Uncertainty Quantification and Control in Power System Security and Operation Via Data-Driven Polynomial Chaos Expansion Based Methods

Xiaoting Wang

Department of Electrical and Computer Engineering
McGill University
Montreal, Canada

January 2024

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in Electrical Engineering.

2024/01/08

Abstract

The global energy situation is shifting towards renewable energy sources (RESs) to promote sustainability and reduce fossil fuel reliance. This shift brings uncertainties from volatile RESs and new forms of loads (e.g., electric vehicles), challenging power system operation and security. Addressing these challenges, this thesis aims to leverage a surrogate modeling method, namely the polynomial chaos expansion method, to systematically investigate and mitigate the impacts of uncertainties on power system transfer capability and economic dispatch (ED). The overarching goal is to offer vital guidance for ensuring and enhancing the security of power systems while maximizing the utilization of transmission assets and economic benefits, considering the high uncertainty level of current and future power grids.

The thesis first studies the impacts of uncertainties brought by volatile RESs, random loads, and unforeseen equipment outages on power system available transfer capability (ATC), a crucial index in power system security analysis. By exploiting polynomial chaos theory and moment-based methods, a data-driven sparse polynomial chaos expansion (DDSPCE) method is developed for probabilistic total transfer capability (PTTC) and ATC assessment. Notably, without requiring preassumed probability distributions of random inputs, the proposed DDSPCE directly exploits data for estimating the probabilistic characteristics of PTTC (e.g., mean, variance, probability density function (PDF), and cumulative distribution function (CDF)), based on which the ATC with a certain confidence level can be readily calculated. An integrated sparse framework further enhances its computational efficiency and accuracy. Simulations on the modified IEEE 118-bus system and the modified PEGASE 1354-bus system validate the DDSPCE method's efficacy in PTTC evaluation. Furthermore, the results underscore the significance of incorporating discrete uncertainties, like equipment outages, in both PTTC and ATC assessments.

The thesis then delves into the impacts of uncertainties, especially from wind power, on ED, a critical aspect of the power system daily operation. A DDSPCE-based surrogate modeling method is developed to estimate the probabilistic characteristics of ED solutions, including their mean, variance, and distribution functions. The developed method can handle extensive random inputs without their predefined probability distributions. Extensive simulation results on an integrated electricity and gas system (IEGS) using real-life wind power data validate the efficiency and effectiveness of the proposed method in quantifying the impacts of uncertainties on the ED solutions, even when the ED solutions are multimodal. These results highlight the DDSPCE method's efficacy and efficiency in addressing general and complex scenarios.

After investigating the impacts of uncertainties on power system static security and ED, the thesis focus turns to mitigating these impacts. To this end, this thesis conducts a global sensitivity analysis to allocate the dominant random inputs to assist in designing the uncertainty-control measures. Particularly, different PCE-based models are developed and compared for global sensitivity analysis within the transfer capability enhancement and ED. Leveraging the insights from the sensitivity information, uncertainty control strategies (e.g., by utilizing energy storage systems) can be designed, thereby mitigating the impacts of uncertainties. These findings offer invaluable direction for uncertainty management and control design in real-world power system operations.

Résumé

La situation énergétique mondiale est en train de s'orienter vers les sources d'énergie renouvelable (SER) dans un souci de durabilité et de réduction de la dépendance aux combustibles fossiles. Cette évolution engendre des incertitudes, notamment à cause des SER volatiles et des charges aléatoires comme les véhicules électriques, posant ainsi des défis pour la gestion et la sécurité des systèmes électriques. Dans cette optique, cette thèse se propose d'utiliser une méthode de modélisation par substitution, spécifiquement l'expansion du chaos polynomial, pour étudier et atténuer systématiquement les effets des incertitudes sur la capacité de transfert et la distribution économique (DE) des systèmes électriques. L'ambition majeure est d'apporter des recommandations essentielles pour assurer et renforcer la sécurité des systèmes électriques, tout en optimisant l'utilisation des infrastructures de transmission et en maximisant les bénéfices économiques, en tenant compte des incertitudes croissantes des réseaux électriques actuels et futurs.

La thèse commence par examiner les conséquences des incertitudes introduites par les SER volatiles, les charges stochastiques et les pannes d'équipement imprévues sur la capacité de transfert disponible (CTD) des systèmes électriques, un indicateur essentiel pour l'analyse de leur sécurité. En exploitant la théorie du chaos polynomial et des méthodes basées sur les moments, une méthode d'expansion du chaos polynomial éparse et orientée données (DDSPCE) est élaborée pour évaluer la capacité de transfert total probabiliste (PTTC) et la CTD. Notablement, sans nécessiter de distributions probabilistes préétablies pour les entrées aléatoires, le DDSPCE utilise directement les données pour estimer les caractéristiques probabilistes du PTTC, telles que la moyenne, la variance, la fonction de densité de probabilité (FDP) et la fonction de distribution cumulative (FDC). Ces estimations permettent de calculer aisément la CTD à un niveau de confiance donné. Un cadre éparse intégré renforce davantage son efficacité computationnelle et sa précision. Des simulations sur des systèmes modifiés, tels que l'IEEE 118-bus et le PEGASE 1354-bus, confirment l'efficacité de la méthode DDSPCE pour évaluer le PTTC. De plus, les résultats soulignent l'importance de prendre en compte des incertitudes discrètes, comme les pannes d'équipement, dans les évaluations du PTTC et de la CTD.

La thèse explore ensuite les impacts des incertitudes, en particulier celles liées à l'énergie éolienne, sur la DE, un élément crucial de l'exploitation quotidienne des systèmes électriques. Une méthode de modélisation substitutive basée sur DDSPCE est mise au point pour estimer les caractéristiques probabilistes des solutions DE, y compris leur moyenne, variance et fonctions de distribution. Cette méthode peut traiter un grand nombre d'entrées aléatoires sans nécessiter de distri-

butions probabilistes prédéfinies. Des simulations approfondies sur un système intégré d'électricité et de gaz (IEGS) utilisant des données réelles d'énergie éolienne attestent de l'efficacité et de la pertinence de la méthode proposée pour quantifier les effets des incertitudes sur les solutions de DE, même lorsque ces solutions présentent plusieurs modes. Ces résultats mettent en évidence la pertinence et l'efficacité de la méthode DDSPCE pour traiter des scénarios variés et complexes.

Après avoir étudié les effets des incertitudes sur la sécurité statique des systèmes électriques et sur la DE, la thèse se concentre sur leur atténuation. À cette fin, une analyse de sensibilité globale est réalisée pour identifier les principales sources d'incertitude et aider à la conception de mesures de contrôle. Plusieurs modèles basés sur le PCE sont développés et comparés pour cette analyse de sensibilité, en particulier dans le contexte de l'amélioration de la capacité de transfert et de la DE. Grâce aux informations obtenues de cette analyse, des stratégies de contrôle des incertitudes, comme l'utilisation de systèmes de stockage d'énergie, peuvent être élaborées, réduisant ainsi les effets des incertitudes. Ces découvertes fournissent des orientations précieuses pour la gestion des incertitudes et la conception de contrôles dans les opérations réelles des systèmes électriques.

Dedicated to My Grandmother

Acknowledgments

First and foremost, I would like to convey my profound appreciation to my supervisors, Professor Xiaozhe Wang and Professor François Bouffard, for their unwavering support and guidance throughout my Ph.D. studies. In particular, I am profoundly thankful to my primary supervisor, Professor Xiaozhe Wang. She provided me with invaluable professional guidance and valuable suggestions as well as financial support for my Ph.D. studies and personal development. Also, I would like to express my great thanks for her encouragement to be confident, whether during exams, research seminars, public speaking occasions, or job searching. Her encouragement played a pivotal role in building my self-assurance throughout this journey. I am truly appreciative of her genuine concern for our well-being within and beyond academic research.

I am very grateful to my co-supervisor, Professor François Bouffard, who offered continuous support and guidance throughout my research journey. His expertise in power system analysis has significantly enriched my knowledge and understanding of the power area. Professor Bouffard consistently provided thoughtful, insightful, and useful comments and suggestions, which played a pivotal role in my Ph.D. research.

I would like to express my deepest gratitude to my Ph.D. committee members, Professor Archer Yi Yang, Professor Ioannis Psaromiligkos, and Professor William Leif Hamilton. Their astute questions, insightful suggestions, and constructive comments during my Ph.D. qualifying exam, plan proposal, and research seminar have been instrumental in refining my research direction and approach. Special thanks to Professor Psaromiligkos for his exceptional teaching in the course ECSE 509, which significantly enhanced my understanding of probability theory. Additionally, I extend my heartfelt gratitude to Professor Hamilton for his expertise in teaching COMP 551, which greatly advanced my knowledge of machine learning.

I am very grateful to Professor Lijun Sun for his role as the Internal Examiner and to Professor Yijun Xu for his contributions as the External Examiner of my thesis. Their thoughtful and insightful feedback has significantly enriched the quality of my thesis. I would also extend my sincere thanks to my oral defence committee members Professor Lijun Sun, Professor Louigi Addario-Berry, Professor Benoit Champagne, and Professor Changhong Cao. Their expertise and insightful questions have been indispensable in the successful completion of my defence and in improving my thesis's quality.

Additionally, I would like to express my deepest gratitude to Professor Peter Caines for his exemplary instruction in course ECSE 516, which enriched my knowledge and understanding of hybrid system optimization control. His continuous support and guidance, both during the course and throughout my job hunting, have been indispensable. I am truly thankful for the knowledge and assistance he has provided.

My heartfelt appreciation goes to all the staff of the Electrical and Computer Engineering Department for their assistance in ensuring that I can smoothly complete my Ph.D. program. I would also like to extend my sincere thanks to the McGill Thesis Writing Lab for reviewing my thesis and for invaluable suggestions to enhance its quality.

My appreciation must also go to the financial support I have received during my Ph.D. studies. Specially thanks for the support from the Natural Sciences and Engineering Research Council of Canada (NSERC), Fonds de Recherche du Quebéc-Nature et Technologies (FRQNT), McGill Engineering Doctoral Awards (MEDA) program, and the FRQNT Ph.D. Scholarship program.

I would like to extend my heartfelt appreciation to my friends and colleagues at the Power Engineering Group of McGill University. Our shared moments have been invaluable, and I am deeply thankful for the assistance they've provided and the great time we've shared. Special mentions go to, Dr. Hao Sheng, Dr. Haihao Jiang, Dr. Chu Sun, Dr. Yuchong Huo, Dr. Sanja Dzeletovic, Dr. Jinling Qi, Dr. Liqun Yang, Dr. Bingyun Li, Dr. Rongpeng Liu, Dr. Xun Gong, Dr. Georgia Pierrou, Dr. Ilias Zenelis, Dr. Jinpeng Guo, Sakirat Wolly, Jingyu Liu, Yaojie Cai, Mingqiu Du, Niño Christopher Ramos, Chuang Qi, Junyuan Zhao, Mohan Du, Sawsan Sayed, Xingjian Wu, Wei Huang, Yiping Liu, Aidan Gerkis, Mahsa Ebadat-Parast, and Kaan Gun. In particular, I would like to thank Dr. Hao Sheng for his technical guidance in starting my first project at McGill University. His insights were instrumental in ensuring a smooth transition into my Ph.D. research. In addition, my sincere thanks go to Dr. Rongpeng Liu, whose meticulous review of my thesis and insightful feedback greatly enhanced its quality.

I would also express my sincere gratitude to my friends at McGill University, Dr. Deng Mao, Jinsong Zhang, and Ningbo Zhu. Special thanks to Dr. Deng Mao for her invaluable assistance in ECSE 509 and for generously sharing useful information and materials with me. Also, I would like to thank Jinsong Zhang and Ningbo Zhu for their help and assistance in the COMP 551 course.

I would also like to thank my dear Words On Wings (WOW) Club members for their kind help and assistance in improving my public presentation skills. Specially, I would like to express my greatest thanks to my friend, Kit, for her support and guidance in my oral defence preparations.

I would like to extend a special thank you to my former roommate, Ariel (Yue Zou). I deeply appreciate the moments we shared, especially during the challenging times of the Covid pandemic. I would like to express my heartfelt gratitude to my friend, Ji Zhong. Her encouragement throughout my Ph.D. studies has been invaluable.

Most of all, my Ph.D. studies would not have been completed without the support of my family. I would like to express my wholehearted appreciation to my family for their belief in me, the constant support, and the unconditional love they showered upon me. My gratitude to them goes beyond what words can convey.

Preface

In the interest of full disclosure, this thesis mainly includes three journals with the candidate being the first author. The candidate's specific contributions involved devising the solutions, conducting the simulations, analyzing and presenting the results, and drafting the papers and response letters. Furthermore, the candidate worked in close collaboration with the supervisors to define the research problems. The contributions of the candidate and other co-authors for each publication are elaborated upon below.

Journal Articles

[J1] **Xiaoting Wang**, Xiaozhe Wang, Hao Sheng, and Xi Lin. A Data-Driven Sparse Polynomial Chaos Expansion Method to Assess Probabilistic Total Transfer Capability for Power Systems with Renewables. IEEE Transactions on Power Systems, vol. 36, no. 3, pp. 2573 - 2583, May 2021.

** Also presented at the 2022 IEEE Power & Energy Society General Meeting (PESGM) **

The candidate defined the problem, devised the solution, built and carried out the numerical studies, performed the data analysis, and wrote the paper. Professor Xiaozhe Wang contributed to supervising the candidate, providing constant technical support including problem definitions, paper modifications, addressing the reviewers' comments, and financial assistance in completing the paper publication. Professor Hao Sheng provided technical support regarding the use of DSATools and the calculation of transfer capability and helped in responding to some of the reviewers' queries. Dr. Xi Lin suggested considering the line trip in the transfer capability assessment.

