieved.

The processing block diagram is shown in Figure 4.3. The basic simulation steps of the PCM is presented, and it includes 2 main steps:

1) The step 1 is based on the proposed collocation modelling method; the correlated data are modelled using a polynomial mapping, hence reducing the number of the system uncertainties. Further on, based on the PDF of reduced uncertainty space, another mapping function is formulated to represent input-output relation of PPF problem. This procedure is also based on the orthogonal polynomials and Gram-Schmidt process orthogonalization, as describe before. This mapping model (i.e. surrogate function) has shorter computation time compared to the original DPF model.

2) The step 2 is based on the MCS. Instead of using the DPF model, the surrogate model from step 1 is applied. Sample points are generated from the reduced system uncertainty space, and number of $N = 40000$ samples (sampled from

system uncertainty space 40000 times) is used for computing with the MCS. All those sample points are applied to polynomial mapping model to compute desired outputs.

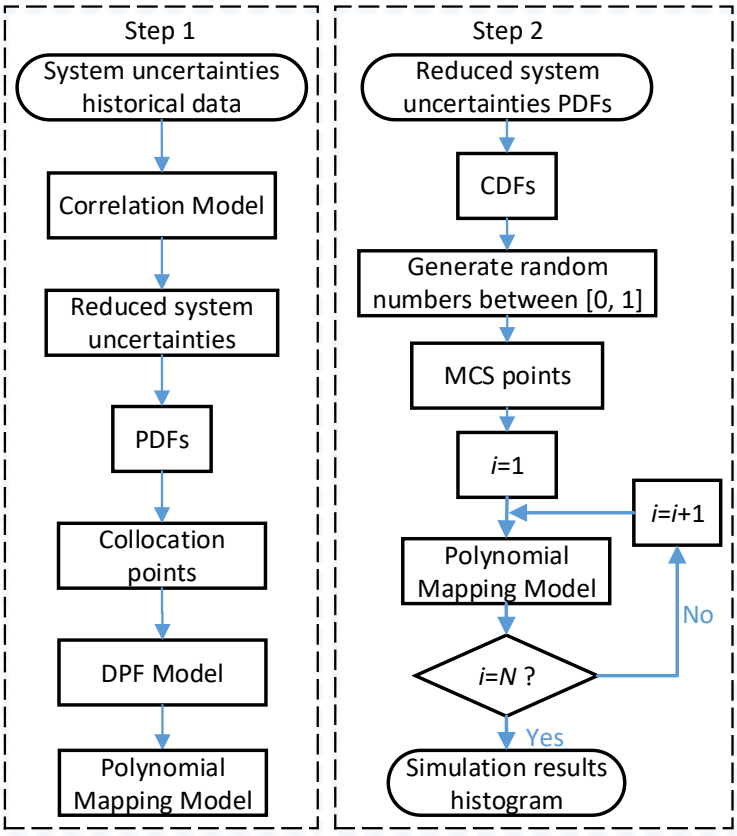

Figure 4.3. Block diagram of the PCM with proposed correlation modelling method.

4.4 Simulation results and discussion

In the simulation example a quadratic PCM is used in constructing the surrogate model. A single uncertain parameter requires $n_i = 3$ simulation points to evaluate the coefficients of the mapping polynomial. Total required simulation points for multi uncertain parameters PCM equals to $\prod_{i=1}^{m} n_i$. Without the correlation model, when $m = 7$, the total number of collocation points is $3^7 = 2187$. With the correlation model, the total number of uncertainties is reduced from $m = 7$ to $m = 5$, and then the total number of simulation points is only $3^5 = 243$. The collocation points required to construct the mapping polynomial is determined by two factors which are the order of the PCM and the number of uncertainties. Therefore, when the order of the PCM is fixed, the collocation points are exponentially increasing with higher number of system

uncertainties, hence, reducing the number of system uncertainties will significantly reduce the number of simulation points needed in constructing the mapping equation. Total number of collocation points when considering correlation model is nearly reduced to one tenth compared to the case when correlation model is not used. In other words, much less computation effort is required to generate the mapping equation when correlations between uncertain parameters are considered.

The MCS method is able to generate all system outputs after simulation, although not every output is of interest to the user. The PCM is also capable to deal with multiple outputs, whereas more computational effort needs to be involved. Hence, in this section single output and multiple outputs cases are demonstrated to compare those two PPF simulation methods. In order to compare the similarity of two histograms, the *FHSI* as expressed in Eq. (3.11) is used to indicate accuracy of the histogram generated by PCM comparing with the one generated by MCS.

4.4.1 PCM with single output

In this section, to compare the two PPF simulation methods, the active power flow P_4 which is power flow from bus 4 in Figure 4.1 is used. Following the block diagram shown in the section 4.3, the histograms of the desired output P_4 calculated with different simulation methods are shown in the Figure 4.4. The PCM based simulation results curve is overlapping with the one obtained by MCS. Using Eq. (3.11), Index $FHSI = 97.68\%$ indicates the similarity of the two histograms is high, in other words, the simulation result based on PCM is matching well with the one from the MCS. Furthermore, this confirms that with only few collocation points the constructed polynomial mapping model is computationally competent in mapping the inputs to desired output with similar accuracy as MCS.

The total computation time of the MCS is composed by $N_{MCS} \times t_{DPF}$, where N_{MCS} is the number of MCS points and t_{DPF} is the simulation time for each input point which is processed by DPF model. While the computation time of the PCM includes two parts:

1) In the step1, time of constructing polynomial mapping model equals $N_{collocation} \times t_{DPF} + t_{other}$, where $N_{collocation} \times t_{DPF}$ is time of using the collocation input points to compute corresponding outputs upon DPF model, and t_{other} is the time to generate the polynomial mapping model based on

collocation input points and corresponding outputs;

2) step 2, the time of creating the frequency histogram of desired output based on the polynomial model which equals to $N_{MCS} \times t_{polynomial}$, where $t_{polynomial}$ is the simulation time for each input point according to the generated polynomial mapping model.

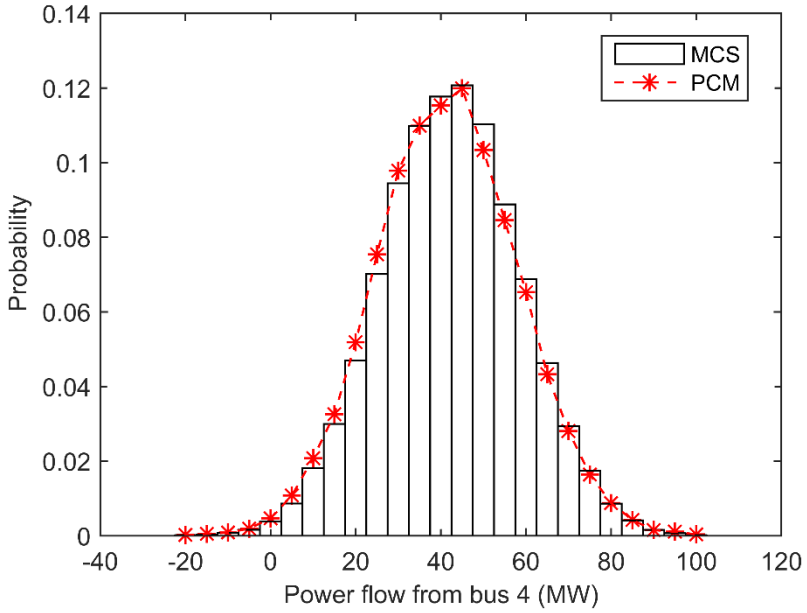

Figure 4.4. Histograms of P_4 using MCS and PCM.

The values of all computation times are summarised in Table 4.2.

Table 4.2 Computation time of 6 buses SA power network with single output.

		MCS	PCM
N_{MCS}		40000	40000
$N_{collocation}$		-	243
Step 1	t_{DPF} (s)	0.0158	0.0158
	t_{other} (s)	-	0.001
Step 2	$t_{polynomial}$ (s)	-	1.22×10^{-5}
t_{Total} (s)		631	4.33

The The values of all computation times are summarised in Table 4.2.

Table 4.2 shows the total simulation time of PCM is less than 1% of the MCS which means the PCM has striking computation efficiency without much trading off the accuracy comparing with MCS. The simulation time of PCM for each step are about 3.84s and 0.49s respectively. So, step 1 occupies about 90% of the total simulation time which is due to high computation cost of DPF simulation, and this also further verifies the

computation efficiency of PCM by properly using relatively small amount of simulation points $N_{collocation}$. Because $N_{collocation}$ is determined by the number of uncertainties in the system, hence, with higher number of uncertainties involved into the PCM, $N_{collocation}$ will exponentially increase, therefore more computation effort is needed at step 1 and more complicated polynomial mapping model will be created. This indicates that the computation efficiency of the PCM is decreasing while increasing the number of uncertainties. With the same number of N_{MCS}, the simulation time of step 2 is determined by $t_{polynomial}$. In here, the $t_{polynomial}$ is about 10^3 smaller than t_{DPF}, hence, for the same number of inputs N_{MCS}, the time saving is incredible. Because more complicated polynomial mapping model will increase $t_{polynomial}$, therefore, increasing the number of uncertainties will increase the simulation time of step 2 as well. However, if increasing the number N_{MCS} to obtain more accurate simulation results, the time saving will be significantly increased. With 7 uncertainties and single output case, the PCM shows incredible time savings compared with the MCS method. Although the advantage of PCM may not be significant for large dimensions of uncertainty space, it is still a preferable method in the PPF computation.

4.4.2 PCM with multiple outputs

In this section, the active power flows, P_2, P_3, P_4, P_5 and P_6, as outputs, are used to compare both proposed methods. These power flows are power flows from each bus shown in Figure 4.1. The Table 4.3 shows the computation times of MCS and PCM. With the same number of MCS points and uncertainties, the time of MCS method is almost the same as single output case as shown in The values of all computation times are summarised in Table 4.2.

Table 4.2 which confirms the numbers of desired outputs will not affect the computation time of MCS method. The time spent on step 1 of PCM is also nearly the same as single output case, whereas, the $t_{polynomial}$ of multiple outputs case is about 2 times larger compared to the single output case. Because the PCM generates different polynomial mapping models for each desired output, then for step 2, the interpolation for each output must be running in sequence, hence, higher computation time is needed to solve for more outputs. In here, with 7 uncertainties, the $t_{polynomial}$ is still far less to t_{DPF}, so the simulation time of step 2 is barely affecting the total simulation time even with higher number of desired outputs. For large number of uncertainty dimensions,

$t_{polynomial}$ will become larger, and then for large number of desired outputs, the further increase in $t_{polynomial}$ may not be significantly smaller than t_{DPF} which means the time saving may not be as remarkable as for lower dimensions. In this case, the computation efficiency can be improved by dividing the simulation into several groups (with different sub-sets of outputs) and running parallel processing on few computers simultaneously.

Table 4.3 Computation time of 6 buses SA power network with multiple outputs.

		MCS	PCM
N_{MCS}		40000	40000
$N_{collocation}$		-	243
Step 1	t_{DPF} (s)	0.0158	0.0158
	t_{other} (s)	-	0.001
Step 2	$t_{polynomial}$ (s)	-	2.21×10^{-5}
t_{Total} (s)		631	4.73

The *FHSI*s of the case with multiple outputs is shown in Table 4.4. The values are varying around 96% to 99% which indicates that the PCM has almost the same accuracy as MCS in handling the PPF computation, whereas PCM has much smaller computation cost than the MCS. Moreover, to obtain higher accuracy of desired outputs, higher number of MCS sampling points can be applied to the system, and this will further increase the simulation time difference between the MCS and PCM due to relatively smaller $t_{polynomial}$ compared with t_{DPF}.

Table 4.4 FHSIs of PCM with multiple outputs.

FHSI	PCM
Power flow P_2	96.83%
Power flow P_3	99.36%
Power flow P_4	97.68%
Power flow P_5	99.16%
Power flow P_6	99.01%

4.4.3 Factors affecting the computation time of the PCM

According to the block diagram shown in Figure 4.3, the computation time of the PCM is split into two steps and shown in Table 4.2 and 4.3. In step 1, a polynomial mapping model is constructed by using the collocation points, and in step 2, the MCS random simulation points are applied to this polynomial mapping model. It is not necessary to repeat step 1 to construct the polynomial mapping model unless the operation conditions of the system significantly change. Hence, the polynomial mapping model generated by PCM has incredible computation efficiency which indicates that in practical

situations it is possible to deal with real-time power system operation and guide operators in control tasks and scheduling. The factors affecting the time saving in using the PCM compared to MCS are summarized in Table 4.5.

Table 4.5 Factors affecting the time saving in using the PCM compared to MCS.

Factors	Time saving
	increase ↑, decrease ↓
Number of uncertainties ↑	↓
Number of outputs ↑	↓
Number of MCS points ↑	↑
Power system size ↑	↑

In order to construct the polynomial mapping model, the collocation points need to be obtained first. Whereas the collocation points are exponentially increasing while more system uncertainties are involved, which means more computation effort is needed to generate the polynomial mapping model. Hence, as shown in Table 4.5, the PCM has higher computational efficiency with lower number of uncertainties. Because the PCM generates different polynomial mapping models for each desired output, to obtain all the histograms of the desired outputs more computation time is required. Hence the total computation time is increasing when solving multiple outputs cases. To achieve high-accuracy simulation results, for a specified number of system uncertainties, more MCS sample points may be adopted, as the simulation time of using the polynomial mapping model is much smaller than applying the DPF model, therefore the total time saving of using the PCM compared to MCS increases. In here, for demonstration purpose, the testing system is relatively small with small amount of system uncertainties, whereas the realistic power grid of SA will be larger than this with more system uncertainties, and more correlated data are involved into the model. Because for larger size of power system, the t_{DPF} will dramatically increase, however, $N_{collocation}$ stays relatively small comparing to N_{MCS}, and $t_{polynomial}$ is also much smaller than t_{DPF}, hence, the simulation time of MCS will increase faster than PCM, and this will lead to the obvious time saving when using PCM for increasing size of power system network. Although the total computation cost of the PCM is not as low as for the small size power system, with the proposed correlation modelling (as shown in Section 3.3.1), the computation efficiency will be improved by reducing the number of system uncertainties.

4.5 Conclusions

In this chapter, a simplified SA power transmission network with historical system data is applied to demonstrate the application of the PCM in handling the PPF computation. To verify the computation efficiency of the PCM, the MCS method as a benchmark is adopt to the testing system as well. Because the computation efficiency of PCM is highly associated with the system dimension, the proposed correlation modelling method (as described in Section 3.3.1) is applied to model the interdependent system uncertainties and reduce the number of system uncertainties. This approach makes the PCM more computationally competitive in solving the PPF problem. Although a relatively small aggregated SA load flow model is used in this case studies, the simulation results clearly show the striking time savings and computation accuracy of the PCM in handling both single output and multi outputs cases compared to the MCS. With the proposed uncertainties correlation modelling method, the number of system uncertainties is reduced which improves the effectiveness of the PCM in dealing with higher dimension system simulation. Due to the high accuracy of the PCM, this method is adequate to handle the power system simulations. Moreover, when the polynomial mapping model created by the PCM is set up, the incredible time savings of this proposed model compared to the DPF model gives the chance to run the probabilistic power system simulations in real-time and guide operator in daily power system operation and scheduling.

Chapter 5

PROBABILISTIC POWER FLOW COMPUTATION USING SPARSE GRID INTERPOLATION METHOD

THIS chapter describes how to use Sparse Grid Interpolation (SGI) method to solve power system PPF problem. The computation mechanism of SGI method is similar to the PCM, where instead of using deterministic power flow model, a surrogate model is constructed to run every Monte Carlo simulation point. Comparing with the DPF model, this surrogate model has striking high computation efficiency. Only small number of collocation points are required to generate this surrogate model. By using the simplified SA power system network and historical data, the case with independent uncertainties is presented first and then the correlation modelling method is incorporated in the final case study. Both of those case studies verify the computation efficiency and accuracy of the SGI method in solving the PPF problems.

5.1 Introduction

The PCM presented in Chapter 4 shows the computation efficiency of handling PPF computation. The PCM is a simple and cost-effective simulation method, but this method can be affected by increase in input-space (i.e. uncertain parameter space) dimension. To overcome this limitation, the SGI method, or the optimized DASGI method, is introduced. The capability of the SGI to solve high dimensional PPF problems is verified in [31], [34]-[40]. Combined with the Copula theory [37], the SGI can accurately solve the interdependent uncertainties cases. According to the case studies in the literature, the SGI is potentially superior in dealing with the PPF computation for either independent uncertainties or correlated uncertainties.

In our study, the aim is to use the SGI method to solve the practical PPF problems of the SA power system. Hence, the same aggregated SA power system network as shown in Figure 4.1 is applied in testing the SGI method. Since the SGI method is more computationally competent to solve high dimensional cases, a larger number of uncertainties will be considered in the SA network model. Meanwhile, both independent uncertainties and correlated uncertainties cases of power system PPF analysis are demonstrated. Although comparing with the PCM the SGI is more adequate to handle high dimension PPF computation, the computation cost still increases when using the SGI with higher system dimensions. Therefore, for correlated cases, a proper correlation modelling method is used as described in section 3.3.2. According to this correlation modelling method, the total number of system uncertainties is reduced, therefore, further increase the computation efficiency of the SGI method.

First, for independent uncertainties case, the SGI method is applied, and the simulation results are compared with the MCS method. For correlated uncertainties case, due to more uncertainties involved, the optimized DASGI method which includes the correlation modelling method is applied, and the results are compared with both the MCS with Copula and DASGI with Copula. Both of those simulation results confirm the computational efficiency and accuracy of the SGI method and the correlation modelling method.

The remainder of this chapter is organised as follows, Sections 5.2 details the information of the testing system. The simulation methods are described in Section 5.3 in terms of different case study scenarios. The simulation results are obtained and

compared in Section 5.4. Discussion and conclusions are presented in Section 5.5.

5.2 Aggregated SA power system model

In this section, based on the same aggregated SA power system network as shown in Figure 4.1, the aggregated SA power system network with large number of uncertainties is shown in Figure 5.1. Type and number of uncertain variables for each bus are listed in Table 5.1 and 5.2 for independent case and correlated case respectively.

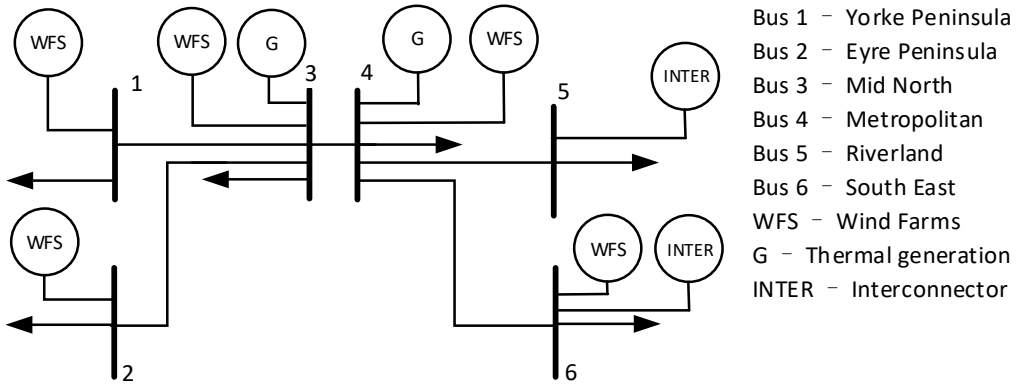

Figure 5.1. The aggregated SA power system with large number of uncertainties.

Table 5.1 Number of uncertainties of 6 buses SA aggregated power network; uncertainties are independent.

Regions	Bus number	Demand	Wind farms	Inter connector	Total
EP	Bus 1	1	2		3
YP	Bus 2	1	1		2
MN	Bus 3	2	9		11
MET	Bus 4	3	1		4
RL	Bus 5	1			1
SE	Bus 6	1	6	1	8
Total number of uncertainties in the system					29

In Table 5.2 another case study is presented where the number of the wind farms is kept the same as in the case shown in Table 5.1. Large number of the uncertain demands in region MN and MET are involved in the second case, therefore increasing the number of the uncertainties. Besides this, the inter connector in region RL is included separately from demand data. The study case in Table 5.2 includes uncertain variables that are correlated.

All the uncertainties shown in Table 5.1 and 5.2 are modelled by the method detailed in section 3.2. For the interdependent uncertainties case, the correlation modelling method

demonstrated in section 3.3.2 is applied to correlated data. All other SA power network parameters can be found in [3].

Table 5.2 Number of uncertainties of 6 buses SA aggregated power network; uncertainties are correlated.

Regions	Bus number	Demand	Wind farms	Inter connector	Total
EP	Bus 1	1	2		3
YP	Bus 2	1	1		2
MN	Bus 3	11	9		20
MET	Bus 4	15	1		16
RL	Bus 5	1		1	2
SE	Bus 6	1	6	1	8
Total number of uncertainties in the system					51

5.3 Simulation methods

The DPF simulation tool is MATPOWER [72] as before. It is called to solve power flow problem for each sample of input data. The toolbox of SGI [73] is used to conduct the SGI simulation. For independent uncertainties case, basic SGI method is applied and compared with the MCS. The simulation results obtained when using either of the methods are also compared with the results from the reference simulation (computed with original SA power network historical data), and in this way the accuracy of those PPF methods is verified. And then those PPF methods are used to evaluate the efficiency and accuracy of our proposed correlation modelling method in correlated uncertainties case. For the second case that includes correlated uncertainties, due to higher number of uncertainties involved, more efficient SGI method described in section 2.2.3 and named DASGI is applied. In the second case that includes correlation, the Copula method is applied to construct correlation model and combined with MCS and DASGI; the presented multivariate PDFs modelling method in section 3.3.2 is applied in DASGI to construct the surrogate model. The accuracy of the proposed DASGI approach with correlation modelling method will be evaluated by comparing the simulation results with MCS and DASGI using Copula.

5.3.1 Independent uncertainties case

The simulation for independent uncertainties case includes reference data simulation, MCS and SGI. Details of each simulation are:

a. Reference data simulation

The reference data simulation is based on the 5 years' historical data set with half hourly resolution, so each uncertainty has 48 data points for one day, and then the total simulation points of 5 years are around $N = 90000$. The simulation based on this reference data is used to compare the accuracy of the PPF methods, either MCS or SGI. The Figure 5.2 shows the block diagram of reference data simulation, where the simulation inputs are system uncertainties historical data listed in Table 5.1 and each uncertainty has identical number of points N. For each point the DPF model is called to obtain corresponding desired simulation outputs such as power flow between buses, bus voltages and so on. Repeating the process until all those data points are applied to DPF model, and then generating the histogram of the desired outputs.

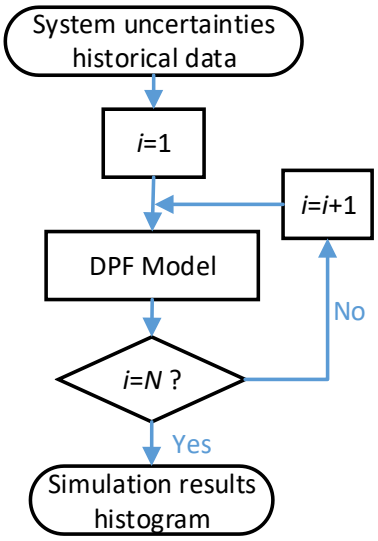

Figure 5.2. Block diagram of reference data simulation.

b. MCS

MCS is a straightforward PPF simulation method with high accuracy. Rather than using whole set of the original data, MCS generates relatively small amount of simulation points and obtains high accuracy results compared with the reference data. The processing block diagram is shown in Figure 5.3.

In MCS method PPF input samples are generated by passing the uniformly distributed random numbers between 0 and 1 through CDFs^{-1} function. Then the DPF model is called to calculate the desired outputs until all simulation points have been applied. Then the histograms of desired outputs are generated.