[J2] **Xiaoting Wang**, Rongpeng Liu, Xiaozhe Wang, Yunhe Hou and François Bouffard. A Data-Driven Uncertainty Quantification Method for Stochastic Economic Dispatch. IEEE Transactions on Power Systems, vol. 37, no. 1, pp. 812-815, Jan. 2022.

The candidate defined the problem, devised the solution, built and carried out the numerical studies, performed the data analysis, and wrote the paper. Dr. Rong-Peng Liu provided the training data for the economic dispatch problem and assisted in formulating the mathematical models and validating the proposed solutions. Professor Xiaozhe Wang played a pivotal role in supervising the candidate, offering consistent technical guidance, ranging from defining problems and refining papers to addressing reviewers' inquiries. She also provided financial support for the paper's publication. Professor Yunhe Hou offered valuable editorial insights. Professor François Bouffard was responsible for supervisory guidance to the candidate, ensuring clarity in the research context and the paper's objectives.

[J3] **Xiaoting Wang**, Rong-Peng Liu, Xiaozhe Wang, François Bouffard. A Comparative Study of Polynomial Chaos Expansion-Based Methods for Global Sensitivity Analysis in Power System Uncertainty Control. To appear in IEEE Transactions on Circuits and Systems II: Express Briefs.

The candidate defined the problem, devised the solution, built and carried out the numerical studies, performed the data analysis, and wrote the paper. Dr. Rong-Peng Liu provided the training data essential for the second case and helped validate the proposed solutions. Professor Xiaozhe Wang played a pivotal role in supervising the candidate, offering consistent technical guidance, ranging from defining problems and refining papers to addressing reviewers' inquiries. She also provided financial support for the paper's publication. Professor François Bouffard was responsible for supervisory guidance to the candidate and offered valuable editorial insights.

Conference Papers

[C1] Jingyu Liu, **Xiaoting Wang**, Xiaozhe Wang. A Sparse Polynomial Chaos Expansion-Based Method for Probabilistic Transient Stability Assessment and Enhancement. The 2022 IEEE Power & Energy Society General Meeting (PESGM), Denver, CO, USA, 2022, pp. 1-5.

Jingyu Liu contributed to defining the problem, proposing the solution, and the mathematical formulation. He also constructed and executed the case studies, performed the data analysis and visualization, and drafted the paper. The candidate offered technical insights and editorial recommendations. Professor Xiaozhe Wang provided technical guidance supervised both Jingyu Liu and the candidate, and contributed to writing and revising the paper.

Contents

List of Figures

List of Tables

List of Acronyms

AC	Alternating Current
ADC	Available Delivery Capability
ANCOVA	ANalysis of COVAriance
ANOVA	ANalysis of VAriance
ATC	Available Transfer Capability
CBM	Capacity Benefit Margin
CDF	Cumulative Distribution Function
CPF	Continuation Power Flow
DDSPCE	Data-Driven Sparse Polynomial Chaos Expansion
DFIG	Doubly Fed Induction Generator
DSM	Direct Search Method
DC	Direct Current
ED	Economic Dispatch
ESSs	Energy Storage Systems
ETC	Existing Transmission Commitments
EVs	Electric Vehicle
FERC	Federal Energy Regulatory Commission
GPR	Gaussian Process Regression
GSA	Global Sensitivity Analysis
HDMR	High-Dimensional Model Representation
HPs	Heat Pumps
IEGS	Integrated Electricity and Gas System
KLE	Karhunen-Loève Expansions
LAR	Least Angle Regression

LHS	Latin Hypercube Sampling
LSA	Local Sensitivity Analysis
LSS	Latin Supercube Sampling
MC	Monte Carlo
MCS	MC Simulations
MG	Microgrid
NERC	North American Electric Reliability Council
NN	Neural Network
NT	Nataf Transformation
OASIS	Open Access Same-time Information System
OLS	Ordinary Least Squares
OMP	Orthogonal Matching Pursuit
OPF	Optimal Power Flow
PCA	Principal Component Analysis
PDF	Probability Density Function
PEM	Point Estimation Method
PCE	Polynomial Chaos Expansion
PPF	Probabilistic Power Flow
PTTC	Probabilistic Total Transfer Capability
QMC	Quasi-Monte Carlo Sampling
RESs	Renewable Energy Sources
RPF	Repeated Power Flow
RT	Rosenblatt Transform
SED	Stochastic-Optimization ED
TRM	Transmission Reliability Margin
TTC	Total Transfer Capability
UC	Unit Commitment
ULTC	Under-Load Tap Changer
UQ	Uncertainty Quantification
VSAT	Voltage Security Assessment Tool

List of Symbols

The primary symbols used in this dissertation are outlined below. Additional symbols may appear throughout the dissertation and will be defined in context as they arise.

Y	Scalar stochastic response
$\boldsymbol{\zeta}$	Vector of random inputs
$G(\boldsymbol{\zeta})$	Stochastic model
$\mathbb{R}$	Real number
$\mathbb{N}$	Natural number
$\mathcal{M}$	Dimension of random inputs
$\mathbb{E}[\cdot]$	Mathematical expectation operator
Ω	Support of $\boldsymbol{\zeta}$
Ω_j	Support of ζ_j
$f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$	Joint PDF of $\boldsymbol{\zeta}$
$F_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$	Joint CDF of $\boldsymbol{\zeta}$
$\mathbb{H}_{\boldsymbol{\zeta}}$	Hilbert space
$G^{\mathrm{pc}}(\boldsymbol{\zeta})$	PCE model
c_k	The k-th PCE coefficient
$\Psi_k(\boldsymbol{\zeta})$	Multivariate orthogonal polynomial basis
H	The order of PCE model
$\phi_j^{(\alpha_k^j)}(\zeta_j)$	Univariate orthogonal polynomial basis of a single input ζ_j
α_k^j	The degree of the univariate polynomial basis for ζ_j at the expansion term k
$\boldsymbol{\alpha}_k$	The multivariate index of $\Psi_k(\boldsymbol{\zeta})$, e.g., $\boldsymbol{\alpha}_k = (\alpha_k^1, \cdots, \alpha_k^{\mathcal{M}}) \in \mathbb{N}^{\mathcal{M}}$
$\boldsymbol{Z}$	Decorrelated random input vector
$\boldsymbol{C}$	PCE coefficients vector
M_p	Number of training samples to build PCE model

$\boldsymbol{\Psi}$	Matrix of multivariate orthogonal polynomial basis Ψ_k
$\boldsymbol{\theta}$	Voltage angles
$\boldsymbol{V}$	Voltage magnitudes
$\boldsymbol{x}$	State vector, e.g., $\boldsymbol{x} = [\boldsymbol{\theta}, \boldsymbol{V}]^T$
$\boldsymbol{b}$	The load-generation variation vector, indicating the direction of power transfer
λ	Transfer capability
N	Number of buses in a transmission system
$\boldsymbol{v}$	Vector of wind speeds in m/s
$P_w(\boldsymbol{v})$	The real power output at wind generator w in MW
$\boldsymbol{r}$	Vector of solar radiations in $\mathrm{W/m^2}$
$P_{pv}(r)$	The real power output at solar generator pv in MW
$\boldsymbol{P_L}$	Vector of load variations
$\boldsymbol{\rho}$	Vector of branch states
ρ_n	State of branch n
$\boldsymbol{U}$	Vector of random inputs, $\boldsymbol{U} = [\boldsymbol{v}, \boldsymbol{r}, \boldsymbol{P_L}, \boldsymbol{\rho}]$
$\mathbb{P}\{\cdot\}$	The probability of event $\{\cdot\}$
λ^{TTC}	Total transfer capability (TTC)
l	$l = \alpha_j^k$, the degree of $\phi_j^{(\alpha_k^j)}(\zeta_j)$ for ζ_j at the expansion term k
$p_{n,j}^{(l)}$	The coefficient of univariate polynomial basis $\phi_j^{(l)}$ in n-th degree
μ_n, j	The n-th raw moment of random input ζ_j
$\psi_j^{(l)}(\zeta_j)$	The one-dimensional orthonormal polynomial basis of ζ_j
e_{cloo}	The corrected leave-one-out cross-validation error
e_{loo}	The leave-one-out cross-validation error
M_s	Number of validation samples
μ	Mean value
σ	Standard deviation
Q	The objective function (minimum cost) of economic dispatch (ED) problem
$\boldsymbol{P_g}$	Vector of conventional generator outputs
$\boldsymbol{P_w}$	Vector of wind generator outputs
G_0	Constant term in the High Dimensional Model Representation (HDMR)
$G_j(\zeta_j)$	The effect of a single random input ζ_j on Y
$G_{j,d}(\zeta_j, \zeta_d)$	The effect of the combined random inputs ζ_j and ζ_d to Y
$\mathrm{Var}[\cdot]$	The mathematical variance operator

$\mathcal{A}$	Random inputs index set $\mathcal{A} = \{1, \cdots, \mathcal{M}\}$
$\emptyset$	Empty set
$\mathrm{Cov}[\cdot]$	Mathematical covariance
S_j	The ANCOVA (ANalysis of COVAriance) global sensitivity indices
$S_j^{(U)}$	The uncorrelated effect of ζ_j on $\mathrm{Var}[Y]$
$S_j^{(C)}$	The correlated effect of ζ_j to $\mathrm{Var}[Y]$
$G_j^{\mathrm{pc}}(\zeta_j)$	Including the terms $\Psi_{\alpha_k^j}$ that only depend on a single input variable ζ_j
$G_{j,d}^{\mathrm{pc}}(\zeta_j, \zeta_d)$	Including the terms $\Psi_{\boldsymbol{\alpha}_k^{j,d}}$ relying on inputs ζ_j and ζ_d
$\widehat{\mathbb{E}}[\cdot]$	Estimated mean value
$\widehat{\mathrm{Var}}[\cdot]$	Estimated variance
$\widehat{\mathrm{COV}}[\cdot]$	Estimated covariance

Chapter 1

Introduction

1.1 Background and Motivation

1.1.1 Challenges of Modernizing Power Systems

Modern power systems are being driven by the increasing need to promote a sustainable environment and reduce reliance on fossil fuels [10], [11]. This transition towards a greener and more renewable energy future is characterized by a shift towards renewable energy sources (RESs) [12], [13]. These RESs possess several advantages, including reduced greenhouse gas emissions [14], improved energy security [15], and enhanced power systems resilience [16]. In recent years, the integration of volatile RESs, such as wind and solar into power systems has experienced remarkable growth [1], [2]. This is evidenced by the substantial increase in renewable generation capacity worldwide (see Fig. 1.1 and Fig. 1.2). For instance, in 2022, solar energy continued to expand capacity with an additional 192 GW [1], followed by wind energy with 75 GW [2]. This significant growth in renewable capacity reflects the global commitment to achieving net-zero carbon emissions and transitioning to a more sustainable energy mix.

However, the intermittent and stochastic nature of wind and solar, caused by various factors, including weather conditions, geographical location, and technological limitations, impact wind and solar farms [17]. This stochastic behavior, characterized by fluctuations in wind speed and solar irradiance, directly affects the power output, leading to volatility and difficulties in accurate prediction and a higher level of uncertainty in power systems [18], [19]. Moreover, the integration of emerging energy demands (e.g., heat pumps (HPs) [20] and electric vehicles (EVs) [21]) introduces additional uncertainties in load forecasting, despite advancements in forecasting technologies. Fur-

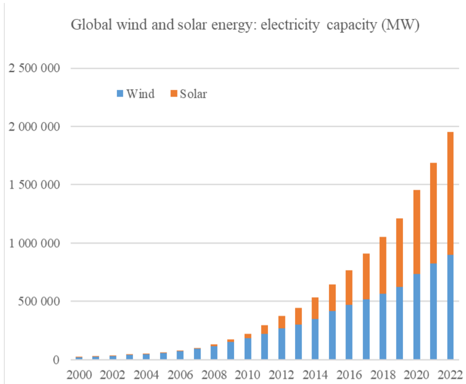

Fig. 1.1 Global installed wind and solar energy: electricity capacity [1]

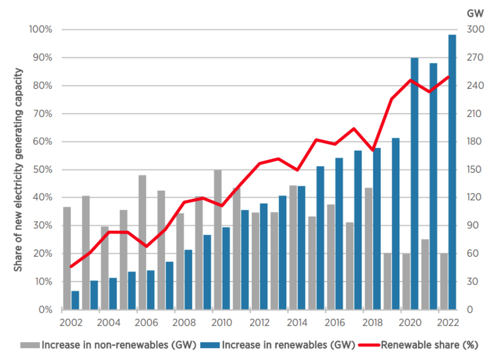

Fig. 1.2 Renewable share of annual power capacity expansion [2]

thermore, the aging power infrastructure raises the likelihood of asset outages, further complicating and adding uncertainty to power systems [22], [23].

The increased uncertainties have posed a myriad of challenges to power systems, particularly during highly stressed operating conditions when power systems are operating close to their limits or facing contingencies [24]. These challenges encompass voltage instability [25, 26], transient stability issues [27], and even system collapse [28]. In such situations, it is crucial to ensure the security of the power system while considering various sources of uncertainty (such as wind and solar power fluctuations, load variations, and equipment failures) and economic factors (minimizing costs, optimizing resource utilization, and maximizing efficiency) [29].