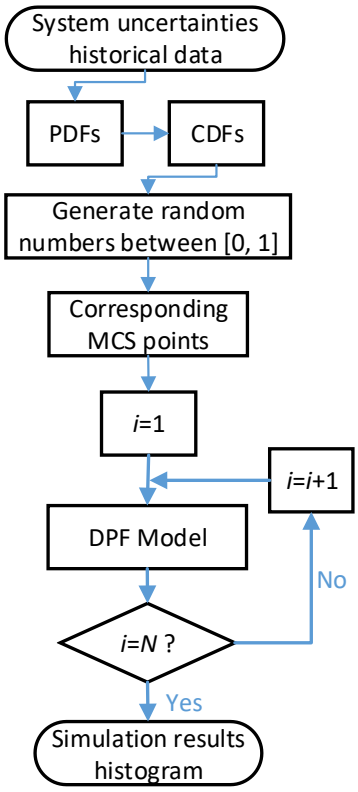

Figure 5.3. Block diagram of MCS.

c. SGI

Instead of calling DPF model for every simulation point, the SGI uses small number of collocation points to generate surrogate model of the DPF model. Then all the simulation points are applied to this surrogate model. Hence, strikingly reduced simulation time is achieved. The block diagram of the SGI is shown in Figure 5.4.

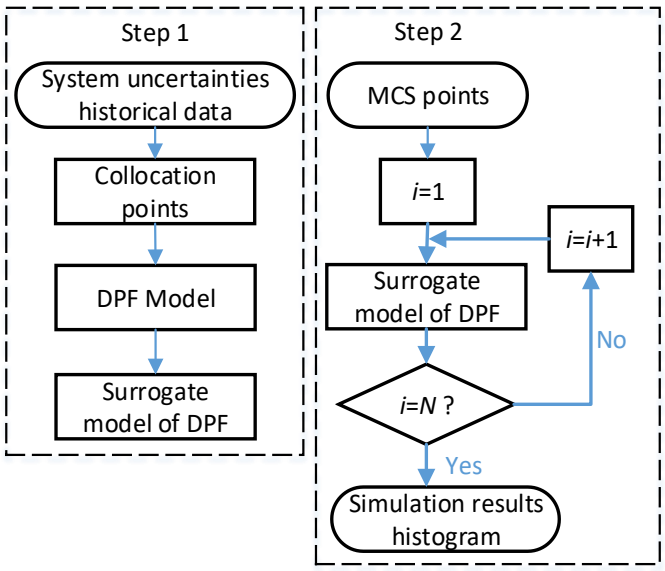

Figure 5.4. Block diagram of SGI.

As shown in Figure 5.4, running the SGI simulation includes 2 main steps:

1) In the first step, based on Smolyak algorithm, a small subset of collocation points from system uncertainties historical data are generated to construct the surrogate model of the DPF; this mapping model has shorter computation time when comparing with the original DPF model;

2) In the second step, all the MCS sample points are applied to the surrogate model to calculate desired outputs.

5.3.2 Correlated uncertainties case

The simulation for correlated uncertainties case includes the MCS with Copula, the DASGI with Copula and the DASGI with proposed multivariate PDFs modelling method in section 3.3.2. Details of each simulation are:

a. MCS with Copula

The procedures of MCS combined with Copula simulation method is the same as the one presented to evaluate the computation efficiency of the PCM. The details can be found in section 4.3 point a. The block diagram of the MCS with Copula is shown in Figure 4.2.

b. DASGI with Copula

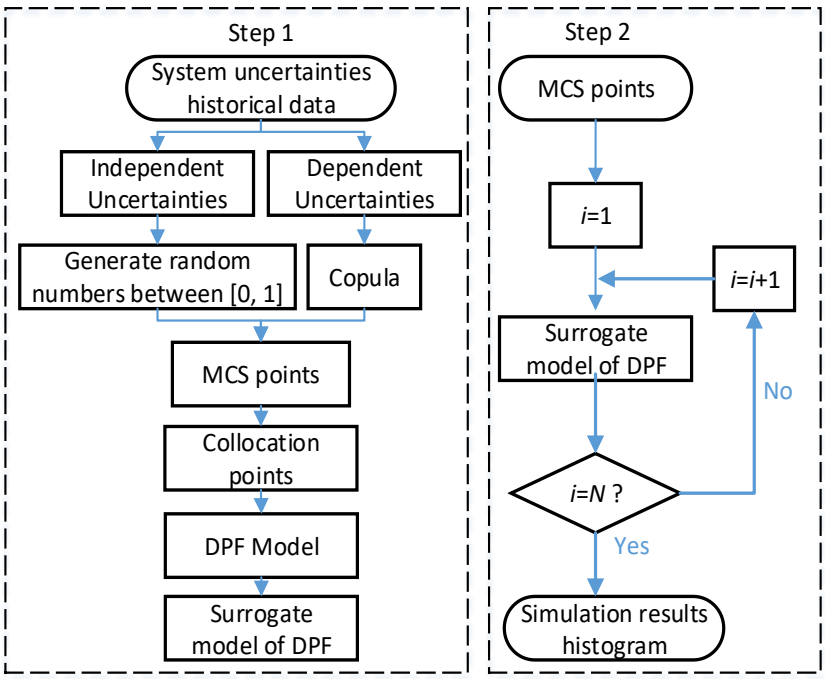

Figure 5.5. Block diagram of DASGI with Copula.

Similar as the basic SGI method, based on a relatively small number of collocation points, a surrogate model that replace DPF is developed. The block diagram of DASGI with Copula is shown in Figure 5.5.

The DASGI method includes 2 steps:

1) In the first step, the simulation points are generated as with MCS method with Copula, and then based on the Smolyak algorithm, a small number of collocation points are used to construct a surrogate model of DPF;

2) In the second step, instead of using DPF model, MCS sample points are applied to the surrogate model to compute desired outputs. The theorical details of DASGI simulation method can be found in [37].

c. DASGI with multivariate PDFs

Comparing with Copula method, in addition to accurately modelling the dependence between uncertainties, the proposed multivariate PDFs method, described in section 3.3.2, is able to reduce the number of system uncertainties. Therefore, this method combining with the DASGI can further increase the computation efficiency of DASGI method. The block diagram of DASGI with multivariate PDFs is shown in Figure 5.6.

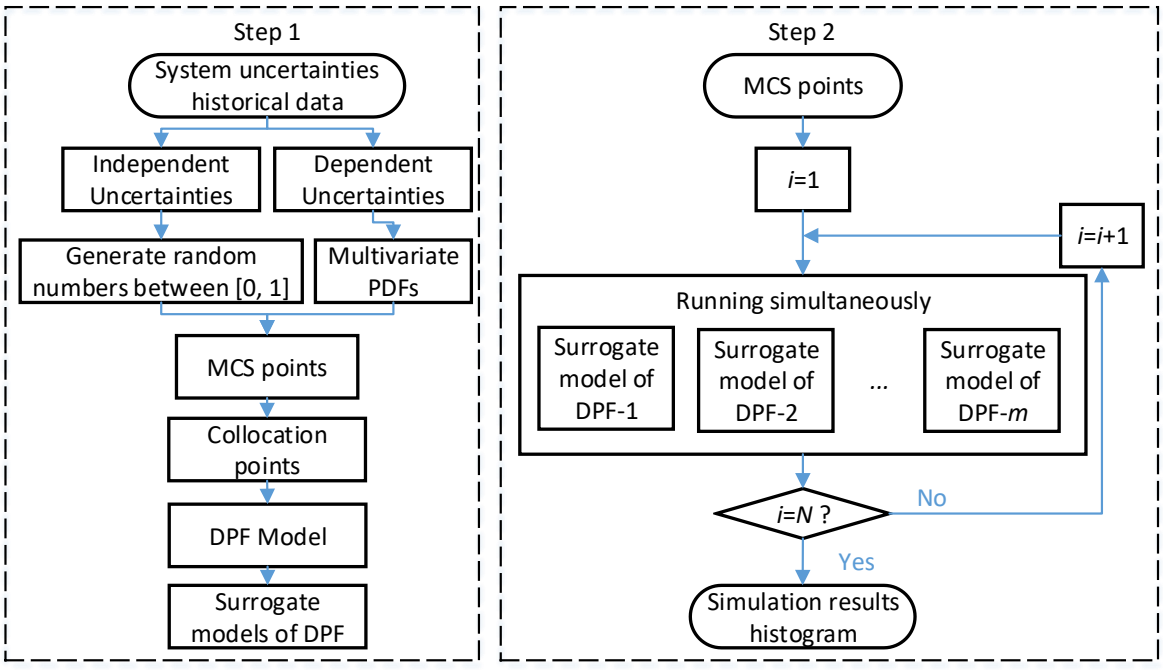

Figure 5.6. Block diagram of DASGI with multivariate PDFs.

As shown in Figure 5.6, this method includes 2 steps:

1) In the first step, the multivariate PDFs method replaces the Copula method to generate the correlation model, and then based on the selected number of mapping points m, m surrogate models are constructed at the same time;

2) In the second step, these m corresponding surrogate models are running in parallel to generate desired outputs. Those simultaneous simulations can be solved by parallel computation techniques or by using multi CPU cores.

Based on this proposed correlation modelling method, due to reduced number of system uncertainties, the computation cost of generating and running those surrogate models in steps 1 and 2 is smaller than the DASGI with Copula method.

5.4 Simulation results and discussion

As a typical collocation method, the SGI has similar computation mechanism as the PCM, therefore the SGI method is also capable of handling multiple outputs, but the computational burden is increasing when more outputs are required. Hence, in our case study, for either independent uncertainties case or correlated uncertainties case, both single output and multiple outputs cases are included and discussed. Meanwhile, the *FHSI* as expressed in Eq. (3.11) is used to indicate how accurate are the histograms generated by SGI or DASGI method comparing with those generated by the MCS.

5.4.1 SGI for independent uncertainties case

In total, 29 system uncertainties are considered in the aggregated SA power transmission network shown in Figure 5.1. Power flows from each of the buses are used as outputs in PPF. Those outputs are used to compare the SGI method and MCS method.

5.4.1.1 SGI with single output

In here, the power flow P_{31} which is power flow from bus 3 to bus 1, is used to compare accuracy of methods tested.

The Figure 5.7 shows the histograms generated by refence data simulation, MCS and SGI. Both tested methods fit well the reference results which indicates the feasibility of using those methods to handle PPF computation. The SGI curve is overlapping well with MCS results which confirms that with only few collocation points the constructed surrogate model is capable to map the inputs to desired output with similar accuracy as the MCS.

In Table 5.3, the values of *FHSI*s for both MCS and SGI are compared. These values are approximately 96%, which confirms that the SGI has the same computation accuracy as MCS. However, the computation cost is hugely different, as illustrated in Table 5.4.

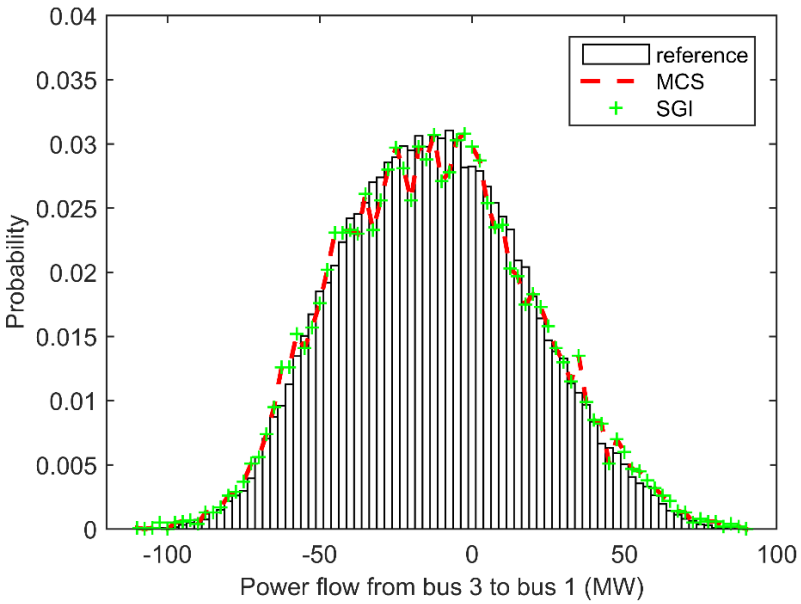

Figure 5.7. Histograms of P_{31} using reference, MCS and SGI.

Table 5.3 The *FHSI* of MCS and SGI with single output.

FHSI	MCS	SGI
Power flow P_{31}	96.14%	96.13%

Table 5.4 Computation time of 6 buses SA power network with single output.

		MCS	SGI
N_{MCS}		10000	10000
$N_{collocation}$		-	145
Step 1	t_{DPF} (s)	0.0145	0.0145
	t_{other} (s)	16.8	0.3
Step 2	$t_{interpolation}$ (s)	-	1.1×10^{-5}
t_{Total} (s)		161.9	2.5

The total computation time of MCS is composed of $N_{MCS} \times t_{DPF}$, where N_{MCS} is the number of MCS points and t_{DPF} is the simulation time for each input point which is determined by DPF model. While the computation time of SGI includes two parts:

1) In the step 1, time of constructing surrogate model equals $N_{collocation} \times t_{DPF} + t_{orther}$, where $N_{collocation} \times t_{DPF}$ is time to generate the pairs of collocation input points and corresponding output using DPF model, t_{orther} is the time to construct the surrogate model based on collocation points;

2) In the step 2, based on the surrogate model, the time of creating the frequency histogram of desired output is $N_{MCS} \times t_{interpolation}$.

The Table 5.4 shows that the total simulation time of SGI is only about 1.5% of MCS which means SGI has striking computation efficiency without trading off the accuracy comparing with MCS. The t_{other} of SGI is not significantly affecting the total time, however, the $N_{collocation}$ and $t_{interpolation}$ are the critical parts to enable much lower computation cost comparing with MCS. The $N_{collocation}$ is determined by the number of uncertainties in the system which is relatively small comparing with N_{MCS}, this results in less computational effort on generating the surrogate model. However, with higher number of uncertainties involved into the SGI computation, $N_{collocation}$ will exponentially increase leading to more complex surrogate model, hence more computation time is needed at step 1; the computation efficiency of SGI is decreasing when increasing the number of uncertainties. $t_{interpolation}$ is another core part of SGI computation which determines the efficiency of generating desired output comparing with MCS. In this power system case study, the $t_{interpolation}$ is 10^3 smaller than t_{DPF}, hence, for the same N_{MCS} points the time saving is incredible. And with increasing number of N_{MCS} to obtain more accurate simulation results, this time saving will significantly increase as well. However, $t_{interpolation}$ is directly determined by surrogate model which means more complexity of surrogate model will cause more computation time for each simulation point, in other words, $t_{interpolation}$ is higher with more uncertainties considered in the system. With this 29 uncertainties and single output case, the SGI has striking computation time savings comparing with MCS. Although the advantage of SGI may not be so significant for large number of uncertainties, it is still a method of choice in PPF computation.

5.4.1.2 SGI with multiple outputs

In this section, power flows between 6 buses, P_{31}, P_{32}, P_{43}, P_{54} and P_{64}, are used as output variables in comparison of tested methods. The computation time is listed in Table 5.5. With the same number of MCS points and uncertainties, the time of MCS is almost the same as single output case shown in Table 5.4 which confirms that the number of desired outputs will not affect the computation time of the MCS. The time spent on step 1 of SGI is not significantly different between single and multiple outputs cases, whereas, the $t_{interpolation}$ of multiple outputs case is nearly 10 times larger

compared to single output case. Because SGI generates different surrogate models for each desired output, in step 2, the interpolation for each output must be computed in sequence, hence, higher computation time is needed to solve for more outputs. In here, with 29 uncertainties, the $t_{interpolation}$ is far less to t_{DPF}, so even with higher number of desired outputs, the SGI is still computationally efficient compared to MCS. However, for large number of uncertainty dimensions, $t_{interpolation}$ will become larger as well, and then for multiple outputs of large uncertainty dimensions' case, the further increased $t_{interpolation}$ may not be significantly smaller than t_{DPF}.

Table 5.5 Computation time of 6 buses SA power network with multiple outputs.

		MCS	SGI
N_{MCS}		10000	10000
$N_{collocation}$		-	145
Step 1	t_{DPF} (s)	0.0145	0.0145
	t_{other} (s)	16.8	0.4
Step 2	$t_{interpolation}$ (s)	-	0.9×10^{-4}
t_{Total} (s)		161.9	3.5

Table 5.6 The *FHSI*s of MCS and SGI with multiple outputs.

FHSI	MCS	SGI
Power flow P_{31}	96.14%	96.13%
Power flow P_{32}	94.46%	94.46%
Power flow P_{43}	93.35%	93.33%
Power flow P_{54}	98.40%	98.40%
Power flow P_{64}	96.83%	96.81%

With 10000 MCS points, both MCS and SGI have the limited resolution to capture probabilistic features of all outputs, hence, the *FHSI*s of multiple outputs shown in Table 5.6 are varying around 93% to 98%. The *FHSI*s of SGI are almost the same as the MCS which indicate SGI has the same accuracy as MCS to handle the PPF computation. The SGI has much smaller computation cost compared to MCS. Moreover, to obtain high accuracy of desired outputs, higher number of MCS points can be applied to the system, and this will further increase the simulation time difference between MCS and SGI due to relatively smaller $t_{interpolation}$ compared with t_{DPF}.

5.4.2 SGI for correlated uncertainties case

In the correlated uncertainties case (i.e. the second case), as shown in Table 5.2, more uncertainties are considered. Total number of uncertainties is 51. The optimised SGI

method, named as DASGI method, is applied to solve this high dimensional PPF problem. Based on the same aggregated SA power transmission network shown in Figure 5.1, the power flows from each buses, are used as outputs to illustrate the performance of the methods tested. The results are showing comparison between the MCS with copula, the DASGI with copula and the DASGI with multivariate PDFs method.

5.4.2.1 DASGI with single output

The Figure 5.8 shows the frequency histograms of power flow from bus 2 to bus 3 generated by the three simulation methods. The simulation results from proposed DASGI-multivariate PDFs method are overlapping well with the results calculated by both MCS-Copula and DASGI-Copula methods which indicates the feasibility of the introduced method to solve PPF problem.

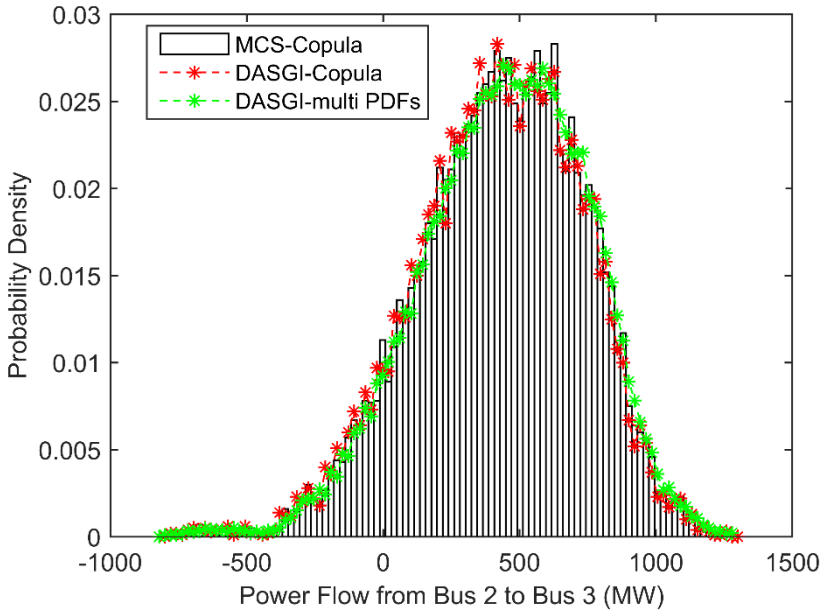

Figure 5.8. Histograms of P_{23} using MCS-Copula, DASGI-Copula and DASGI-multivariate PDFs.

The accuracy for the two DASGI methods, presented by using the *FHSI*, is shown in Table 5.7.

Table 5.7 The *FHSIs* of two DASGI method with single output comparing with MCS method.

FHSI	DASGI-Copula	DASGI-multivariate PDFs
Power flow P_{31}	96.93%	96.65%

Comparing with the MCS-Copula method, both DASGI methods have superior accuracy where the *FHSI*s are all over 95%, which again confirms that the proposed DASGI-multivariate PDFs method has similar computation accuracy as DASGI-Copula method.

Although the proposed method has similar accuracy as DASGI-Copula method, due to reduced number of system uncertainties involved into the DASGI model the computational burden is smaller. The simulation time of three methods for the main part of PPF computation are determined by different elements, which are:

a. MCS with Copula

   $N_{MCS} \times t_{DPF}$, where N_{MCS} represents the number of MCS points, and t_{DPF} is original simulation time by applying each simulation point to DPF model.

b. DASGI with Copula

   Two steps:

   1) In the step 1, constructing the surrogate model, $N_{collocation} \times t_{DPF} + t_{other}$, first, $N_{collocation}$ collocation points are selected and applied to the DPF model to obtain $N_{collocation}$ corresponding outputs, and then, based on these paired data the interpolation procedure can generate a simplified surrogate model with extra time cost t_{other};

   2) In the step 2, generating desired output, $N_{MCS} \times t_{interpolation}$, instead of using original DPF model, all the MCS points are applied to this surrogate model to evaluate the interpolated values at these simulation points.

c. DASGI with multivariate PDFs

   The same simulation steps are used as DASGI with Copula method, depending on the number of PDFs m, m surrogate models are constructed and running in parallel, hence, the related time cost for 1 surrogate model is considered in the comparison.

The Table 5.8 shows the simulation times of the three methods, where both DASGI methods have relatively smaller computational burden comparing with the MCS method. They benefit from the surrogate model with much shorter simulation time than DPF model. Small amount of collocation points is used to construct the surrogate model, and this enables reduction of the computation cost. Therefore, $N_{collocation}$ and $t_{interpolation}$ are the core parts to determine computation cost of DASGI. Because

$N_{collocation}$ is directly proportional to the number of system uncertainties, hence, the time cost of DASGI-multivariate PDFs method for step 1 is smaller than the method with Copula due to reduced number of system uncertainties resulting in the required $N_{collocation}$down to 119 from 303. Meanwhile, simplified surrogate model is constructed according to this system with reduced dimension, leading to further computation cost savings in step 2. As a result, the total time cost of proposed method is less than half of the time spent on Copula method. With this single output case, DASGI has striking computation cost reduction compared to MCS. Moreover, the proposed method further increases the computation efficiency by reducing the system dimension.

Table 5.8 Computation time of 3 simulation methods with single output.

		MCS-Copula	DASGI-Copula	DASGI-multivariate PDFs
N_{MCS}		10000	10000	10000
$N_{collocation}$		-	303	119
Step 1	t_{DPF} (s)	0.0169	0.0169	0.0169
	t_{other} (s)	-	0.374	0.163
Step 2	$t_{interpolation}$ (s)	-	4.47×10^{-4}	2.07×10^{-4}
t_{Total} (s)		169.8	10.0	4.3

5.4.2.2 DASGI with multiple outputs

The DASGI method can also handle multiple outputs cases at the expense of extra computational effort. In this section, all the power flows in the region MN are considered. In section 5.4.2.1, MCS-Copula, DASGI-Copula and DASGI-multivariate PDFs methods are compared, so in this section, those three methods are compared as well.

Table 5.9 Computation time of 3 simulation methods with multiple outputs.

		MCS-Copula	DASGI-Copula	DASGI-multivariate PDFs
N_{MCS}		10000	10000	10000
$N_{collocation}$		-	303	119
Step 1	t_{DPF} (s)	0.0169	0.0169	0.0169
	t_{other} (s)	-	0.504	0.288
Step 2	$t_{interpolation}$ (s)	-	1.04×10^{-3}	5.46×10^{-4}
t_{Total} (s)		169.8	16.1	7.8

In total, 16 desired outputs are calculated and the computation time for each method is listed in Table 5.9. In the MCS method, the simulation time will not be affected by the number of desired outputs with certain number of MCS points selected, therefore, the simulation time listed in Table 5.9 is the same as the one shown in Table 5.8. However, increasing the number of desired outputs will affect the computation time of both DASGI

methods. In step 1, one surrogate model is paired to one output, hence, increasing the number of desired outputs will require more surrogate models to be constructed, therefore, larger t_{orther} is required as shown in Table 5.9. This is nearly twice longer time than time required for the single output case. In step 2, for the same reason, extra time is needed to evaluate the interpolated values for each output based on the MCS points. So, comparing with Table 5.8, $t_{interpolation}$ is increasing. In here, the $t_{interpolation}$ is still relatively smaller than t_{PDF}, hence the total computation time of both DASGI methods are still far less than the time required for MCS method. However, in the case of further increase of outputs, the computation efficiency of DASGI may not be improvement over the MCS method.