In this regard, the development of accurate and fast uncertainty quantification (UQ) and management tools and technologies is essential. These tools and technologies should be capable of providing real-time information, such as security margins, system state estimation, contingency analysis, and optimization capabilities [30]. By quantitatively measuring the impacts of uncertainties on power systems and mitigating their adverse effects, these tools can furnish system operators with valuable information. This, in turn, assists them in making well-informed and real-time decisions that strike the right balance between ensuring system security and promoting economic efficiency.

1.1.2 Power System Security Assessment

Power system security is crucial for the reliable and stable operation of power systems. It is defined as the ability to maintain power supply and adhere to operational limits, ensuring system safety in the face of unforeseen events (e.g., equipment failures, fuel shortages, or demand fluctuations) [31, 32]. Power system security assessment involves conducting an analysis to evaluate the extent to which a system is reasonably protected from significant disruptions to its operation [33]. In a broad sense, the security assessment in power systems can be classified into two types: static and dynamic [34]. The static security assessment checks for violations of equipment, thermal, and voltage limits in steady-state conditions following disturbances. On the other hand, the dynamic security assessment evaluates the system's ability to reach a new stable equilibrium operating point after a disturbance.This thesis primarily focuses on *static security assessment*, thereby not delving into the research challenges related to dynamic security assessment.

The increasing uncertainties impact power system static security in many aspects, which encompass the stochastic optimal power flow [35], stochastic power flow calculations [36], stochastic economic dispatch [37], security margin [38], transfer capability [39]). Therefore, there is a pressing need for more refined methods for power system static security assessment considering uncertainties. Specifically, the first part of this thesis considers evaluating the impacts of uncertainties on *available transfer capability* (ATC). This capability denotes the maximum power that can be transmitted between two nodes through a subset of transmission lines without compromising system security limits for further commercial uses beyond existing commitments [4]. The determination of ATC is closely related to another transfer capability concept, *total transfer capability* (TTC), which is defined as the maximum power that can be securely transferred between two nodes through a subset of transmission line considering all possible contingencies [4].

Conventionally, deterministic approaches have been widely used in power system security assessment [40–44]. These methods generally provided deterministic security criteria, for example, operational limits of devices (e.g., lines, buses, transformers, and generators) or fixed security margins after severe contingencies (e.g., equipment outages like lines, transformers, generators). Yet, the deterministic security assessment shows its inability to comprehensively account for uncertainties, often failing to capture the full spectrum of potential operating conditions, uncertainties (e.g., from RESs and loads), and contingencies that can impact system security. As a result, it may provide underestimated or overestimated risks, security margins, and vulnerabilities, potentially leading to suboptimal operational decisions and insufficient mitigation strategies in the face of unforeseen events. This highlights the limitations of traditional deterministic power security assessment in power systems [29] , prompting the adoption of probabilistic methods to deal with challenges brought by uncertainties and contingency with probabilities, which enable security assessment at a predefined confidence level.

1.1.3 Economic Dispatch

This thesis also addresses another critical aspect, the economic dispatch (ED) problem, which plays a pivotal role in power system operation. The ED problem involves the allocation of the overall demand among various generating units while minimizing the production costs [32]. Mathematically, ED is formulated as a constrained optimization problem. The objective function typically aims to minimize the generation costs, while adhering to various physical and network constraints (e.g., power flow constraints, voltage limits, thermal limits, and generator capacity constraints). Unlike conventional controllable power generations (e.g., thermal and reservoir hydroelectric power), variable RESs (e.g., wind power) are easily affected by environmental weather conditions, exhibiting highly uncertain properties. See Fig. 1.3 for the wind and solar generation outputs with uncertainty. The power outputs of RESs generation, such as wind and solar power, cannot be controlled in the same manner as fossil-fueled power generation, which is characterized by high controllability and dispatchability. Furthermore, the short-term variations of these RESs cannot be predicted with absolute precision. Uncertainties persist regarding the quantity of power that these sources will deliver in the upcoming hour or day. [45], [46].

The uncertain power outputs from variable RESs will lead to different dispatch solutions and operating costs [47] for a power system. The second part of this thesis aims to quantify the impacts of these uncertainties on the ED problem. More detailed literature reviews will be presented in the following section.

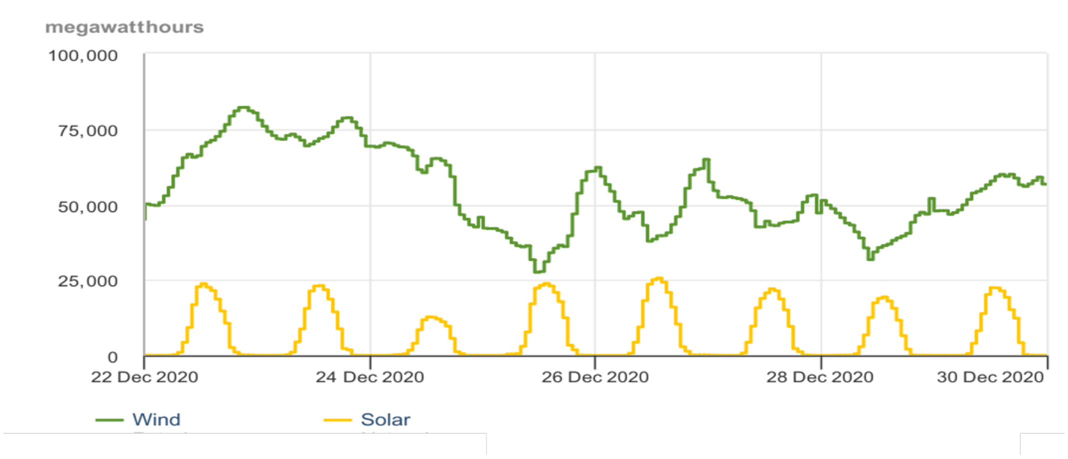

Fig. 1.3 Uncertainty in wind and solar generations power output [3]

1.2 Literature Review

This section delves into the methods used for quantifying uncertainty in power system static security assessment in Section 1.2.1 and ED in Section 1.2.2. Additionally, Section 1.2.3 offers a succinct overview of current strategies to mitigate the effects of uncertainties.

1.2.1 Uncertainty Quantification in Power System Static Security Assessment

As discussed in Section 1.1.2, conventional deterministic static security assessment methods may struggle with evolving system conditions and increasing uncertainties in power systems. This section provides an overview of UQ methods and their associated challenges in power system static security assessment considering uncertainties.

Monte Carlo Simulations

The Monte Carlo (MC) simulations rely on repeated random sampling, which is the most common and direct sampling-based method designed to assess uncertainties in power system security assessment [48–50]. The MC simulations frequently serve as a benchmark in evaluating and comparing alternative probabilistic methods. Intriguingly, while the accuracy of MC simulations remains independent of the the dimension of random inputs, it hinges critically on the number of stochastic samples deployed and the complexity level of systems [51]. A typical requirement involves the execution of tens of thousands of MC simulations to attain high precision. This significant computational demand remains a key limitation of MC simulations, even when advanced sampling techniques (e.g., Latin hypercube sampling [52, 53] (LHS), Latin supercube sampling (LSS) [54] and Quasi-MC Sampling (QMC) [55, 56]) are employed to enhance computational speed, compromising its applicability in real-time applications.

Analytical Methods

To reduce the computational time, analytical methods have been proposed for power system static security assessment. The fundamental concept of these methods involves utilizing specific algorithms, such as convolution techniques [57] or cumulant techniques [58] to estimate the probability density function (PDF) and cumulative distribution function (CDF) of the system response (e.g., PPF solutions [59, 60]). Among the analytical techniques, convolution methods involve the convolution of all random variables, which can be computationally intensive and require significant memory storage. In contrast, cumulant-based methods leverage properties of series expansions, (e.g., Gram-Charlier [61] and Cornish-Fisher [59]). These methods depend on the probability distributions of random inputs and are less computationally demanding than MC simulations, yet they maintain a commendable level of accuracy. However, these methods have some limitations since they often treat random input variations as minor perturbations around equilibrium points. This assumption facilitates the linearization of the power system model, while it may not be true in practical applications and this perspective can limit their applicability in real-world applications, especially when dealing with large-scale and nonlinear systems [57, 59–61].

Point Estimation Methods

The point estimation method (PEM) offers an alternative to improve computational efficiency compared with MC simulations for uncertainty quantification. PEM addresses uncertainties by determining the statistics and PDF of stochastic responses through the computation of their statistical moments [62–64]. As such, PEM has lowered the computation burden caused by MC simulations. Su et al. proposed the PEM method to solve the probabilistic power flow (PPF) [65] and transfer capability [66]. Later, a 2PEM (two-point estimation method) was applied to account for uncertainties in the optimal power flow (OPF) problem. However, it might not capture all the nuances of uncertainties, especially when there is a large dispersion in uncertain variables and with the Gaussian assumption of the distribution of uncertainties [67]. A discrete PEM method was developed by combining the Gram–Charlier expansion technique [68] to solve the PPF problem, enabling the free assumption of probability distributions of random inputs. Despite its efficiency, the higher moments of responses derived using PEM often lack accuracy. This can result in less accurate estimations of probability distributions and response statistics since the statistics and PDFs are derived based on estimated input moments [69]. Additionally, it may not be suitable for responses with thick-tailed distributions, as PEM-based methods may struggle to provide accurate estimates of higher-order moments, such as kurtosis and skewness [70].

Surrogate Models

In the realm of power system security assessment, surrogate modeling techniques have been introduced as a popular means to alleviate the computational demands associated with MC simulations. Surrogate models offer many advantages in power system security assessment. They offer superior computational efficiency compared to MC simulations, especially for systems with a moderate number of uncertainties, without sacrificing accuracy. In contrast to the PEM method, surrogate models provide detailed statistics (e.g., mean and variance) and a full spectrum of probabilistic response distributions. Moreover, in comparison to the analytical method, they excel in managing systems with pronounced non-linearities, eliminating the need for linearization of system models. Examples of these techniques include polynomial chaos expansion (PCE) [38] [71], Gaussian process regression (GPR) [72, 73], and neural network (NN)-based models [74, 75], among others. PCE, in particular, is a prominent example that provides a computationally-efficient approximation. Compared with GPR, it may provide more accurate statistics (e.g., mean and standard deviations) and the entire shape and the tail of probability distributions. Furthermore, PCE may outperform GPR in estimating noisier and more multi-modal distributions [76]. PCE is represented as a weighted sum of orthogonal polynomials of uncertainties, which are regarded as random inputs with predefined probability distributions [6] [77]. This model can be constructed with a limited number of evaluations, capturing the primary physical model's input-output mapping. Notably, PCE-based models allow for the direct extraction of the system response's sample mean and variance from the weights of polynomials [78].

Given its benefits, PCE has found applications both within and outside the power community (e.g., fields like Civil Engineering [79–82] and Mechanical Engineering [83–85]). Within the power community, PCE's applications span across various facets of power system security assessment [23, 86–93]. Yet, the computational efficiency may be limited for high dimensional uncertainties due to the "curse of dimensionality" issue and correlation exist between random inputs may complicate the UQ [23, 87, 88]. Despite these issues have been alleviated in [89, 93] to some extent, a common limitation among these methods is the assumption that all random variables adhere to specific parametric distribution functions. In real-world power system applications, the knowledge of probability distributions of random inputs might be limited or inaccurate, while raw data, like wind speed and solar radiation, is more readily available.

Overall, as delineated in this section, existing UQ methods for examining the impacts of uncertainties on power system static security, especially the ATC, exhibit several research gaps. These include:

1. The escalating integration of RESs amplifies the dimensionality of uncertainties, rendering many UQ techniques computationally intensive.

2. The presence of correlations among uncertainties complicates the achievement of accurate UQ results.

3. Dealing with uncertainties that have unknown distributions is a challenging task.

4. Only limited work has considered the impacts of uncertainties brought about by unexpected equipment outages on power system security assessment.

In essence, there is a lack of a systematic method that can effectively and accurately assess ATC while addressing the limitations mentioned above. The current landscape necessitates a thorough exploration of uncertainties brought by RESs, load variations, and unexpected equipment outages in ATC assessment. To this end, there is a pressing need for the development of a data-driven UQ method that is both computationally efficient and highly accurate, to reinforce real-time ATC assessment considering uncertainties.

The first part of this thesis (Chapter 3) addresses the challenges of UQ in power system security assessment, with a spotlight on available transfer capability evaluation. Chapter 3 also offers an in-depth literature review on transfer capability assessment.

1.2.2 Uncertainty Quantification in Economic Dispatch

Power system operators employ an OPF to fulfill the power demand at the lowest possible cost, first involving selecting a subset of generators that can satisfy the demand for a specific time period, typically ranging from 1 to 2 days. That is the so-called *unit commitment* (UC). Once the UC decisions are determined, the ED problem is solved to determine the optimal generation levels for the committed generators, aiming to meet the expected demand at the lowest economic cost [94]. This thesis considers that the ED problems are solved under a series of future time intervals (e.g., a day-ahead window with a one-hour interval), to provide a more proactive solution in the face of increasing uncertainties (e.g., arising from RESs). In this context, ED refers to the allocation of generation among different units to fulfill the power demand cost-effectively. It ensures adherence to physical and network constraints while also accounting for uncertainties. The detailed mathematical formulation of the ED problem is given in Chapter 4.

Given the significant integration of variable RESs in power systems, the impacts of these uncertainties in ED have become more pronounced. To address and quantify them, numerous formulations have been proposed in the existing literature. These can be broadly classified into three categories: 1) deterministic ED formulations; 2) robust-optimization ED formulations; and 3) stochastic-optimization ED formulations.

Deterministic ED formulations focus on minimizing operational costs based on pre-determined reserve capacities that account for selected realizations of the uncertainties arising from renewable generation power outputs and stochastic loads [95–97]. For instance, Chen et al. [96] proposed a direct search method (DSM) to coordinate wind and thermal generation dispatch, minimizing the production cost in ED. This approach uses fixed reserves to address uncertainties from wind power generation and has demonstrated high efficiency. Liang et al. [97] introduced a genetic algorithm for the ED utilizing fuzzy sets to represent uncertainties from stochastic loads and RESs. While this method is effective, it can be computationally intensive, especially for large-scale systems. Furthermore, issues may arise when the number of representations or the spread of potential uncertainties is insufficient to capture the true possible realizations of the uncertainties.

The robust-optimization formulations can serve as an alternative methodology for ED considering uncertainties [98–103]. The optimization objective of these formulations is to minimize the operational cost under the worst-case scenario. Accordingly, system operators can acquire the most severe consequences brought about by uncertainties within a predefined range. Sasak et al. [101] and Li et al. [102] developed confidence interval-based robust optimization formulations for ED, maximizing robustness against uncertainties associated with RESs (e.g., wind), which offered frequent real-time updates of generation schedules. However, the accuracy of these methods heavily relies on accurate forecasting of wind distributions, and any significant deviation in forecasts can affect its efficiency. Furthermore, these formulations set bounds on the variability of uncertainties and optimize for the worst-case scenario, resulting in overly conservative outcomes to ensure risk resilience. To address these issues, Ding et al. [103] proposed a multi-stage distributed robust optimization model directly using the historical wind data to address the uncertainties in the ED efficiently, which guaranteed the risk resilience of robust optimization. However, striking a balance between the economic efficiency and conservatism of dispatch solutions often relies on the personal experience of the system operators.

Stochastic-optimization ED formulations are designed to minimize the expected value of a loss function (e.g., operation cost) impacted by decision variables (e.g., generator power outputs) and by exogenous uncertainties (e.g., wind power) [104]. A prevalent method to address the uncertainties is the scenario-based approach, which relies heavily on multiple scenarios to represent the uncertainties in renewable generations [105–109]. Advanced methods have also been proposed to reduce the scenarios. For example, Feng et al. [110] introduced a parallel dual-DQAM to decouple the multi-scenario ED problem both in terms of scenarios and time periods, which can significantly reduce computational burden and enhance solving efficiency. However, these approaches assume that these uncertainties conform to known/inferred probability distributions. While the stochastic-optimization formulations offer cost-effectiveness and can provide high-risk resilience against uncertainties, they present two challenges. The first challenge is the accuracy and efficiency trade-off: the accuracy of this method is directly related to the number of scenarios. Employing a vast number of scenarios can lead to reduced computational efficiency. This poses a dilemma in achieving a balance between solution accuracy and computational speed, an area that warrants further exploration. The second challenge arises in practical applications, where obtaining an accurate PDF for renewables is impractical. The distributions assumed or inferred (e.g., wind speed follows Weibull distribution, solar radiation follows Beta distribution) may not represent the actual data [111, 112], causing inaccurate uncertainty quantification results in ED.

As pinpointed in this section, a comprehensive UQ method for ED that is computationally effective while maintaining accuracy becomes a critical task considering the increasing penetration of RESs (e.g., wind power). Key research gaps bypass:

1. Current stochastic-optimization ED formulations typically demand a large number of scenarios to achieve good accuracy, which reduces computational efficiency.

2. The increasing uncertainty dimension further lowers the computation efficiency, hindering the applications of many UQ techniques.

3. The unknown distribution and correlation information of uncertainties also increase the difficulty in UQ on ED.

Thus, a comprehensive data-driven UQ method, that is computationally effective and highly accurate, for ED in power systems under high dimensional uncertainties is needed. Therefore, Chapter 4 centers on the impacts of uncertainties on ED, in particular, dedicated to addressing the limitations on UQ based on stochastic-optimization ED formulations.

1.2.3 Global Sensitivity Analysis-Based Uncertainty Control in Power System Static Security and Economic Dispatch

As discussed in Section 1.2.1- 1.2.2, rising levels of uncertainties have profound implications for power systems static security and ED. Several methods have been proposed to mitigate the impacts of uncertainties, which can be categorized into two types: with and without energy storage units. For example, direct control methods, such as pitch angle control [113] and inertia control [114] have been applied to smooth the wind generator power outputs, without the need for energy storage units. However, these methods can increase system complexity and may not always harness the maximum wind power [115], decreasing the utilization of RESs.

To address these issues, energy storage systems (ESSs) have been used as buffers to flatten wind power fluctuations [116] [117]. Various energy storage technologies have been reviewed for their efficacy in mitigating the uncertainties of ESSs [118, 119]. However, the renewable uncertainties (i.e., random inputs) become high-dimensional, and it is not practical to install energy storage units for every random input factor (e.g., near wind farms). Therefore, how to manage the ESSs to mitigate the impacts of uncertainties brought by RESs became a key issue. This leveraged the utilization of different optimization models for energy scheduling and management [120–122]. For example, an interval optimization-based dispatch approach has been developed to handle the variability introduced by RESs via managing the system's flexibility [120]. However, these methods assumed predetermined distributions of RESs and load variations, which may be not true in practical applications.

Another uncertainty mitigation approach that has gained popularity stems from sensitivity analysis [123, 124]. Sensitivity analysis is classified into local and global methods. Local sensitivity analysis (LSA) examines the impact of variations in a single input parameter on the model output, focusing on a specific point in the input space, which assumes the model is linear around the point of interest (i.e., small perturbations around stable equilibrium points and determines a linear relationship between inputs in the local region around the nominal operating points) [125]. In contrast, global sensitivity analysis (GSA) assesses the effect of one or multiple variables on the output across the entire input parameter space, accounting for interactions among all inputs. By conducting a sensitivity analysis of uncertainties, critical inputs that will affect power system security and ED can be found. Many sensitivity analysis-based approaches have been proposed in power system applications. For example, LSA-based approaches have been proposed in load modeling for wide-area power systems [126] and drive train parameters identification of a doubly fed induction generator (DFIG) [127]. However, the LSA evaluates the sensitivity of a model's output to variations of its inputs at a specific point (e.g., a nominal value) and assumes a linear relationship between input and output changes, which may not be true for practical power systems.

Unlike LSA, GSA examines the effects across the entire spectrum of potential variations and incorporates the system's nonlinear characteristics (i.e., the effects of the variation of inputs on whole uncertain space on the system outputs). While MC simulations are commonly used for GSA, the high computational time makes it impractical [128]. To address this issue, Ye et al. [129] proposed a Gaussian process modeling method for GSA to measure the impacts of independent uncertainties (e.g., load variations) on the changes in voltage magnitude. Xu et al. [130] and Ni et al. [131] proposed a PCE surrogate GSA method to perform a priority ranking of correlated uncertainties from

RESs that affected power system voltage stability and load flow, respectively. Later, a PCE-based GSA method was proposed in [132] to rank the impacts of the maximum loadability of an islanded microgrid (MG). Based on the sensitivity information, an effective uncertainty mitigation measure has been designed. Nevertheless, the method applied requires accurate probability distributions of random inputs, which might not always be accessible in real-world scenarios. Additionally, in practical applications, random inputs (uncertainties) often exhibit correlations, e.g., temporal and spatial dependence observed between different wind speeds [133]. While some efforts have been made, conducting global sensitivity analysis of correlated random inputs is still challenging. Overall, discussions on uncertainty control related to ATC and ED might be lacking.

As emphasized in this section, discussions on uncertainty control related to power system security and ED might be lacking. Current uncertainty control methods present some limitations:

1. Current GSA methods often relying on MC simulations, may necessitate a substantial number of scenarios for accurate estimations.

2. The presence of correlations among uncertainties further complicates GSA.

3. Optimization-based or GSA-based uncertainty control methods require pre-assumed or inferred probability properties of uncertainties.

4. Uncertainty control measures based on GSA require fast and accurate GSA, which is still challenging.

In this regard, an effective control measure to mitigate the uncertainty impacts, which addresses the aforementioned limitations is required.

Specially, this thesis emphasizes and addresses the limitations of existing methods that use global sensitivity analysis as a tool to mitigate the effects of uncertainties on ATC and ED, as detailed in Chapter 5.

1.3 Research Objectives and Methodology

To effectively address the research questions identified in Section 1.2, it is crucial to establish the following primary research objectives:

- The first part of this thesis (Chapter 3) aims to address research gaps in power system static security assessment (see Section 1.2.1), focusing on ATC. It involves the development of a data-driven PCE-based method that directly utilizes data to assess ATC while considering various correlated uncertainties from RESs, load variations, and unexpected line outages.

- The second part of this thesis (Chapter 4) intends to fill gaps in research concerning ED problems under high-dimensional uncertainties (see Section 1.2.2). It involves the practical application of a data-driven PCE-based method to efficiently evaluate the objective value of ED problems, even in situations where uncertainty distributions and correlations remain unknown.

- The third part of this thesis (Chapter 5) is dedicated to the development of effective uncertainty control strategies for power system static security and ED to target research limitations mentioned in Section 1.2.3. Leveraging PCE-based methods for GSA, they aim to design uncertainty control measures based on the estimated sensitivity information to mitigate the impacts of uncertainties and enhance both power system security and economic efficiency.

1.3.1 Methodology

This thesis leverages surrogate modeling techniques, polynomial chaos theory, and global sensitivity analysis to quantify and mitigate the impacts of uncertainties in the probabilistic assessment of power system security and ED. A primary focus is the development of data-driven PCE methods to quantify and mitigate the effects of uncertainties on power systems security and ED.

PCE serves as a powerful tool to approximate stochastic responses of power system models [38,90]. It establishes statistically-equivalent relationships between uncertainties (inputs) and their corresponding responses (outputs), requiring only a small number of model evaluations (i.e., input-output sample pairs used to train models). By employing PCE-based surrogate models, one can swiftly evaluate the stochastic responses of power system models, such as transfer capabilities and the objective values of ED problems. The data-driven nature of this approach further enhances the feasibility of constructing PCE for large-scale power systems in real-world scenarios [38, 134].

Additionally, this thesis proposed tailored PCE-based methods for global sensitivity analysis, a crucial instrument for measuring the significance of an individual or a group of uncertainties. By extracting key insights from sensitivity data and harnessing energy storage technologies, effective control strategies are developed to alleviate the impacts of uncertainties on power system security and ED.

1.3.2 Research Tools

In this thesis, simulations were conducted using Matlab scripts, integrated with the Voltage Security Assessment Tool (VSAT) [135], a pivotal component of DSATools. DSATools is recognized as a robust suite of power system analysis tools, with a track record of effectively assessing power system security. However, current commercial power system analysis tools often dismiss the integration of uncertainties, particularly those arising from RESs, stochastic loads, and unexpected line outages. Therefore, an interface between Matlab and VSAT/DSATools was developed, enabling the integration of uncertainties in VSAT/DSATools. Based on this, VSAT was capable of evaluating the ATC of the power system accurately. MC simulations based on VSAT/DSATools are conducted to establish benchmark results.

The simulations of UQ and global sensitivity analysis for the proposed methods in this thesis were executed using Matlab scripts combined with UQLab Toolbox [136], which aims to make cutting-edge UQ techniques more universally accessible. The developed Matlab scripts were combined with the Polynomial Chaos Expansions [137] package, and the Sensitivity Analysis package [138] from UQLab, ensuring the attainment of the research objectives.

1.4 Claims of Originality

This thesis introduces several contributions in the field of data-driven methods for uncertainty quantification in power systems, specifically in the areas of probability transfer capability assessment, stochastic optimization, and uncertainty-aware control. The originality of these contributions, as demonstrated in the published papers, can be classified into three key perspectives.

- This thesis proposes a data-driven sparse polynomial chaos expansion (DDSPCE) method, which leverages available data of random inputs (e.g., RESs, random loads, and unexpected equipment outages), to study the impacts of uncertainties on power system ATC. The proposed method, requiring no preassumed probability distributions of random inputs, can accurately and effectively estimate the stochastic characteristics (mean, variance, PDF, and CDF) of probabilistic TTC (PTTC), based on which, the ATC with a certain confidence level can be readily calculated. An integrated sparse scheme further enhances its computational efficiency and accuracy. Numerical studies conducted on the modified IEEE 118-bus system and the modified PEGASE 1354-bus system demonstrate the DDSPCE method's efficacy in

PTTC and ATC evaluation. Furthermore, the results highlight the significance of incorporating discrete random inputs in PTTC and ATC assessment [J1].

- This thesis develops a DDSPCE-based surrogate model to investigate the impacts of uncertainties, particularly from wind power, on the ED during daily power system operations. This surrogate model is capable of estimating the probabilistic characteristics of ED solutions (e.g., the objective values), including their mean, variance, and distribution functions. Moreover, the proposed method can handle a vast number of random inputs without requiring predefined probability distributions. Extensive simulation results on an Integrated Electricity and Gas System (IEGS) using real-life wind power data, underscore the method's efficiency and effectiveness in quantifying the impacts of uncertainties on ED solutions, even when the ED solutions exhibit multimodality properties. These findings emphasize the DDSPCE method's efficacy and efficiency in tackling both general and intricate scenarios [J2].

- This thesis designs effective global sensitivity analysis-based uncertainty control to mitigate the impacts of uncertainties on power system security and ED. Particularly, different PCE-based methods are developed to conduct GSA, enabling to identification of the dominant random inputs. Leveraging the insights from the sensitivity information, this thesis designs uncertainty control strategies (e.g., by utilizing ESSs) for transfer capability enhancement and ED. These findings provide essential guidance for managing uncertainties and designing controls in practical power system operations [J3].