The computation accuracy measure of both DASGI methods is represented as *FHSI*s and it is shown in Figure 5.9. Comparing with the MCS method, both DASGI methods have acceptable accuracy with all the *FHSI*s over 95%. The *FHSI*s of DASGI-multivariate PDFs method are very close to the results calculated by DASGI-Copula method which indicates that the proposed method has similar accuracy as commonly used DASGI-Copula method. Whereas, as reported in Table 5.8 and 5.9, the proposed method further reduces the computation cost.

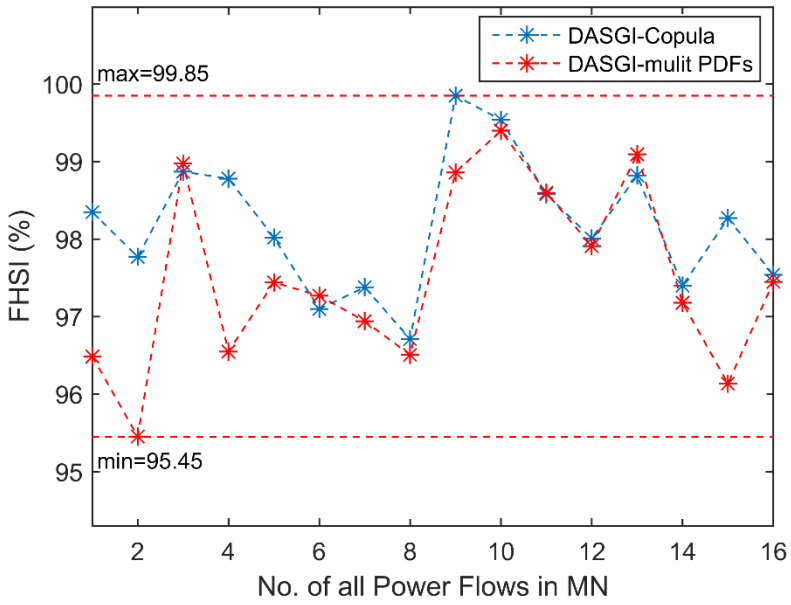

Figure 5.9. The *FHSI*s of both DASGI methods for the power flows in MN.

5.4.3 Factors affecting the computation time of the SGI based method

The computation time for SGI based method, as shown in Table 5.4, 5.5, 5.8 and5.9, is mainly affected by factors $N_{collocation}$ and $t_{interpolation}$ where $N_{collocation}$ is determined by the number of uncertainties and $t_{interpolation}$ is determined by both the number of uncertainties and the system size. Therefore, with certain number of system uncertainties, $N_{collocation}$ is not change, and increasing the size of power system will lead to the increasing of $t_{interpolation}$. Meanwhile for MCS method t_{DPF} will increase as well, as $t_{interpolation}$ is much smaller than t_{DPF}, hence, the simulation time of the MCS will increase faster than for the SGI based method, and this will lead to more simulation time savings of SGI based method with large power systems. In [37], the simulation results of IEEE 39-bus and 118-bus testing system show that with identical number of system uncertainties the time savings is increasing as increasing the system size.

In conclusion, the factors affecting the time saving of SGI compared to MCS is summarized in Table 5.10.

Table 5.10 The f**actors affecting the time saving from SGI based method to MCS method.**

Factors		Time saving
increase ↑, decrease ↓		
Certain number of uncertainties	Number of outputs ↑	↓
	Number of MCS points ↑	↑
	Power system size ↑	↑
Certain size of system	Number of uncertainties ↑	↓
	Number of outputs ↑	↓
	Number of MCS points ↑	↑

The SGI based PPF computation method constructs the surrogate model to replace the traditional DPF model which significantly reduced the computation cost when solving PPF problems. Whereas, from Table 5.10, the system dimension and number of desired outputs apparently limit the advantage. Hence, in correlated uncertainties case study, reducing the number of system uncertainties is the most direct and effective way to improve the capability of SGI based method in dealing with high-dimensional problems or multiple outputs cases. Therefore, with the proposed correlation modelling method, Table 5.8 and 5.9 show the shorter computation time is achieved comparing with the one with Copula method without much trade off in accuracy. Meanwhile, once the surrogate model is generated, it is only needed to be updated when the operation conditions of the system are modified.

5.5 Conclusions

In this chapter, upon the SA power system historical data and the system uncertainties' probabilistic model, the case studies of using SGI based method to handle PPF computation are demonstrated. Both, independent uncertainties and correlated uncertainties cases are considered.

For independent uncertainties case, the basic SGI is used as the key PPF computation method and compared with MCS. All the simulations are based on the aggregated SA power system network which has high significance in evaluating the capability of the SGI method in practical "real life" systems. The SGI method constructs surrogate model to replace the traditional DPF model. This gives striking time savings in solving PPF problems. Both the single output and multiple outputs cases indicate, comparing with MCS, that the SGI has incredible time saving without trading off the accuracy. Although SGI is highly affected by the number of uncertainties and desired outputs, this method is still the best for lower and medium uncertainty dimensions of input space in PPF computation.

For correlated uncertainties case, more system uncertainties are involved, hence, the DASGI method, which is developed from standard SGI method to further reduce the computation effort in dealing with high dimension problems, is applied to the SA power system model. Furthermore, multivariate PDFs method is introduced to model dependences between power system uncertainties. The aim of the proposed method is to reduce dimension of the problem, hence, to improve the capability of DASGI method in dealing with high-dimensional and multiple outputs cases. Comparing with the Copula correlation modelling method, the simulation results from both single and multiple outputs indicate that the proposed method improves the computation efficiency of DASGI method without much trade off in accuracy. The study confirms that the DASGI combined with multivariate PDFs method is capable to solve PPF problems with high-dimensional input space.

Chapter 6

PROBABILISTIC POWER SYSTEM PLANNING

THIS chapter demonstrates a practical application of using collocation method to manage the uncertainty in the power flow studies and power system planning of realistic transmission network. Nowadays, the impact of uncertainties to the power system makes the PPF analysis more valuable compared with the DPF analysis. In the previous chapters, the Sparse Grid Interpolation deterministic sampling technique is proved as the most competent and efficient method to solve the probabilistic power flow problem. Therefore, in this chapter, according to the historical and forecasted power system data for the SA transmission grid, and based on the SGI deterministic sampling technique the PPF model is constructed and intended for use in probabilistic planning studies. Meanwhile, the effectiveness and practicability of this proposed methodology are evaluated by comparison with the commonly used Monte Carlo Simulation method.

6.1 Introduction

Due to the large number of wind farms installed in South Australia and the variability of wind speed and consequently output power of those plants, a significant increase of uncertainty is expected to impact network planning and operation. Meanwhile, the number of installed rooftop PV generation is growing year by year, furthermore, multiple solar farms are also introducing more uncertainties to transmission system planning. The traditional Deterministic Power Flow model that is currently used has the limitation of considering the probabilistic nature of power system uncertainties. In addition, when performing deterministic planning based on the DPF model, the probability of a contingency event occurring is not taken into account. The deterministic approach simply requires that the network continues to provide an adequate and secure supply of electricity to customers for any contingency from the pre-selected list. This planning methodology can deliver a higher level of network redundancy, which may not be needed for many customers.

Probabilistic Power Flow analysis was first presented in 1974 [6], and is widely used in probabilistic planning nowadays. The PPF analysis not only calculates values of system variables (e.g. bus voltages), but also quantifies the probability of those variables being impacted by uncertain inputs [6].

As described in previous chapters, the most common approach in solving the PPF model is based on the Monte Carlo probabilistic sampling technique. This method is a straightforward method with high accuracy. However, it requires a large number of simulation samples, and therefore has a very high computation cost and computation time. This limitation is not always acceptable in practical studies of large transmission networks, which have many sources of uncertainty. Hence, an alternative to the MCS based on a deterministic sampling technique, named Sparse Grid Interpolation, is proposed and detailed in Chapter 5. This method is suitable for solving high dimensional problems with high accuracy and low computation cost. Implementation of this sampling method enables PPF analysis with a high computation efficiency compared to the MCS technique.

According to the PPF model, the proposed probabilistic planning methodology is based on an economic cost-benefit analysis, where the economic viability of each proposed network augmentation is evaluated [74]. In this context and under significantly

increasing uncertainty, use of the PPF model as a main tool for computation in probabilistic planning is seen to be of great potential importance for the South Australian transmission grid operator. By using the PPF model a network planner can, for each contingency, calculate the probability of constraints and load shedding occurrences and predict the impact of load shedding. The Expected Unserved Energy (EUSE) can be computed, and for a known Value of Customer Reliability (VCR), a balance can be found between delivering secure and reliable supply of electricity and maintaining reasonable costs for customers [39], [75], [76].

The remainder of this chapter is organised as follows, Section 6.2 details the information of the testing system and different case study scenarios. Based on the simulation results, the probabilistic power system planning is presented in Section 6.3. Conclusions are presented in Section 6.4.

6.2 Deterministic sampling technique for PPF model computation

The SA transmission network, shown in Figure 5.1, is applied to the PPF analysis. It consists of 6 buses, each bus representing one region of SA. The total number of the system uncertainties is 20, as listed in Table 6.1.

Table 6.1 Number of uncertainties of 6 buses SA aggregated power network.

Regions	Bus number	Demand	Wind farms	Solar farms	Inter connector	Total
EP	Bus 1	1	1	1		3
YP	Bus 2	1	2			3
MN	Bus 3	1	2	1		4
MET	Bus 4	1	1			2
RL	Bus 5	1	1	1	2	5
SE	Bus 6	1	1		1	3
Total number of uncertainties in the system						20

In this section, the purpose of running PPF computation is to test the deterministic sampling technique in the application of power system planning. The renewable generation has significant impaction on power system planning and operation. So, the testing data used here considers not just wind farms but also the newly installed solar farms. All the uncertainties are represented using PDFs constructed using the method demonstrated in Section 3.2. Besides representing the demand and wind power generation as system uncertainties, the uncertainty of the inter connector between SA and VIC is accounted for by considering it as a system demand. The solar farm data is

combined with regional system demand.

The deterministic sampling technique applied here is based on the SGI method. To verify accuracy and computation efficiency of this technique, the MCS is used as benchmark technique. Good feature of MCS is its high accuracy, but as a trade-off the computation burden is quite high. The processing block diagram shown in Figure 5.3 details the basic steps involved in MCS.

As shown in the Figure 5.3, the MCS calls the DPF model for every simulation point to generate outputs, which leads to the high computation cost. Instead of using for every simulation point DPF model, the deterministic sampling method separates the simulation into 2 steps, as shown in Figure 5.4. A surrogate model of the DPF is constructed based on a small number of collocation points. This input-output mapping model has higher computation efficiency compared with the original DPF model. Then applying all the MCS points to this surrogate model, the desired outputs are computed.

As discussed in previous chapters, the SGI based deterministic sampling method has the drawback that each desired system output requires a respective surrogate model. Therefore, increasing the number of desired outputs leads to a growing computation time when creating all the corresponding surrogate models. However, due to the relatively small computation cost of each surrogate model compared with the original DPF model, the deterministic sampling method is still computationally more efficient compared to the MCS method in the case of multiple outputs. The power flow between two buses of the SA network is used to demonstrate both single output and multiple outputs cases.

6.2.1 Single output case

The power flow P_{31} which represents the power flow from bus 3 to bus 1, is used to compare the PPF computation methods. The Figure 6.1 shows the PDF of the simulation results of both the deterministic sampling method and the MCS method.

The PDF obtained by using the deterministic sampling method closely maps the results computed by using the MCS method with a relative error of around 1.04×10^{-4}, indicating the feasibility of applying the SGI method to accurately determine PPF, while also enormously improving computation efficiency, as illustrated in Table 6.2.