1.5 Thesis Outline

This thesis is structured as follows.

- **Chapter 2** presents the foundational mathematical concepts essential to the thesis. It covers the PCE method, including the basic concepts of PCE, strategies to build orthogonal polynomial bases, PCE coefficient calculations, providing a comprehensive understanding of these techniques.

- **Chapter 3** proposes a novel DDSPCE method for assessing PTTC and ATC. It begins by providing definitions of transfer capability and an overview of state-of-the-art probabilistic transfer capability assessment methods. The mathematical formulation of TTC based on continuation power flow (CPF) and the modeling of uncertainties (such as wind, solar, loads, and unexpected line outages) are presented. The proposed method utilizes a moment-based approach to construct the polynomial basis, leveraging the available data instead of relying on the probability distribution of random inputs. An integrated sparse scheme based on the least angle regression (LAR) and a modified truncation norm is exploited to enhance computational efficiency while maintaining accuracy. Chapter 3 is closely related to the publication [J1].

- **Chapter 4** extends the DDSPCE method to construct a statistically equivalent surrogate model for ED problems. It starts with existing surrogate model-based methods for solving the stochastic-optimization ED problems, followed by the mathematical formulation of the stochastic optimization ED formulation as a complex-constrained optimization problem. This chapter then elaborates on the proposed DDSPCE-based surrogate model to estimate ED solutions (e.g., objective values) by exploiting data directly. Simulation results on an IEGS using real-world wind power data are presented to validate the efficiency and effectiveness of the proposed method. The research presented in this chapter is closely linked to the publication [J2].

- **Chapter 5** presents a comprehensive investigation of various PCE-based methods for ANCOVA (ANalysis of COVAriance) indices-based global sensitivity analysis in power system security and ED considered correlated random inputs. This chapter begins with a thorough analysis of the existing literature in this field. Subsequently, a high-dimensional decomposition representation (HDMR) of power systems under uncertainty is introduced. Two distinct PCE-based techniques for computing the ANCOVA sensitivity indices are then detailed. Furthermore, this chapter then designs an efficient strategy for uncertainty control by leveraging the sensitivity information obtained from PCE-based methods to mitigate the impacts of random inputs on ATC and ED. The findings of this chapter offer valuable insights and guidance for the management and control of uncertainty in power system operation and security assessment. The work presented in Chapter 5 is closely related to the publication [J3].

- **Chapter 6** serves as a conclusion to the thesis, summarizing the main contributions and discussing potential areas for future research.

Chapter 2

Polynomial Chaos Theory

This chapter will introduce the *polynomial chaos expansion* (PCE) method, a widely recognized metamodel for uncertainty quantification in power systems and various other fields. The concept of PCE was initially introduced by Wiener [139], where the stochastic response was approximated using Hermite polynomials for Gaussian random inputs. This method was later generalized by Xiu et al. [6] to solve the stochastic differential equations with random inputs that follow some standard distributions (e.g., Beta, Uniform, and Gamma). Since then, PCE has been advanced by integrating adaptive sparse scheme [78], moment-based methods [140], response surface methodology [86], and Bayesian inference [38]. These PCE-based approaches have gained traction in power systems (e.g., PPF problems [71, 89, 90, 93], ED [37, 141], load-margin assessment [38], and dynamic simulations [142, 143]). These methods offer the dual benefits of reduced computational burdens and the ability to handle systems with pronounced non-linearities. Capitalizing on the strengths of PCE, this thesis presents tools for uncertainty quantification and control based on the data-driven PCE method. The subsequent sections provide the foundational concepts of the PCE method.

2.1 Polynomial Chaos Expansion

2.1.1 The Generalized PCE for Model Response

Consider a stochastic response model $Y = G(\boldsymbol{\zeta})$ with an input vector $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \cdots, \zeta_{\mathcal{M}}] \in \mathbb{R}^{\mathcal{M}}, \mathcal{M} \geq 1$, where inputs $\boldsymbol{\zeta}$ could be volatile renewables (e.g., wind speed and solar radiations), load variations, unexpected line outages, etc. Y may be power flow solutions (e.g., bus voltage magnitudes or angles), PTTC, or the solution to ED problems. Without loss of generality, let's consider Y as a scalar-valued output equipped with finite second-order moments [78], i.e.,

$$\mathbb{E}[Y^2] = \int_\Omega G^2(\boldsymbol{\zeta})dF_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) = \int_\Omega G^2(\boldsymbol{\zeta})f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})d\boldsymbol{\zeta} < +\infty \tag{2.1}$$

where $\mathbb{E}[\cdot]$ represents the expectation operator. Ω is the support of $\boldsymbol{\zeta}$;$F_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ is the the joint CDF and $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ is the joint PDF of $\boldsymbol{\zeta}$, respectively.

Remark 2.1.1. **Hilbert Space**: Let $\mathbb{H}_{\boldsymbol{\zeta}} = \mathcal{L}^2_{f_{\boldsymbol{\zeta}}}(\Omega, \mathbb{R})$ be the real Hilbert space associated with the probability measure $f_{\boldsymbol{\zeta}}$. This space is characterized by following inner product [144, 145]:

$$\langle u, v\rangle_{\mathbb{H}_{\boldsymbol{\zeta}}} = \int_\Omega u(\boldsymbol{\zeta})v(\boldsymbol{\zeta})f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})d\boldsymbol{\zeta} \tag{2.2}$$

Based on ***Remark*** 2.1.1, (2.1) is equivalent to $\mathbb{E}[G(\boldsymbol{\zeta})^2] = \langle G(\boldsymbol{\zeta}), G(\boldsymbol{\zeta})\rangle < +\infty$, indicating that stochastic responses with finite second-order moments belong to the Hilbert space $\mathbb{H}_{\boldsymbol{\zeta}}$.

As it was shown in [6, 139, 146], any stochastic response with finite second-order moments can be represented by a weighted sum of orthogonal polynomial basis functions of random inputs $\boldsymbol{\zeta}$:

$$Y = G(\boldsymbol{\zeta}) = G^{\mathrm{pc}}(\boldsymbol{\zeta}) = \sum_{k=1}^{+\infty} c_k\Psi_k(\boldsymbol{\zeta}) \tag{2.3}$$

where $G^{\mathrm{pc}}(\boldsymbol{\zeta})$ represents the PCE-based models, c_k are PCE coefficients, and $\Psi_k(\boldsymbol{\zeta})$ are multidimensional orthogonal polynomial bases, which are orthogonal with respect to the joint PDF $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ of random inputs $\boldsymbol{\zeta}$, i.e., $\Psi_k(\boldsymbol{\zeta})$ satisfies the following orthogonal condition:

$$\int_\Omega \Psi_k(\boldsymbol{\zeta})\Psi_m(\boldsymbol{\zeta})f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) = \gamma_{km}\delta_{km} \tag{2.4}$$

where γ_{km} is a positive constant and δ_{km} is the Kronecker delta, i.e., if $k = m, \delta_{km} = 1$ and if $k \neq m, \delta_{km} = 0$. Particularly, if the orthogonal condition in (2.4) is satisfied, $G^{\mathrm{pc}}(\boldsymbol{\zeta})$ converges in the sense of $\mathcal{L}^2$-norm according to the Cameron–Martin theorem [147], i.e.,

$$\lim_{L\to+\infty}\left(\left\|Y - \sum_{k=0}^{L-1} c_k\Psi_k(\boldsymbol{\zeta})\right\|_{\mathcal{L}^2}\right)^2 = \lim_{L\to+\infty}\mathbb{E}\left[\left(Y - \sum_{k=0}^{L-1} c_k\Psi_k(\boldsymbol{\zeta})\right)^2\right] \tag{2.5}$$

Note that (2.3) is impractical due to the summation of infinite terms. To tackle this issue, the

PCE in (2.3) is generally truncated with finite expansion terms [78]:

$$Y = G(\boldsymbol{\zeta}) \approx G^{\mathrm{pc}}(\boldsymbol{\zeta}) = \sum_{k=0}^{L-1} c_k \Psi_k(\boldsymbol{\zeta}) \tag{2.6}$$

where $L = \frac{(\mathcal{M}+H)!}{\mathcal{M}!H!}$, with $\mathcal{M}$ the dimension of random inputs and H the order of PCE model, i.e., the maximum degree of the polynomial basis functions. The key points to building a PCE model include the construction of polynomial basis functions $\Psi_k(\boldsymbol{\zeta})$ and the calculation of PCE coefficients c_k, which will be introduced in Section 2.1.2 and 2.1.3, respectively.

2.1.2 The Construction of Polynomial Bases

Independent Random Inputs

If random inputs $\boldsymbol{\zeta}$ have mutually independent components, the classical polynomial bases $\Psi_k(\boldsymbol{\zeta})$ can be produced by the full tensor product of the one-dimensional orthogonal polynomial bases $\phi_j^{(\alpha_k^j)}$ [145]:

$$\Psi_k(\zeta_1, \cdots, \zeta_{\mathcal{M}}) \;=\; \prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(\zeta_j) \tag{2.7}$$

$$\sum_{j=1}^{\mathcal{M}} \alpha_j^k \leq H, \qquad k = \{0, 1, \cdots, L-1\} \tag{2.8}$$

where $\Psi_k(\boldsymbol{\zeta}$ produced can easily satisfy the orthogonal condition in (2.4); α_k^j denotes the index of the j-th univariate polynomial basis ϕ_j at expansion term k, i.e., α_j^k is the degree of the univariate polynomial basis for random input ζ_j on the expansion term k. Equation (2.8) refers to a standard truncation scheme typically applied to truncate the PCE in (2.3).

Conventionally, in the generalized PCE, the univariate polynomial bases $\phi_j^{(\alpha_k^j)}$ are chosen based on the distribution types of random inputs as introduced by Xiu. et al. [6]. If random input ζ_j follows some typical distribution types, the univariate polynomial basis $\phi_j^{(\alpha_k^j)}(\zeta_j)$ can be selected by utilizing the Wiener–Askey polynomials (Table 2.1 [6]), where the optimal convergence rate is achievable. Particularly, an existing method, SPCE [36] selecting the univariate orthogonal polynomial basis ϕ_j based on Table 2.1 is used for comparison.

Once the univariate orthogonal polynomial bases ϕ_j are determined, the multivariate orthogo-

Table 2.1 Standard forms of continuous distributions and their corresponding Wiener–Askey polynomials [6]

Distribution	Polynomials	Support
Gaussian	Hermite	$(-\infty, +\infty)$
Uniform	Legendre	$[-1, +1]$
Beta	Jacobi	$[-1, +1]$
Gamma	Generalized Laguerre	$[0, +\infty]$

nal polynomial bases $\Psi_k(\boldsymbol{\zeta})$ can be constructed through (2.7) and (2.8). An example of the construction of multivariate polynomial bases $\Psi_k(\boldsymbol{\zeta})$ with selected univariate orthogonal polynomial bases $\phi_j^{(\alpha_k^j)}(\zeta_j)$ is given below (see Table 2.2). Consider a 3-dimensional random input vector $\boldsymbol{\zeta} = \{\zeta_1, \zeta_2, \zeta_3\}$, where $\zeta_j, j = 1, 2, 3$ follow standard Gaussian distribution. Then, the multivariate orthogonal polynomial bases $\Psi_k(\boldsymbol{\zeta})$ with the PCE order $H \leq 3$ (i.e., the possible degree of the polynomial basis functions could be $\{0, \cdots, 3\}$) are as follows:

Remark 2.1.2. Note that $\boldsymbol{\alpha}_k = (\alpha_k^1, \cdots, \alpha_k^{\mathcal{M}}) \in \mathbb{N}^{\mathcal{M}}$ denotes the multivariate index of $\Psi_k(\boldsymbol{\zeta})$, indicating how the individual polynomial basis $\phi_j^{(\alpha_k^j)}(\zeta_j)$ is combined to form the multivariate polynomial bases $\Psi_k(\boldsymbol{\zeta})$. The index $\boldsymbol{\alpha}$ can be regarded as a $L \times \mathcal{M}$ matrix, where each element (i.e., α_k^j) indicates the degree of the j-th univariate polynomial basis for expansion term k. E.g., in Table 2.2, $\mathcal{M} = 3$, $H = 3$, then $L = 20$ according to (2.8). As such, $\boldsymbol{\alpha} \in \mathbb{N}^{20\times 3}$, with $\boldsymbol{\alpha}_k$ corresponding to each row of $\boldsymbol{\alpha}$.

Dependent Random Inputs

In the general case, random inputs in the physical space are correlated in practice (e.g., the spatial and temporal correlation may exist between wind speeds [18, 148]). Several methods have been proposed to solve the correlation between random inputs. The most common way used in literature is to convert the correlated random inputs into independent ones using some transformation techniques, such that the salient property described in (2.14) is retained. For linearly correlated inputs, methods like principal component analysis (PCA) [149] or Karhunen-Loève expansions (KLE) [37] can be used. Nataf transformation (NT) is effective for decorrelating inputs with correlations modeled by Gaussian Copula [89]. However, NT may not handle uncertainties with highly nonlinear or thick tail dependence. In such cases, the Rosenblatt transform (RT) offers an alternative approach to transforming dependent inputs into independent ones, often combined with vine copula to model the dependence structure of uncertainties [150]. The PCE model built based on data after decor-

Table 2.2 An example of the multivariate polynomial bases $\Psi_k(\boldsymbol{\zeta})$ construction

$\Psi_k(\boldsymbol{\zeta})$	$\prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(\zeta_j)$	$\boldsymbol{\alpha}_k = (\alpha_k^1, \alpha_k^2, \alpha_k^3)$	H
$\Psi_0(\boldsymbol{\zeta})$	$\phi_1{}^{(0_0^1)}(\zeta_1) \times \phi_2{}^{(0_0^2)}(\zeta_2) \times \phi_3{}^{(0_0^3)}(\zeta_3)$	$\boldsymbol{\alpha}_0 = (0,0,0)$	0
$\Psi_1(\boldsymbol{\zeta})$	$\phi_1{}^{(1_1^1)}(\zeta_1) \times \phi_2{}^{(0_1^2)}(\zeta_2) \times \phi_3{}^{(0_1^3)}(\zeta_3)$	$\boldsymbol{\alpha}_1 = (1,0,0)$	1
$\Psi_2(\boldsymbol{\zeta})$	$\phi_1{}^{(0_2^1)}(\zeta_1) \times \phi_2{}^{(1_2^2)}(\zeta_2) \times \phi_3{}^{(0_2^3)}(\zeta_3)$	$\boldsymbol{\alpha}_2 = (0,1,0)$	1
$\Psi_3(\boldsymbol{\zeta})$	$\phi_1{}^{(0_3^1)}(\zeta_1) \times \phi_2{}^{(0_3^2)}(\zeta_2) \times \phi_3{}^{(1_3^3)}(\zeta_3)$	$\boldsymbol{\alpha}_3 = (0,0,1)$	1
$\Psi_4(\boldsymbol{\zeta})$	$\phi_1{}^{(1_4^1)}(\zeta_1) \times \phi_2{}^{(1_4^2)}(\zeta_2) \times \phi_3{}^{(0_4^3)}(\zeta_3)$	$\boldsymbol{\alpha}_4 = (1,1,0)$	2
$\Psi_5(\boldsymbol{\zeta})$	$\phi_1{}^{(1_5^1)}(\zeta_1) \times \phi_2{}^{(0_5^2)}(\zeta_2) \times \phi_3{}^{(1_5^3)}(\zeta_3)$	$\boldsymbol{\alpha}_5 = (1,0,1)$	2
$\Psi_6(\boldsymbol{\zeta})$	$\phi_1{}^{(0_6^1)}(\zeta_1) \times \phi_2{}^{(1_6^2)}(\zeta_2) \times \phi_3{}^{(1_6^3)}(\zeta_3)$	$\boldsymbol{\alpha}_6 = (0,1,1)$	2
$\Psi_7(\boldsymbol{\zeta})$	$\phi_1{}^{(2_7^1)}(\zeta_1) \times \phi_2{}^{(0_7^2)}(\zeta_2) \times \phi_3{}^{(0_7^3)}(\zeta_3)$	$\boldsymbol{\alpha}_7 = (2,0,0)$	2
$\Psi_8(\boldsymbol{\zeta})$	$\phi_1{}^{(0_8^1)}(\zeta_1) \times \phi_2{}^{(2_8^2)}(\zeta_2) \times \phi_3{}^{(0_8^3)}(\zeta_3)$	$\boldsymbol{\alpha}_8 = (0,2,0)$	2
$\Psi_9(\boldsymbol{\zeta})$	$\phi_1{}^{(0_9^1)}(\zeta_1) \times \phi_2{}^{(0_9^2)}(\zeta_2) \times \phi_3{}^{(2_9^3)}(\zeta_3)$	$\boldsymbol{\alpha}_9 = (0,0,2)$	2
$\Psi_{10}(\boldsymbol{\zeta})$	$\phi_1{}^{(1_{10}^1)}(\zeta_1) \times \phi_2{}^{(1_{10}^2)}(\zeta_2) \times \phi_3{}^{(1_{10}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{10} = (1,1,1)$	3
$\Psi_{11}(\boldsymbol{\zeta})$	$\phi_1{}^{(2_{11}^1)}(\zeta_1) \times \phi_2{}^{(1_{11}^2)}(\zeta_2) \times \phi_3{}^{(0_{11}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{11} = (2,1,0)$	3
$\Psi_{12}(\boldsymbol{\zeta})$	$\phi_1{}^{(2_{12}^1)}(\zeta_1) \times \phi_2{}^{(0_{12}^2)}(\zeta_2) \times \phi_3{}^{(1_{12}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{12} = (2,0,1)$	3
$\Psi_{13}(\boldsymbol{\zeta})$	$\phi_1{}^{(1_{13}^1)}(\zeta_1) \times \phi_2{}^{(2_{13}^2)}(\zeta_2) \times \phi_3{}^{(0_{13}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{13} = (1,2,0)$	3
$\Psi_{14}(\boldsymbol{\zeta})$	$\phi_1{}^{(0_{14}^1)}(\zeta_1) \times \phi_2{}^{(2_{14}^2)}(\zeta_2) \times \phi_3{}^{(1_{14}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{14} = (0,2,1)$	3
$\Psi_{15}(\boldsymbol{\zeta})$	$\phi_1{}^{(1_{15}^1)}(\zeta_1) \times \phi_2{}^{(0_{15}^2)}(\zeta_2) \times \phi_3{}^{(2_{15}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{15} = (1,0,2)$	3
$\Psi_{16}(\boldsymbol{\zeta})$	$\phi_1{}^{(0_{16}^1)}(\zeta_1) \times \phi_2{}^{(1_{16}^2)}(\zeta_2) \times \phi_3{}^{(2_{16}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{16} = (0,1,2)$	3
$\Psi_{17}(\boldsymbol{\zeta})$	$\phi_1{}^{(3_{17}^1)}(\zeta_1) \times \phi_2{}^{(0_{17}^2)}(\zeta_2) \times \phi_3{}^{(0_{17}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{17} = (3,0,0)$	3
$\Psi_{18}(\boldsymbol{\zeta})$	$\phi_1{}^{(0_{18}^1)}(\zeta_1) \times \phi_2{}^{(3_{18}^2)}(\zeta_2) \times \phi_3{}^{(0_{18}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{18} = (0,3,0)$	3
$\Psi_{19}(\boldsymbol{\zeta})$	$\phi_1{}^{(0_{19}^1)}(\zeta_1) \times \phi_2{}^{(0_{19}^2)}(\zeta_2) \times \phi_3{}^{(3_{19}^3)}(\zeta_3)$	$\boldsymbol{\alpha}_{19} = (0,0,3)$	3

relation is described in ***Remark* 2.1.3**. The literature presents alternative approaches that bypass the dependence on random inputs. [134, 151]. Wang et al. [134] developed multivariate polynomials $\Psi_k(\boldsymbol{\zeta})$ by employing Gram-Schmidt-based methods with correlated inputs. However, their effectiveness in high-dimensional contexts remains unproven. Torre et al. [151] suggested ignoring the dependence structure and building the multivariate polynomial bases $\Psi_k(\boldsymbol{\zeta})$ using (2.7) for the response estimation, reasonable good response estimations can be obtained, yet it failed to provide the closed-form representations of estimated statistics.

Remark 2.1.3. Assume that correlated random inputs $\boldsymbol{\zeta}$ after decorrelation using the transforms (e.g., NT, RT or PCA technique), denoted by $\mathcal{T}^{-1}(\boldsymbol{Z})$ [37, 89, 149, 150] are independent. As such, the response Y (e.g., PTTC, objective values of ED) is approximated by the PCE model built with random inputs after decorrelation:

$$Y = G(\boldsymbol{\zeta}) = G(\mathcal{T}^{-1}(\boldsymbol{Z})) \approx G^{\mathrm{pc}}(\boldsymbol{Z}) = \sum_{k=0}^{L-1} c_k \Psi_k(\boldsymbol{Z}) \tag{2.9}$$

where $\Psi_k(\boldsymbol{Z})$ is calculated using (2.7), that is:

$$\Psi_k(Z_1, \cdots, Z_{\mathcal{M}}) \;=\; \prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(Z_j) \tag{2.10}$$

Specially, this method was applied in Chapter 3 and Chapter 4 using PCA to decorrelate for response estimation. The NT and RT are used in Chapter 5. It's crucial to note that the choice of a decorrelation method is contingent upon the types of random input distributions and their underlying dependence structures.

Remark 2.1.4. The detailed introduction of PCA, NT, and RT can be found in Appendix A and Appendix B.

2.1.3 The Calculation of PCE Coefficients

In this thesis, the least-square regression method is applied to calculate the expansion coefficients c_k. For a given set of sample pairs $[\boldsymbol{\zeta}_p, \boldsymbol{Y}_p]$ with random input samples $\boldsymbol{\zeta}_p = \left\{\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)} \cdots, \boldsymbol{\zeta}^{(M_p)}\right\}$ and the corresponding response $\boldsymbol{Y} = \left\{Y^{(1)}, Y^{(2)}, \cdots, Y^{(M_p)}\right\}$, the expansion coefficients c_k can be computed by minimizing the following cost function [78]:

$$\begin{aligned} J(\boldsymbol{C}) &= \mathbb{E}\left[\sum_{s=1}^{M_p}\left[Y^{(s)} - \sum_{k=0}^{L-1} c_k \Psi_k(\boldsymbol{\zeta}^{(s)})\right]^2\right] \\ &= \mathbb{E}\Big[(\boldsymbol{Y} - \boldsymbol{\Psi C})^{\mathrm{T}} (\boldsymbol{Y} - \boldsymbol{\Psi C})\Big] \end{aligned} \tag{2.11}$$

where $\boldsymbol{C} = [c_0, c_2, \ldots, c_{L-1}]^{\mathrm{T}}$ with L denotes PCE coefficients in vector form. The matrix $\boldsymbol{\Psi}$ is formulated by the multivariate polynomials:

$$\begin{bmatrix} \Psi_0(\boldsymbol{\zeta}^{(1)}) & \Psi_1(\boldsymbol{\zeta}^{(1)}) & \cdots & \Psi_{L-1}(\boldsymbol{\zeta}^{(1)}) \\ \Psi_0(\boldsymbol{\zeta}^{(2)}) & \Psi_1(\boldsymbol{\zeta}^{(2)}) & \cdots & \Psi_{L-1}(\boldsymbol{\zeta}^{(2)}) \\ \cdots & \cdots & \cdots & \cdots \\ \Psi_0(\boldsymbol{\zeta}^{(M_p)}) & \Psi_1(\boldsymbol{\zeta}^{(M_p)}) & \cdots & \Psi_{L-1}(\boldsymbol{\zeta}^{(M_p)}) \end{bmatrix} \tag{2.12}$$

Then, the coefficients can be obtained by solving the least-square problem in (2.11) using ordinary least-square (OLS):

$$\hat{\boldsymbol{C}} = \left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}\right)^{-1} \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{Y} \tag{2.13}$$

2.1.4 Moments of the Model Response

If random inputs $\boldsymbol{\zeta}$ are independent, closed-forms of the mean and variance of response (i.e., $\mathbb{E}[Y]$ and $\mathrm{Var}[Y]$) can be obtained from the PCE coefficients due to the orthonormality of polynomial basis $\Psi_k(\boldsymbol{\zeta})$.

$$\mathbb{E}[Y] \approx \mathbb{E}\left[\sum_{k=0}^{L-1} c_k \Psi_k(\boldsymbol{\zeta})\right] = c_0, \quad \mathrm{Var}[Y] \approx \mathrm{Var}\left[\sum_{k=0}^{L-1} c_k \Psi_k(\boldsymbol{\zeta})\right] = \sum_{k=1}^{L-1} \gamma_k c_k^2 \tag{2.14}$$

where c_0 is the coefficient of the constant term in (2.9).

2.2 Conclusions

In this section, the general form of the PCE method is introduced. Specifically, a generalized PCE-based model constructed using known input distributions is presented. In subsequent chapters, the PCE-based methods will be used to assess ATC and PTTC, to solve the stochastic-optimization

ED problem, and for GSA-based uncertainty control. Moreover, the moment-based method will be introduced in Chapter 3 to construct the polynomial bases. These bases will also be utilized in Chapter 4 and Chapter 5. PCE models, constructed using the bases introduced in Table 2.1, will be used for comparison. Furthermore, different coefficient calculation methods will be introduced in the following chapters.

Chapter 3

A Data-Driven Sparse PCE Method for UQ in ATC Assessment

As introduced in Chapter 1, high penetration of RESs, new loads (e.g., EVs), and unexpected equipment outages bring a high level of uncertainty in power systems. The impacts of these uncertainties on power system static security assessment need to be studied. Specially, this chapter investigates the impacts of these uncertainties on available transfer capability (ATC). This chapter will first introduce some power system transfer capability concepts. Besides, this chapter introduces a data-driven sparse PCE (DDSPCE) method for assessing the probabilistic total transfer capability (PTTC), based on which, ATC with a certain confidence interval is evaluated. The ideas and findings presented in this chapter are predominantly based on previously published research [J1].

3.1 Introduction

3.1.1 Transfer Capability Definitions

Power transfer capability quantifies the maximum power that can be transmitted across a network under specific assumptions and conditions. It is significantly important in safeguarding security and optimizing the efficiency of power systems, particularly in deregulated environments. According to the definitions of the North American Electric Reliability Council (NERC) [4], *available transfer capability* (ATC), quantifies the remaining transfer capacity within the physical transmission network that can be utilized for future commercial activity over and above already committed uses [4]. It represents the capability available for additional power transfers in the market. The determination of ATC is closely related to *total transfer capability* (TTC). TTC refers to the maximum amount of electric power that can be reliably transmitted along a specific path within the transmission network while satisfying predetermined pre- and post-contingency system conditions, i.e.,TTC is the transfer capability of the transmission network that may be limited by physical and electrical constraints, including thermal, voltage, and stability limits. Therefore, TTC can be determined by:

$$\text{TTC} = \text{Minimum of } \{\text{thermal limit, voltage limit, stability limit}\}$$

The most restrictive limit on TTC may vary with system operating conditions changing (see Fig.3.1).