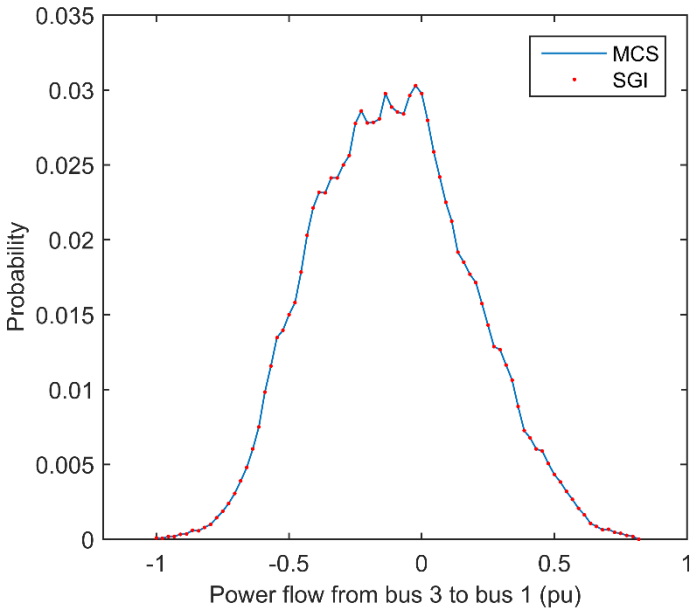

Figure 6.1. PDFs of P_{31} calculated using Monte Carlo (MCS – bold line) and deterministic sampling (SGI - dots). Results are normalized using maximum power flow (pu base).

Table 6.2 Computation times of the PPF model of aggregated power network in SA of single output case.

	Monte Carlo	Deterministic sampling
Simulation time (s)	161.9	2.5

The total computation time of deterministic sampling method is only about 1.5% of the simulation time using the MCS method. This is because the computational effort in running surrogate model generated by the deterministic sampling method is much less than the computation effort of running the traditional DPF model.

6.2.2 Multiple outputs case

For multiple outputs case, the power flows between 6 buses, P_{31}, P_{32}, P_{43}, P_{54} and P_{64}, are used to compare each method. The relative error of each power flow computed by the deterministic sampling method compared to the MCS method is shown in Table 6.3.

Table 6.3 Relative error of each power flow computed by deterministic sampling method comparing with the MCS method.

Power flow	P_{31}	P_{32}	P_{43}	P_{54}	P_{64}
Relative error	1.04×10^{-4}	1.06×10^{-4}	2.26×10^{-4}	1.01×10^{-4}	3.01×10^{-4}

The relative errors of the power flows shown in Table 6.3 vary, depending on accuracy of the corresponding surrogate model. Although some of the relative errors are larger, which is caused by selecting collocation points or interpolating, all the relative errors are less than 10^{-3}, confirming the accuracy of the proposed method compared to the MCS method. The simulation time is shown in Table 6.4.

Table 6.4 Computation times of the PPF model of aggregated power network in SA of multiple outputs case.

	Monte Carlo	Deterministic sampling
Simulation time (s)	161.9	3.5

With 5 desired outputs, the computation cost of the deterministic sampling method is about 2% of the MCS method. Compared to the single output case, although the number of desired outputs is increased by 5 times, the simulation time only increases by about 1.4 times; this is attributed to the computational efficiency of the surrogate model.

The accurate simulation results and improved computation cost demonstrated above show the effectiveness, efficiency and accuracy of deterministic sampling method comparing with the commonly used MCS method. Adopting the proposed simulation technique in probabilistic power system planning will lead to significant practical benefits.

6.3 Probabilistic power system planning

As a basis for economic planning, a probabilistic planning approach is applied to evaluate power system risks and to decide whether an investment maximises the net present value of the market benefit. The proposed deterministic sampling method has an incredible time saving in PPF computation without a significant reduction in accuracy. Therefore, the hugely reduced computation cost makes the method well suited to handle either operational or expansion planning tasks, and to help facilitate day to day operation of the system.

To adequately justify a network investment, the expected unserved energy is selected as an output parameter. It is defined as the amount of energy that is expected to be unsupplied in a year due to network constraints. According to PPF computation results calculated in previous section, the power flow P_{31} is used to demonstrate the application in probabilistic planning study. Based on the computed PDF of P_{31}, and only considering

the magnitude of the power flow, the CDF of P_{31} is obtained, and it is shown in Figure 6.2.

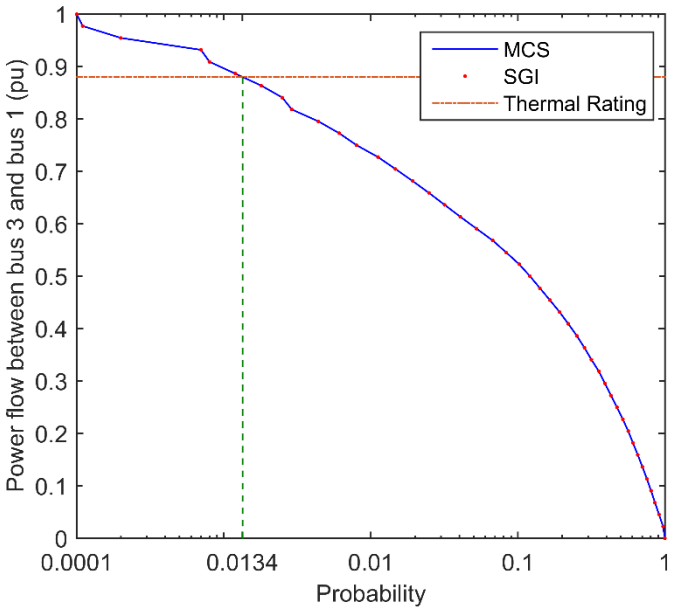

Figure 6.2. The CDFs of P_{31} calculated using Monte Carlo (MCS – bold line) and deterministic sampling (SGI - dots). Results are normalized using maximum power flow (pu base).

In this case study, the CDF is used to evaluate the probability of exceeding a notional thermal rating of the transmission lines between buses 1 and 3. As seen in the previous section, the deterministic sampling method has almost equivalent accuracy compared with MCS results; hence, the CDF calculated by the deterministic sampling method fits well the MCS results, and the results of the deterministic sampling method and surrogate model may be directly used in probabilistic planning studies.

Based on the CDF shown in Figure 6.2, if the notional thermal rating between buses 1 and 3 is 0.88pu (using maximum power flow as a base), then the probability of exceeding the thermal rating is 0.00134. According to this simulation result, for a 1-year time period, the final result is 11.7 hours of expected overload per year.

In addition, the expected unserved energy can be calculated by determining the area above the thermal rating line and below the CDF curve of power flow shown in Figure 6.2. So, about 50MWh of expected unserved energy is calculated.

Finally, the minimum amount of energy at risk is assessed, and that is needed to

economically justify a network investment. To overcome the above constraint, a new 275kV transmission line would be installed in parallel with the existing transmission lines between bus 1 and bus 3. Assuming an equivalent annual cost over this additional line's expected life of around \$2.5 million per annum [75] and a Value of Customer Reliability (VCR, representing customers' willingness to pay for reliability of electricity supply) of \$26880/MWh [76] yields a minimum required expected unserved energy to justify the investment: \$2.5 million/\$26880 = 93MWh, which is higher than the actual value previously determined. Therefore, this investment is not yet economically justified. However, if the power system demand is forecast to grow by 0.5% each year, calculating the CDFs of the power flow P_{31} for the next 10 years, the curves are obtained as shown in Figure 6.3.

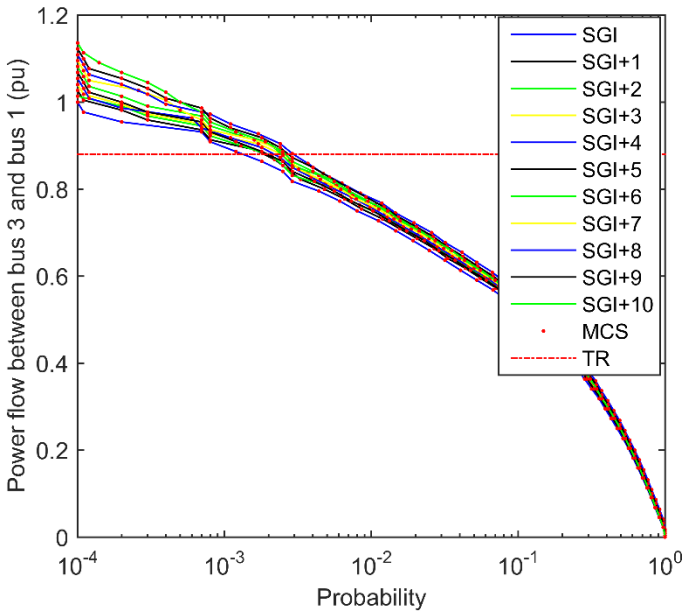

Figure 6.3. The CDFs of P_{31} calculated using Monte Carlo (MCS – bold line) and deterministic sampling (SGI - dots) based on 10 years forcasting. Results are normalized using maximum power flow (pu base).

The Figure 6.3 shows the simulation results obtained by using deterministic sampling method, based on historical data (SGI), and forecasting year 1 (SGI+1), forecasting year 2 (SGI+2), …, and forecasting year 10 (SGI+10). The results match well the corresponding MCS results (dots), and show that with growing load, the probability of exceeding the thermal rating is increasing. Hence, the expected unserved energy becomes larger over time, as shown in Table 6.5.

Table 6.5 Expected unserved energy with 0.5% yearly load increasing rate between bus 3 and bus 1 of aggregated power network in South Australia.

0.5% yearly load increasing rate	Historical data	Forecasting year									
		1	2	3	4	5	6	7	8	9	10
Time (h)	11.7	12.4	14.1	15.8	16.7	18.9	20.1	21.9	23.0	23.8	25.4
Expected unserved energy (MWh)	50	67	71	89	92	100	122	130	155	160	172

As the load grows, both the demand at risk above the thermal rating and the period of exposure are increasing. After 5 years, the total duration of predicted overload has increased to 18.9h, compared to 11.7h based on the historical data. The calculated expected unserved energy in year 5 is up to 100 MWh, which exceeds the economic threshold value of 93MWh, and now, the cost of adding parallel transmission line capacity is justified. This indicates that an investment to build a new transmission line between bus 3 and bus 1 will become justified 5 years after the initial exposure.

While the above case studies do not consider the increasing penetration of renewable generation in South Australia, and assumes this transmission capacity is the only constraint in the system, the principle is well demonstrated. The proposed deterministic sampling method and application of surrogate model has the same practical accuracy as the MCS method, whereas the computation cost is greatly reduced.

6.4 Conclusions

In this chapter, the deterministic sampling technique, which is based on the SGI method, is introduced to PPF computation and power system probabilistic planning. Based upon historical data and an aggregated model of the SA transmission system, the PPF computation results obtained by using the proposed method have been compared with the commonly used MCS method. A similar accuracy of simulation results is obtained by both methods, while the deterministic sampling method and surrogate model shows a striking improvement in computation efficiency. Results from a small probabilistic planning case study have been used to demonstrate use of the PPF model. Due to high computing efficiency, the proposed method is expected to be adequate to handle large and challenging probabilistic planning tasks. An example area in which the method may prove particularly suitable is for performing planning studies of future potential constraints on renewable generation output that may need to be properly addressed. A vastly reduced computation cost of the proposed technique compared to the MCS

method will improve the feasibility and practicality of performing probabilistic planning studies of the large transmission network.

Chapter 7

CONCLUSIONS AND FUTURE WORK

THIS chapter concludes the thesis by firstly reviewing and summarising the results and findings of each chapter. Two types of collocation methods, the probabilistic collocation method and sparse grid interpolation method, are proposed to handle the PPF analysis. In addition, the sparse grid interpolation method is applied to power system planning. This chapter also presents the possible future research in this field.