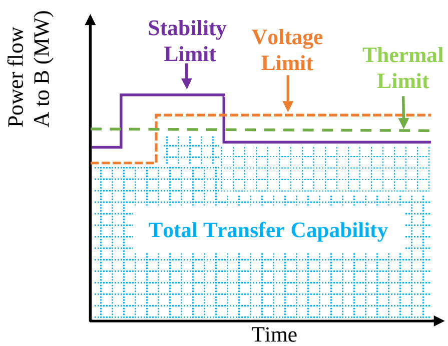

Fig. 3.1 Limits to Total Transfer Capability [4]

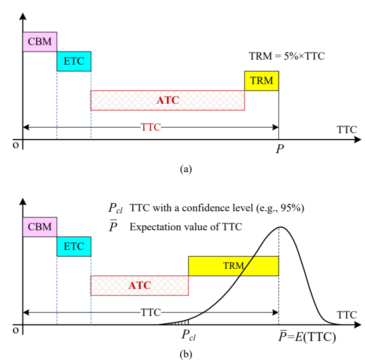

Fig. 3.2 ATC and its related concepts

The determination of TTC and ATC is vital for ensuring secure and reliable power transmission. The deregulation of the electricity industry aims to foster a sustainable and competitive market for electricity trading, ensuring fairness and transparency in power system operations. To achieve this goal, the Federal Energy Regulatory Commission (FERC) mandates the posting of ATC on the Open Access Same-time Information System (OASIS) [152], which requires accurate and efficient ATC and TTC assessment.

Mathematically, ATC can be calculated by:

$$\text{ATC} = \text{TTC} - \text{TRM} - \text{ETC} - \text{CBM} \tag{3.1}$$

where TRM is the *transmission reliability margin* which is the reserve margin for dealing with uncertainties in system conditions to guarantee the secure operation of the interconnected transmission network. ETC is the *existing transmission commitments* (base case power flow), including retail customer service. CBM is the *capacity benefit margin*, which guarantees access to generation from interconnected systems that can meet generation reliability requirements. Generally, the relation between these concepts is described by a traditional deterministic framework (see Fig.3.2(a)), where TRM is typically a fixed value or at a certain percentage of TTC (e.g., TTC$\times 5\%$), ETC can be obtained from base power flow, and CBM can be typically specified based on utility's market model [153].

3.1.2 The Probabilistic TTC Assessment Problem

After reviewing some basic definitions of transfer capability-related concepts, we now focus on the challenges required to be addressed in the TTC and ATC evaluation problem. Traditionally, the assessment of TTC and ATC has been conducted within a deterministic framework. Nevertheless, the social awareness of a more eco-friendly society and growing load demand have resulted in high penetration of RESs (e.g., wind and solar farms), which push the system to operate closer to its limits. The increasing integration of intermittent RESs introduces uncertainties on the generation side due to the varied power output from renewable generators. Besides, the predictions of loads are imperfect even with improvements in recent forecasting technology, which further brings uncertainties to the load side. Furthermore, the rising probability of line outages resulting from aging transmission networks has also introduced substantial uncertainty into power grids [154].
These uncertainties greatly affect the system's stability and security, which leads the TTC to be an uncertain quantity [155]. Consequently, it is more reasonable to consider TTC as a random variable, referred to as probabilistic TTC (PTTC), rather than a deterministic quantity, as depicted in Fig. 3.2 (b). Moreover, the determination of TRM becomes crucial to incorporate the increased randomness arising from intermittent RESs and equipment failures. In other words, it is necessary to adopt a probabilistic framework to address these uncertainties. As such, the calculation of ATC relies on evaluating the statistical properties of PTTC, which, in turn, allows for estimating the TRM (e.g., the difference between the mean value and the 95% confidence interval of the TTC probability distribution). Based on these estimations, the evaluation of ATC under a probabilistic framework can be summarized as follows:

$$\mathrm{ATC} = \mathbb{E}[\mathrm{PTTC}] - \mathrm{TRM} - \mathrm{ETC} - \mathrm{CBM} \tag{3.2}$$

Besides, in PTTC calculation, the maximum transfer capability considering many scenarios based on $N-1$ contingencies will be evaluated. However, due to the unexpected equipment outages, only considering $N-1$ contingencies may not be sufficient to cover all possible scenarios that will occur in the future [156], particularly for a large system with a huge amount of equipment. Therefore, the impact of unexpected outages in PTTC assessment requires more work. Moreover, typical voltage control devices such as adjustable transformers (e.g., Under-Load Tap Changer (ULTC) transformer) may significantly change the value of ATC, hence they must be considered in TTC and ATC computation.

A clear comparison between Fig. 3.2 (a) and Fig. 3.2 (b) reveals that the introduction of increased randomness from RESs and unforeseen outages can lead to larger TRM values and, consequently, smaller ATC values. Achieving accurate estimations of PTTC, TRM, and ATC is crucial for maintaining the secure operation of the grid while maximizing the utilization of transmission assets and reaping economic benefits.

3.1.3 Existing Methods

In this section, a literature review of existing methods, including deterministic and probabilistic approaches for TTC and ATC assessment, will be given.

1) Deterministic Approaches

In previous works, a number of methods have been proposed in the literature for deterministic TTC and ATC assessment. These methods can be divided into two main types, one is based on the direct current (DC) flow, and the other is based on alternating current (AC) power flow. Hamoud et al. proposed a DC load flow-based method to assess ATC with high computation efficiency, while only the real power flows through lines are modeled [157]. Christie and Ejebe et al. introduced the linear DC-based distribution factors to calculate ATC quickly, yet only applicable to systems operating close to the base case [158], [159]. Continuation power flow (CPF) [160–163], repeated power flow (RPF) [164], optimal power flow (OPF) [165] are categorized in the AC power flow group. Among these methods, CPF is the most widely used due to its salient property of not encountering the numerical difficulty of ill-conditioning at or around the critical point, despite RPF being much easier to implement and having a fast convergence rate. CPF method introduces a load parameter and determines the solution from the augmented Jacobian matrix, which contributes to its well-conditioning feature. However, as discussed in Section 3.1.2, uncertainties brought by RES, load variations, and branch outages turn TTC into a random variable, which requires more work to deal with the uncertainties. Thus, researchers have focused on the probabilistic approaches, and we will briefly introduce these approaches in the following section.

2) Probabilistic Approaches

Besides the deterministic approaches discussed in section 3.1.3, many probabilistic approaches have been developed to account for the uncertainties discussed in Section 3.2.2. Currently, a traditional simulation-based method has been the most widely used to deal with the uncertainties in PTTC evaluation, namely, MC simulations (MCS) [166], [167]. MCS can be implemented through a straightforward procedure that have been utilized with the deterministic AC power flow approaches mentioned in Section 3.1.3. However, this method is not practically attractive due to tens of thousands of simulations required to achieve sufficient accuracy for the results, despite some advanced sampling techniques being integrated such as LHS [52] and LSS [168]. To lower the computation time, Ramezani et al. [169] combined the clustering method with MCS while the accuracy deteriorated. Later on, Chang et al. [170] developed a Bootstrap method that is professional in estimating the confidence interval of ATC, while the re-sampling procedure results in high computation time. Therefore, an accurate and efficient PTTC evaluation method is still a prerequisite.

To cut down the computing time while ensuring accuracy, PCE [6], a popular uncertainty quantification method, is adopted in [171], [172]. The PCE model can construct a statistically-equivalent surrogate model for PTTC evaluation through a small number of model response pairs, from which the sample mean, sample variance, PDF, and CDF can be obtained from the PCE weights. Particularly, Fei et al. [89] and Hao et al. [93] have further enhanced the PCE-based method by integrating an adaptive sparse scheme to tackle the "curse of dimensionality" and using the Nataf transformation to deal with the highly correlated random inputs with various types of marginal distributions. Likewise, a low-rank approximation method is proposed to assess the available transfer capability in [173]. More recently, Liu et al. [174] proposed a nonparametric analytic method to evaluate the dynamic TTC, where the model is built by a group Lasso regression-based training scheme with the probabilistic distributions of state variables assumed to be Gaussian. However, in all the methods discussed above, the marginal distributions of random variables are inferred though some parametric distribution functions (e.g., wind speed follows Weibull distribution [8], solar irradiance follows Beta distribution [175]), while the real world data may not follow the parametric distributions or with unknown distributions. Hopefully, the raw data of random inputs such as wind speed and solar irradiance are more likely to be obtained [176].

To tackle these challenges, researchers are currently focusing on how to exploit data directly to assess PTTC. An interval optimization-based model was proposed in [177] to assess ATC, which only requires the boundaries of random variables. A point estimate method was adopted in [153] [66] to estimate the standard deviation of PTTC using only data information. However, the probabilistic distribution of PTTC cannot be estimated in [153] [177] [66], which may result in an inaccurate estimation of ATC. Zhang et al. [178] and Xu et al. [38] developed PTTC assessment methods with the capability to estimate its probability distribution. However, the probabilistic distribution approximation procedures in their methods may introduce additional errors in PTTC estimation. A TTC calculation model based on a deep learning method (stacked denoising autoencoder) was proposed in [179], which requires a large number of training data and intractable future selection. To overcome these difficulties, recently, Wang et al. [180] developed a data-driven based PCE method to solve the probabilistic power flow, while probability distribution assumption and approximation are not required, yet this method only considers the continuous random variables.

Furthermore, discrete random variables (e.g., line outages) are required to be carefully handled in PTTC evaluation. Despite $N-1$ contingencies are generally considered in the PTTC assessment according to past blackout reports, it is still not enough to guarantee the systems' secure operation [181] since the $N-1$ contingencies may not be able to cover all the possible outage scenarios, resulting in an overestimated ATC. Besides, the control devices are also not considered in the aforementioned methods, which may affect the ATC values dramatically, such as the integration of ULTC [182]. Therefore, situations with more than one line outage (e.g., $N-2$ or more) and control devices are also significant and require to be addressed in the TTC and ATC assessment.

To address the challenges associated with volatile renewables, load variations, and unexpected branch outages in power systems, this chapter aims to propose a data-driven sparse Polynomial Chaos Expansion (DDSPCE) method for assessing PTTC and ATC. This method builds a surrogate model for the CPF-based model, enabling efficient PTTC assessment, based on which, ATC with certain confidence intervals are evaluated. The main contributions of this chapter are as follows.

- The introduced DDSPCE method directly uses available data to estimate the probabilistic characteristics of PTTC, such as mean, variance, and probability distribution, without the need for pre-assumed probability distributions of random inputs.

- DDSPCE is capable of managing a vast array of mixed random inputs, both continuous and discrete (e.g., wind speed, solar irradiance, imminent line outages) when evaluating PTTC.

- By integrating a sparse PCE scheme, specifically the Least Angle Regression (LAR), DDSPCE offers an accurate estimation of PTTC in a fraction of the computational time required by MCS.

- A numerical study highlights the critical importance of incorporating discrete variables in ATC assessment, an aspect that has previously been overlooked.

The remainder of this chapter is as follows. Section 3.2 presents the mathematical formulation of the PTTC problem based on the CPF-based method and uncertainty modeling. Subsequently, this chapter explores the application of polynomial chaos theory, introduced in Chapter 2, to construct a DDSPCE-based model. The theoretical formulation of the proposed DDSPCE method is established in Section 3.3, laying the foundation for its practical implementation. Section 3.4 provides the overall ATC computation procedures. Section 3.5 shows the simulation studies on the modified IEEE 118-bus system including different scenarios involving continuous random inputs, mixed random inputs, $N-K$ contingencies, and adjustable transformers. Furthermore, Section 3.6 demonstrates the scalability of the proposed method by extending it to large-scale power systems, specifically, the PEGASE 1354-bus system with a larger number of random variables. The simulation results validate the accuracy and efficiency of the proposed method in PTTC and ATC assessment, even when dealing with numerous renewable sources, stochastic loads, and topological changes. The results are compared with the benchmark MCS results and those obtained using the SPCE method presented in [93].

3.2 CPF-Based Mathematical Formulation of the PTTC Assessment

This thesis adopts the CPF to model the PTTC assessment problem due to its salient feature of well-conditioning at and around the critical point.