7.1 Conclusions

The power transmission network is embedded with large number of uncertain factors, such as the system demand which is deeply associated with human living habit, the renewable generation which has fluctuation and intermittent feature, etc. Deterministic power flow analysis lacks models that consider the probabilistic nature of those power system uncertainties. Hence, to overcome this issue, the probabilistic power flow analysis was introduced. According to the literature review, variety of PPF analysis methodologies have been developed, and each comes with pros and cons. Among those methods, the collocation method shows potentially incredible computation efficiency and high accuracy while solving the PPF problems. Besides that, the collocation method has the ability to construct the correlation model of interdependent uncertainties. Furthermore, this correlated uncertainty modelling method further increases the capability of collocation method in handling high dimensional PPF analysis. According to those features of collocation method, the aim of this thesis was to use the collocation method as basic for PPF analysis, and to solve a realistic power transmission network problem. In addition, objective was to verify the effectiveness and computation efficiency of this method.

It is noted that a deep and thoughtful understanding of the theories behind the collocation method is prerequisite to conduct the PPF analysis. Therefore, in Chapter 2 the basic theories behind the collocation method are described. In our study, two types of collocation method are adopted, and they are probabilistic collocation method (PCM) and sparse grid interpolation (SGI) method. The PCM is based on orthogonal polynomials and the Gaussian quadrature integration. Whereas, the SGI method is based on Smolyak's construction in multi-dimensional interpolation. Although these two methods are based on different theories, they all intend to construct a cost-effective surrogate model to replace traditional DPF model. Hence, the PPF computation efficiency is strikingly improved with those models. The computation efficiency of both collocation techniques is affected by the parameter space dimension. The PCM can build up the input-output mapping equation between uncertain inputs to desire outputs with less computation effort than MCS, however this method is highly affected by input parameter space dimension, hence, it is more suitable for low or medium dimension PPF analysis. To solve high dimensional cases, the SGI method breaks the so-called curse of dimensionality and is more computationally competent than the PCM, whereas more

complicated steps are applied to construct the surrogate model comparing with the PCM. Based on the pros and cons of those methods, in our study: the PCM is applied to solve low dimension PPF cases, and to reveal the relationship between interdependent uncertainties; and the SGI method is used to high dimensional PPF analysis to achieve better computation efficiency.

Based on the review in Chapter 2, the accuracy of constructing the surrogate model is highly depended on the accuracy of probabilistic model of the system uncertainties. Therefore, the Chapter 3 presented details of power system uncertainties modelling. In our study, the historical data of South Australia power system are used, and the system uncertainties mainly considered in our work are system demand and wind farms generation. In previously proposed method, the system demand and wind speed are modelled using normal distribution and Weibull distribution respectively. However, based on our research, in some realistic cases the PDF of neither system demand nor wind speed is following any known distribution. Hence, Gaussian Mixture Model (GMM) is applied to our work which gives an accurate approximation to the non-Gaussian distributions. The constructed PDFs of those uncertainties are compared with the original data which confirms the ability of the GMM method to model the non-Gaussian distributions. Meanwhile, instead of using either typical wind power curve or theoretical wind power formula, according to the features of PCM, the realistic relationship between wind speed and wind power generation is constructed by using the PCM.

Besides the system uncertainties modelling, another challenge in our work was to consider the interdependences between system uncertainties which is significantly affecting the analysis. Based on the literature review, the Copula theory is a proper and efficient method to construct the nonlinear dependent relationship between correlated data. In addition, according to different scenarios, two PCM based correlation modelling methods are proposed, and their applications are demonstrated. They are:

- for small number of correlated uncertainties, PCM combined with Fuzzy logical optimization;
- for large number of correlated uncertainties, a special multivariate PDFs modelling method which is also based on PCM is applied to build the correlation models.

Comparing those two novel methods with the Copula method, besides accurately revealing the nonlinear relationship between correlated uncertainties, both of those

newly developed methods have the ability to reduce the number of system uncertainties, hence further improve the computation efficiency in conducting the PPF analysis.

With the power system uncertainties model and the correlation model, the two types of collocation method based PPF analysis are detailed in Chapter 4 and Chapter 5. First, the PCM based PPF analysis is presented in Chapter 4. In order to verify the effectiveness and efficiency of the PCM, a simplified SA power transmission network is used to demonstrate how the PCM handles PPF computation. According to the basic theories behind the PCM, this method is highly affected by the parameter space dimension. Hence, in our study, for small number of correlated uncertainties, the proposed correlation modelling method is incorporated to reduce the system dimension, thus making the PCM more computationally competent to solve the PPF problems. Meanwhile, because for every desired output a separated input-output mapping model is required, by increasing number of desired outputs will lead to higher computation cost. So, in our case studies, both single output and multiple outputs cases were demonstrated. The results are compared with the MCS method, and they clearly show the computation efficiency is vastly improved even with multiple outputs case. Once those input-output mapping models are created, the power system simulation can be run in real-time, and that can guide operator in daily power system operation and scheduling.

The only limitation of PCM analysis method is that when input parameter space dimension increases the computation cost will exponentially increase. To overcome this issue, another collocation method, so-called SGI method, is applied to our study. Details were provided in Chapter 5. Based on the same aggregated SA power system network, the SGI method is first applied to solve independent uncertainties case. Similar as the PCM, for each desired output, the SGI method needs to construct the surrogate model separately, therefore, the computation cost will increase when more outputs are required. Thus, in our study, both single output and multiple outputs cases are considered. By comparing with the MCS method, with incredible computation time saving of the surrogate model, the SGI method achieves similar computation accuracy as the MCS results. Furthermore, the number of the system uncertainties is increased which include more correlated data. In other to mitigate the effect of increasing of parameter space dimension, the DASGI method, which is developed from the standard SGI method, is applied to further improve PPF analysis. Besides this, for large number of

correlated uncertainties, the proposed novel correlation modelling method is applied to the system as well. To verify the effectiveness and accuracy of this PPF analysis method that includes a novel correlation model, results are compared with the one obtained with the Copula modelling method. The simulation results from both single and multiple outputs indicate that the proposed correlation modelling method combined with the DASGI has high computation efficiency without much trade off in accuracy and confirms its ability of solving higher dimensions of parameter input space in PPF computation.

In Chapter 4 and Chapter 5 the effectiveness and efficiency of the collocation method are verified. For large power system, the SGI method is more computationally competent to conduct the PPF computation. To extend the application of SGI method, Chapter 6 demonstrates applying this method to the power system planning. The proposed probabilistic planning methodology, discussed in this chapter is based on an economic cost-benefit analysis, where the economic viability of each proposed network augmentation is evaluated. In this context and under significantly increasing uncertainty, use of the proposed PPF model, as a main tool for computation in probabilistic planning, is seen to be of great potential importance for the South Australian transmission grid operator. By using this PPF model a network planner can, for each contingency, calculate the probability of constraints and load shedding occurrences and predict the impact of load shedding. Furthermore, the Expected Unserved Energy can be computed, and with a known Value of Customer Reliability, a balance can be found between delivering secure and reliable supply of electricity and maintaining reasonable costs for customers. An example area in which the method may prove particularly suitable is for performing planning studies of future renewable generation and associated system constrains. A vastly reduced computation cost of the proposed technique compared to the MCS method will improve the feasibility and practicality of performing probabilistic planning studies of large transmission networks.

7.2 Future work

This thesis proposed a generic approach of applying collocation method based PPF analysis to SA transmission network. Several possible research studies can be carried out using this thesis as a starting point. They are described in the following paragraphs:

- **Large power system PPF computation:** Comparing with the commonly and wildly used MCS method, the significant advantage of the collocation method is

its computation efficiency. It is expected that such method will be adequate to handle large-size power system. When increasing the size of the power system, the computation cost of using MCS method will increase faster than when using the collocation methods, hence, this advantage will become more conspicuous. Although through the literature review, the ability of collocation method in handling large power system PPF computation has been verified, those are all based on theoretical testing system. More efforts and attentions should be on large real-size power system. Based on the theoretical evaluation, the experimental validation will be the next step to justify the practical effectiveness of this newly developed PPF analysis methodology. Moreover, more efficient programming languages should be considered to further increase the ability of collocation method to handle realistic power system PPF analysis.

- **Evaluate impact of uncertainties in power system planning:** In our probabilistic power system planning study, a generic case study is detailed which shows the effectiveness and efficiency of collocation method in dealing with probabilistic power system planning. However, it only demonstrates a small part of that area. With the increasing of renewable generation, the impact of those uncertain power system injections will be a challenge to power system planning studies. Properly evaluating the impact of those uncertainties to the power system will help the planner to figure out future potential constrains on renewable generation outputs. Furthermore, the steady state power system security assessment can also be conducted based on this impact studies.

- **Integrating with power system simulation software:** In the current work, system simulation has been done for steady state conditions. The dynamic and transient analysis of power system is another important type of simulation where large number of uncertainties should be considered. The power system simulation software, such as PowerWorld, PSS/E and so on, are the proper way to implement those dynamic and transient analysis. These tools are used in a wide range of planning and operational studies. Therefore, the question is, how to integrate the collocation methods with those software tools instead of combining them with MCS, and to how to improve the simulation speed compared with MCS. Meanwhile, the programming languages used in writing those software tools are more efficient than MATLAB and it is expected that the

overall probabilistic versions of such tools will be computationally faster compared to results presented in this thesis.

- **Power system uncertainties quantification:** Besides applying the collocation method to conduct the PPF analysis, this method can also be employed to quantify the power system uncertainties. In dynamic simulation of power system, a single machine infinite bus system and SA transmission network can be used to test this approach. Compared with MC method, the collocation method applies a small number of collocation points which could provide the possibility to reduce the number of required simulations. And in state estimation, according to the small number of collocation points applied to the simulation, the collocation method has the ability to quantify many power system uncertain variables with relatively lower computation cost. This approach can also be compared with MC method. Meanwhile the collocation method can be applied to study the uncertainty in transient behaviour of power system. Furthermore, based on the aggregated model of SA transmission network or a larger real-size power system, an index for identification of key uncertain parameters can be demonstrated.

BIBLIOGRAPHY

[1] '2016 Renewable Energy Data Book'. Available at https://www.nrel.gov/docs/fy18osti /70231.pdf, accessed February 2018.

[2] 'South Australian Electricity Report 2019'. Available at https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/SA_Advisory/2019/2019-South-Australian-Electricity-Report.pdf, accessed June 2020.

[3] 'Transmission Annual Planning Report 2019'. Available at https://www.electranet.com.au/wp-content/uploads/2019/06/2019-ElectraNet-TAPR_WEB.pdf, accessed June 2020.

[4] J. C. Smith, M. R. Milligan, E. A. DeMeo, and B. Parsons, "Utility wind integration and operating impact state of the art," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 900-908, Aug. 2007.

[5] R. Mamlook, O. Badran, and E. Abdulhadi, "A fuzzy inference model for short-term load forecasting," *Energy Policy.*, vol. 37, no. 4, pp. 1239-1248, Apr. 2009.

[6] B. Borkowska, "Probabilistic load flow," *IEEE Trans. Power App. Syst.*, vol. PAS-93, no. 3, pp. 752–759, May 1974.

[7] P. Chen, Z. Chen, and B. Bak-Jensen, "Probabilistic load flow: A review," in Proc. 3rd Int. Conf. Electric Utility Deregulation and Restructuring and Power Technologies, 2008 (DRPT 2008), Apr. 2008, pp. 1586-1591.

[8] D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of Monte Carlo Methods.* New York, NY, USA: Wiley, 2011.

[9] H. Yu, C. Y. Chung, K. P. Wong, H. W. Lee, and J. H. Zhang, "Probabilistic load flow evaluation with hybrid Latin hypercube sampling and Cholesky decomposition," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 661-667, May 2009.

[10] Y. Chen, J. Wen, and S. Cheng, “Probabilistic load flow method based on Nataf transformation and Latin hypercube sampling,” *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 294-301, Apr. 2013.

[11] D. Cai, D. Shi, and J. Chen, “Probabilistic load flow computation with polynomial normal transformation and Latin hypercube sampling,” *IET Gener., Transm., Distrib.*, vol. 7, no. 5, pp. 474-482, May 2013.

[12] A. Owen, “Latin supercube sampling for very high-dimensional simulations,” *ACM Trans. Model. Comput. Simul.*, vol. 8, no. 1, pp. 71-102, Jan. 1998.

[13] M. Hajian, W. D. Rosehart, and H. Zareipour, “Probabilistic power flow by Monte Carlo simulation with Latin supercube sampling,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1550-1559, May 2013.

[14] R. N. Allan, A. M. L. da Silva, and R. C. Burchett, “Evaluation methods and accuracy in probabilistic load flow solutions,” *IEEE Trans. Power App. Syst.*, vol. PAS-100, no. 5, pp. 2539–2546, May 1981.

[15] A. M. L. da Silva and V. L. Arienti, “Probabilistic load flow by a multilinear simulation algorithm,” *IEEE Proc. Gener., Transm., Distrib. C*, vol 137, no. 4, pp. 276-282, Jul. 1990.

[16] P. Zhang and S. T. Lee, “Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion,” *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 676–682, Feb. 2004.

[17] J. Usaola, “Probabilistic load flow in systems with wind generation,” *IET Gener., Transm., Distrib.*, vol. 3, no. 12, pp. 1031-1041, Dec. 2009.

[18] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, “Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants,” *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2251-2261, Nov. 2012.

[19] J. He and G. Sallfors, “An optimal point estimate method for uncertainty studies,” *Appl. Math. Model.*, vol. 18, no. 9, pp. 494-499, Sep. 1994.

[20] H. Hong, “An efficient point estimate method for probabilistic analysis,” *Reliab. Eng. Syst. Saf.*, vol. 59, no. 3, pp. 261-267, Mar. 1998.

[21] C.-L. Su, “Probabilistic load-flow computation using point estimate method,” *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1843-1851, Nov. 2005.

[22] J. M. Morales and J. Perez-Ruiz, ”Point estimate schemes to solve the probabilistic power flow,” *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1594-1601, Nov. 2007.

[23] J. M. Morales, L. Baringo, A. J. Conejo, and R. Minguez, "Probabilistic power flow with correlated wind sources," *IET Gener., Transm., Distrib.*, vol.4, no. 5, pp. 641-651, May 2010.

[24] M. Aien, M. G. Khajeh, M. Rashidinejad, and M. Fotuhi-Firuzabad, "Probabilistic power flow of correlated hybrid wind-photovoltaic power systems," *IET Renew. Power Gener.*, vol. 8, no. 6, pp. 649-658, Aug. 2014.

[25] J. R. Hockenberry and B. C. Lesieutre, "Evaluation of uncertainty in dynamic simulations of power system models: the probabilistic collocation method," *IEEE Trans. Power Sys.*, vol. 19, no. 3, pp. 1483-1491, Aug. 2004.

[26] J. R. Hockenberry, "Evaluation of uncertainties in dynamic, reduced-order power system models," Ph D. dissertation, MIT, Cambridge, MA, Sep. 2000.

[27] D. Han and J. Ma, "Effect of uncertainties in parameters of load model on dynamic stability based on probabilistic collocation method," *IEEE Lausanne Power Tech*, pp. 1100-1104, July. 2007.

[28] C. Zheng and M. Kezunovic, "Impact of wind generation uncertainty on power system small disturbance voltage stability: A PCM-based approach," *Electr. Power Syst. Res.*, vol. 84, no. 1, pp. 10-19, Mar. 2012.

[29] R. Preece and J. V. Milanovi, "The probabilistic collocation method for dealing with uncertainties in power system small disturbance studies," *2012 IEEE Power and Energy Society General Meeting*, pp. 1-7, July. 2012.

[30] R. Preece, N. C. Woolley, and J. V. Milanovic, "The probabilistic collocation method for power-system damping and voltage collapse studies in the presence of uncertainties," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2253-2262, Aug. 2013.

[31] G. Lin, N. Zhou, T. Ferryman, *et al*, "Uncertainty quantification in state estimation using the probabilistic collocation method," *2011 IEEE/PES Power System Conf. and Exposition*, pp. 1-8, Mar. 2011.

[32] H. Yin and R. Zivanovic, "An application of probabilistic collocation method in wind farms modelling and power system simulation," *2016 IEEE Innovative Smart Grid Technologies – Asia,* Melbourne, Australia, Dec. 2016, pp. 681-686.

[33] H. Yin and R. Zivanovic, "Using probabilistic collocation method for neighboring wind farms modelling and power flow computation of South Australia grid," *IET Generation, Transmission & Distribution*, vol. 11, no. 14, pp. 3568-3575, Oct. 2017.

[34] T. Gerstner and M. Griebel, "Dimension-adaptive tensor-product quadrature," *Computing*, 71(1), pp. 65-87, Jun. 2003.

[35] W. A. Klimke, “Uncertainty modeling using fuzzy arithmetic and sparse grids,” Ph.D. dissertation, mathematics, Univ. Stuttgart, Stuttgart, Germany, Nov. 2005.

[36] X. Ma and Z. Nicholas, “An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations,” *J. Computat. Phys.*, vol. 228, no. 8, pp. 3084–3113, May 2009.

[37] J. Tang, F. Ni, F. Pconci and A. Monti, “Dimension-adaptive sparse grid interpolation for uncertainty quantification in modern power systems: probabilistic power flow,” *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 907–919, Mar. 2016.

[38] H. Yin and R. Zivanovic, “Practical application of collocation method in power flow study of South Australia grid,” *International Journal of Electrical Power & Energy Systems*, vol. 94, pp. 160-170, Jan. 2018.

[39] M. Crnkovic, B. Parker, R. Korte, H. Yin and R. Zivanovic, “Managing uncertainty in the power flow studies of South Australian transmission network,” *CIGRE 2018*, Paris, France, Aug. 2018, C1-202.

[40] H. Yin and R. Zivanovic, “Probabilistic power flow computation using collocation method and including correlation modeling,” *International Transaction on Electrical Energy System*, vol. 29, no. 4, e2796, Apr. 2019.

[41] G. Papaefthymiou, “Integration of stochastic generation in power systems,” Ph.D. dissertation, Electrical Engineer, Delft Univ. Technol., Delft, The Netherlands, 2006.

[42] G. Papaefthymiou and D. Kurowicka, “Using copulas for modeling stochastic dependence in power system uncertainty analysis,” *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, Feb. 2009.

[43] S. Hagspiel, A. Papaemannouil, M. Schmid, and G. Andersson, “Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid,” *Appl. Energy*, vol. 96, pp. 33–44, Aug. 2012.

[44] Q. Xiao and S. Zhou, “Probabilistic power flow computation considering correlated wind speeds,” *Appl. Energy*, vol. 231, pp. 677-685, Dec. 2018.

[45] M. Webster, M. A. Tasang and G. J. McRae, “Application of the probabilistic collocation method for an uncertainty analysis of numerical geophysical models,” Joint Program on the Science and Policy of Global Change, MIT, Rep. 4, Jan. 1996.

[46] M. A. Tasang, W. Pan, R. G. Prinn and G. J. McRae, “An efficient method for parametric uncertainty analysis of numerical geophysical models,” *J. Geophys. Res.-Atmos*, vol. 102, no. 18, pp. 21925-21932, Sep. 1997.

[47] P. J. Davis and P. Rabinowitz, “Methods of Numerical Integration,” New York: Academic, 1975.

[48] Y. Zhou, Y. Wan, S. Roy, *et al*, “Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic flow management,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 44, no. 10, pp. 1347-1363, Oct. 2014.

[49] S. A. Smolyak, “Quadrature and interpolation formulas for tensor products of certain classes of functions,” *Soviet Math. Dokl.*, vol. 4, pp. 240-243, 1963.

[50] B. Volker, N. Erich and R. Klaus, “High dimensional polynomial interpolation on sparse grids,” *Advances in Computational Mathematics*, vol. 12, no. 4, pp. 273-288, 2000.

[51] D. Xiu, “Efficient collocational approach for parametric uncertainty analysis,” *Commun. Comput. Phys*, vol. 2, no. 2, pp. 293-309, Apr. 2007.

[52] F. Nobile, R. Tempone and C. Webster, “A sparse grid collocation method for elliptic partial differential equations with random input data,” *SIAM J. Number. Anal.*, vol. 46, no. 5, pp. 2309-2345, 2008.

[53] B. Bak-Jensen, J. Bech, CG. Bjerregaard and PR. Jensen, “Models for probabilistic power transmission system reliability calculation,” *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1166-1171, Aug. 1999.

[54] R. Pallabazzer, “Evaluation of wind-generator potentiality,” *Solar Energy*, vol. 55, no. 1, pp. 49-59, Jul. 1995.

[55] T.-H. mont and L. Wang, “A study on generator capacity for wind turbines under various tower heights and rated wind speeds using Weibull distribution,” *IEEE Trans. Energy Convers.*, vol. 23, no. 2, pp. 592-602, Jun. 2008.

[56] G. Orfanos, P. Georgilakis and N. Hatziaryriou, “Transmission expansion planning of systems with increasing wind power integration,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1355-1362, May. 2013.

[57] R. Singh, B. C. Pal and R. Jabr, “Statistical representation of distribution system loads using Gaussian Mixture Model,” *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 29-37, Feb. 2010.

[58] G. Valverde, A. T. Saric and V. Terzija, "Probabilistic load flow with non-Gaussian correlated random variables using Gaussian Mixture Models," *IET Generation, Transmission & Distribution*, vol. 6, no. 7, pp. 701-709, July. 2012.

[59] C. Cristina, R. Esther, R. Jesus, "Probabilistic load flow with versatile non-Gaussian power injections," *Electric Power Systems Research*, vol. 119, pp. 266-277, Feb. 2015.

[60] M. Webster, M. Tatang and G. McRae, "Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model," Joint Program on the Science and Policy of Global Change, MIT, Cambridge, MA, Tech. Rep 1996.

[61] J. Calbo, W. Pan, M. Webster, R. Prinn and G. McRae, "Parameterization of urban subgrid scale processes in global atmospheric chemistry models," *Journal of Geophysical Research: Atmospheres*, vol. 103, no. D3, pp. 3437-3451, Feb. 1998.

[62] D. C. Montgomery and G. C. Runger, "Applied statistics and probability for engineers," John Wiley & Sons, Inc., 3rd ed, New York, 2003.

[63] H. R. Sirisena and E. P. M. Brown, "Representation of non-Gaussian probability distributions in stochastic load-flow studies by the method of Gaussian sum approximations," *IEEE Proc.*, vol. 130, no. 4, pp. 165-171, July. 1983.

[64] D. J. Salmond, "Mixture reduction algorithms for point and extended object tracking in clutter," *IEEE Trans. Aerosp. Electron.*, vol. 45, no. 2, pp. 29-37, Jun. 2009.

[65] J. A. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian Mixture and hidden Markov models," Technical Report, ICSI-TR-97-021, International Computer Science Institute, 1998.

[66] W. Decoursey, "Statistics and probability for engineering applications," Burlington: Elsevier Science, 2003.

[67] T. Enric, E. Luka, "Fuzzy logic: an introductory course for engineering student," Springer International Publishing, Cham, 2015.

[68] 'Bureau of Meteorology Australia', http//www.bom.gov.au/climate/data/, accessed March 2018.

[69] D. Kurowicka and R. Cooke, "Uncertainty analysis with high dimensional dependence modelling," Chichester, U.K.: Wiley, Feb. 2006.

[70] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of Royal Statistical Society. Series B (Methodological).*, vol. 26, no. 2, pp. 211-252, 1964.

[71] MATLAB User's Guide, MathWorks, 1995.

[72] 'MATPOWER User's Manual', Available at http//www.pserc.cornell.edu/mat-power, accessed October 2016.

[73] 'Sparse Grid Interpolation Toolbox User's Guide', Available at https://people.sc.fsu.edu/~jburkardt/m_src/spinterp/doc/spinterpdoc.pdf, accessed January 2017.

[74] CIGRE Working Group C4.601, publication 434, "Review of the current status of tools and techniques for risk-based and probabilistic planning in power system", Oct. 2010.

[75] Victorian Energy Networks Corporation, "Electricity Transmission Network Planning Criteria", July 2003, available at www.aer.gov.au, accessed December 2018.

[76] Australian Energy Market Operator, "Value of Customer Reliability Review", September 2014, available at www.aemo.com.au, accessed December 2018.