3.2.1 The Deterministic CPF Based TTC Formulation

Continuation methods, also known as branch tracing, as discussed in [162,183], can be employed to determine the steady-state limit of a power system by tracing a solution path of a set of power flow equations, which is called CPF. This path evolves from a given base point, along with a specified load-generation pattern. Consider a N-bus transmission system, with deterministic power flow equations described as follows (see Chapter 3 [32]).

$$\boldsymbol{f}(\boldsymbol{x})=\begin{bmatrix}P_{Gi}-P_{Li}-P_{Ti}(\boldsymbol{x})\\ Q_{Gi}-Q_{Li}-Q_{Ti}(\boldsymbol{x})\end{bmatrix}=0,\quad i=\{1,2,\cdots,N\} \tag{3.3}$$

where $\boldsymbol{x}=[\boldsymbol{\theta},\boldsymbol{V}]^T$ is the state vector, $\boldsymbol{\theta}$ and $\boldsymbol{V}$ are voltage angles and magnitudes for all buses, respectively; P_{Gi} and Q_{Gi} denote the active and reactive power injections from the traditional generator; P_{Li} and Q_{Li} denote the active and reactive load power at bus i, respectively; $P_{Ti}(\boldsymbol{x})$ and $Q_{Ti}(\boldsymbol{x})$ are the total real and reactive power injections at bus i with the following forms.

$$\begin{aligned}P_{Ti}(x)&=V_i\textstyle\sum_{j=1}^{N}V_j\left(G_{ij}\cos\theta_{ij}+B_{ij}\sin\theta_{ij}\right),\quad i=\{1,\cdots,N\}\\ Q_{Ti}(x)&=V_i\textstyle\sum_{j=1}^{N}V_j\left(G_{ij}\sin\theta_{ij}-B_{ij}\cos\theta_{ij}\right),\quad i=\{1,\cdots,N\}\end{aligned} \tag{3.4}$$

where (3.4) represents the AC load flow equations, V_i denotes the voltage magnitude at bus i; θ_{ij} denotes the angle difference between bus i and bus j; G_{ij} (i.e., the conductance) and B_{ij} (i.e., the susceptance) are the real and the imaginary part of the element Y_{ij} in the bus admittance matrix $\boldsymbol{Y}_{\text{bus}}$.

To determine the transfer capability, we define the load-generation variation vector to represent the direction of power transfer:

$$\boldsymbol{b}=\begin{bmatrix}\Delta P_{G,i}-\Delta P_{L,i}\\ -\Delta Q_{L,i}\end{bmatrix},\quad i=\{1,2,\cdots,N\} \tag{3.5}$$

where $\boldsymbol{b}$ specifies the power transaction under considered; $\Delta P_{G,i}$, $\Delta P_{L,i}$, and $\Delta Q_{L,i}$ denote the increase of active generation power, active load power, and reactive load power at bus i, respectively. By introducing a continuation parameter λ into (3.3), we can use the structured CPF to determine the total power that can be transferred under a specific direction $\boldsymbol{b}$:

$$\boldsymbol{f}(\boldsymbol{x},\lambda)=\boldsymbol{f}(\boldsymbol{x})-\lambda\boldsymbol{b}=0 \tag{3.6}$$

where $\lambda\in\mathbb{R}$ indicates the transfer capability under the specific direction $\boldsymbol{b}$. In the deterministic CPF (3.6), TTC can be calculated by starting with an initial value and gradually increasing the continuation parameter λ until the limit (e.g., thermal, voltage, and stability limits) is reached. At each step, the algorithm predicts the solution based on the previous solution and then uses the Newton-Raphson technique to make corrections. This process iterates until the limit is reached, at which point the algorithm terminates [162]. As discussed in [184], the maximum value of λ without violating the physical and electrical characteristics of the systems (e.g., thermal, voltage, and stability limits) gives TTC.

Conventionally, TTC is calculated in the deterministic CPF method (i.e., a deterministic way), with a fixed percentage (e.g., 5% by NERC [185]) of TTC reserved as TRM to account for uncertainties in the system. The growing level of uncertainty associated with the expansion of RESs, load variations, and aging transmission networks which typically result in higher probabilities of failure (see Chapter 6 [22]), may lead to a larger TRM. A deterministic framework with fixed TRM may not be sufficient for accounting for the uncertainties. Therefore, it is important to carefully model and consider these uncertainties in the calculation of PTTC to obtain a more accurate TRM, and hence a more accurate ATC value. In the following subsection, stochastic modeling of the uncertainties brought by RESs, load variations, and unexpected outages will be presented.

3.2.2 Modelling Uncertainties in PTTC Problem

In this thesis, the volatile RESs (e.g., wind and solar), load variations, and unexpected branch outages are regarded as sources of uncertainties. The uncertainty sources are regarded as random variables, including continuous random variables (wind speed $\boldsymbol{v}(\mathrm{m/s})$, solar irradiance $\boldsymbol{r}(\mathrm{W/m^2})$, and load variations $\boldsymbol{P_L}$), and discrete random variables (branch states $\boldsymbol{\rho}$). In particular, the random input variables in this thesis are not limited to any specific probability distributions. They can be obtained either from measured data or generated using existing probability models.

1) Wind Generation

This thesis assumed that specific buses are connected to wind farms where the wind generator power output $P_w(v)$ is uncertain and depends on the wind speed v. The wind resource exhibits significant variability, which fluctuates randomly with time and geographic locations. The wind generator's real power output $P_w(v)$ uncertainty can be modeled as follows.

Step 1. Obtain the wind speed data. The wind speed data can either be assessed from historical data [186] or generated by a known probability distribution model. For example, the wind speed v in many locations around the world can be modeled by Weibull distribution [8, 18, 133].

Step 2. Determine wind generator's real output power. Once wind speed data is obtained, the wind turbine generator's real output power $P_w(v)$ can be determined using the wind speed-power curve [187]:

$$P_w(v) = \begin{cases} 0 & v \leq v_{in} \text{ or } v > v_{out} \\ \frac{v - v_{in}}{v - v_{rated}} P_{rw} & v_{in} < v \leq v_{rated} \\ P_{rw} & v_{rated} < v \leq v_{out} \end{cases} \tag{3.7}$$

where v_{in}, v_{out} and v_{rated} are the cut-in, cut-out, and rated wind speed (m/s); P_{rw} is the rated wind power (MW).

Step 3. Calculate the wind generator's reactive output power. The reactive power output of the wind generator can be calculated by modeling the wind turbine as a constant P-Q bus (e.g., with a power factor 0.85 lagging in Section 3.5 Case II) or a constant P-V bus with a specified reactive power limit [188].

2) Solar Generation

In recent years, Solar PV plants have been increasingly installed worldwide due to their ability to produce clean energy and low cost. The power output of the solar generator is highly uncertain due to the variability of solar irradiance, which depends on various factors, including weather, environmental, and different time conditions. The solar generator's real power output $P_{pv}(r)$ uncertainty modeling is presented below.

Step 1. Obtain the data of solar irradiance. Similarly, the data of solar irradiance can be either obtained from measurements [189] or generated from an existing probability model (e.g., Beta distribution [190]).

Step 2. Determine the solar cell generator's real power output P_{pv}. Once the solar irradiance data is obtained, P_{pv} can be determined according to the radiation–power curve [187, 190]

$$P_{pv}(r) = \begin{cases} \frac{r^2}{r_c r_{std}} P_{rs} & 0 \leq r < r_c \\ \frac{r}{r_{std}} P_{rs} & r_c < r \leq r_{std} \\ P_{rs} & r > r_{std} \end{cases} \tag{3.8}$$

where r_c denotes the certain radiation at $150\ \mathrm{W/m^2}$; r_{std} denotes the standard solar irradiance at $1000\ \mathrm{W/m^2}$; P_{rs} denotes the rated power of PV panel. In accordance with [191], solar generation is typically injected into the power grid at a unity power factor. Therefore, this thesis assumes that Q_{pv} is equal to zero.

3) Load Variation

Indeed, the load is the most noticeable uncertain variable in the power system, fluctuating with time, weather conditions, and electricity prices, among other factors. Similarly, the load data can be obtained from historical data [192] or generated from a known probability distribution (e.g., Gaussian distribution [193, 194]). Generally, only the active power is predicted by the load forecaster, whereas the reactive power is determined under the assumption of constant power factor [153].

4) Unexpected Branch Outages

An outage, which refers to the disconnection of a transmission line from the grid, is one of the most common faults in power systems. In this thesis, the nth branch outage with probability q_n is modeled by a discrete random variable, which follows a Bernoulli distribution [195]:

$$\mathbb{P}\{\rho_n = 1\} = q_n = 1 - \mathbb{P}\{\rho_n = 0\} \tag{3.9}$$

where ρ_n represents the state of the nth branch. Additionally, credible historical data can also be used to model the uncertainty of line outages. It is worth noting that equipment failures, such as generation outages, can also be modeled using the same approach.

3.2.3 The Probabilistic CPF-Based PTTC Formulation

As mentioned above, increasing penetration of RESs, stochastic loads, and unexpected branch outages will affect the transfer capability. To investigate the impacts of these uncertainties, in this section, we integrate the uncertainties described in Section 3.2.2 into the deterministic power flow equations (3.3). Let $\boldsymbol{U} = [\boldsymbol{v}, \boldsymbol{r}, \boldsymbol{P_L}, \boldsymbol{\rho}]$, then (3.3) becomes a set of probabilistic power flow (PPF) equations: $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{U}) = 0$. Specially, for an N-bus transmission system, the probabilistic AC power flow equations $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{U})$ have the following form for P-Q type buses:

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{U}) = \begin{bmatrix} P_{Gi} + P_{wi}(v_i) + P_{pvi}(r_i) - P_{Li}(P_{Li}) - P_{Ti}(\boldsymbol{x}, \boldsymbol{\rho}) \\ Q_{Gi} + Q_{wi}(v_i) - Q_{Li}(P_{Li}) - Q_{Ti}(\boldsymbol{x}, \boldsymbol{\rho}) \end{bmatrix} = 0, i = \{1, \cdots, N\} \tag{3.10}$$

and for P-V type buses, the corresponding PPF equations are

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{U}) = \begin{bmatrix} P_{Gi} + P_{wi}(v_i) + P_{pvi}(r_i) - P_{Li}(P_{Li}) - P_{Ti}(\boldsymbol{x}, \boldsymbol{\rho}) = 0 \\ V_i = V_{i0} \\ Q_{Gi} = -Q_{wi}(v_i) + Q_{Li}(P_{Li}) + Q_{Ti}(\boldsymbol{x}, \boldsymbol{\rho}) \\ Q_{\min,i} \leq Q_{Gi} \leq Q_{\max,i} \end{bmatrix}, i = \{1, \cdots, N\} \tag{3.11}$$

where P_{Gi},$P_{wi}\left(v_i\right)$,$P_{pvi}\left(r_i\right)$ and P_{Li} are the real power injection from the conventional generator, wind farm, the solar PV power plant and the load at bus i; Q_{Gi}, $Q_{wi}\left(v_i\right)$ and Q_{Li} are the corresponding reactive power injections. The wind generators are modeled as P-Q type nodes (e.g. the power factor is 0.85 lagging [187]), while the solar generator is assumed to have a unity power factor [191]. Note that the transition from P-V to P-Q operation at the terminal bus will be activated if Q_{Gi} exceeds its limits at either $Q_{\min,i}$ or $Q_{\max,i}$, indicating that the corresponding generator is not able to perform voltage regulation.$P_{Ti}\left(\boldsymbol{x},\boldsymbol{\rho}\right)$ and $Q_{Ti}\left(\boldsymbol{x},\boldsymbol{\rho}\right)$ are given by (3.4). Similarly, by incorporating the random vector $\boldsymbol{U}$, the probabilistic CPF equations can be expressed compactly as follows:

$$\boldsymbol{f}\left(\boldsymbol{x},\boldsymbol{\eta},\lambda,\boldsymbol{U}\right)=\boldsymbol{f}\left(\boldsymbol{x},\boldsymbol{\eta},\boldsymbol{U}\right)-\lambda\boldsymbol{b}=0 \tag{3.12}$$

where $\boldsymbol{x}$ denotes the state vector; $\boldsymbol{\eta}$ denotes a vector of control parameters (e.g., tap ratios of adjustable transformers); λ denotes the transfer capability; $\boldsymbol{U}$ is the mixed random vector.

This thesis takes into account the thermal limits, voltage limits, generator active and reactive power limits [184], as well as a credible contingency list of size N_c when calculating TTC. Using the preceding notation, the mathematical formulation of the probabilistic TTC calculation based on the CPF method can be expressed as follows:

$$\begin{aligned}
\max \quad & \lambda^{(\kappa)} \\
\text{s.t.} \quad & \boldsymbol{f}^{(\kappa)}\left(\boldsymbol{x},\boldsymbol{\eta},\boldsymbol{U}\right)-\lambda^{(\kappa)}\boldsymbol{b}=0 && \text{(3.13a)}\\
& V_{\min,i}^{(\kappa)}\le V_i\left(\boldsymbol{x},\boldsymbol{\eta},\lambda^{(\kappa)},\boldsymbol{U}\right)\le V_{\max,i}^{(\kappa)}, && \text{(3.13b)}\\
& S_{ij}^{(\kappa)}\left(\boldsymbol{x},\boldsymbol{\eta},\lambda^{(\kappa)},\boldsymbol{U}\right)\le S_{ij,\max}^{(\kappa)}, && \text{(3.13c)}\\
& P_{\min,i}\le P_{Gi}\left(\boldsymbol{x},\boldsymbol{\eta},\lambda^{(\kappa)},\boldsymbol{U}\right)\le P_{\max,i}, && \text{(3.13d)}\\
& Q_{\min,i}\le Q_{Gi}\left(\boldsymbol{x},\boldsymbol{\eta},\lambda^{(\kappa)},\boldsymbol{U}\right)\le Q_{\max,i} && \text{(3.13e)}
\end{aligned}$$

where $\kappa=\{0,1,\cdots,N_c\}$ with $\kappa=0$ corresponding to the normal operation state (pre-contingency) and $\kappa=1,2,\cdots,N_c$ corresponding to the contingency cases; $\lambda^{(\kappa)}$ denotes the transfer capability in the κ-th case; Constraint (3.13a) denotes the probabilistic CPF and $\boldsymbol{f}^{(\kappa)}\left(\boldsymbol{x},\boldsymbol{\eta},\boldsymbol{U}\right)$ are the PPF equations with $\kappa=0$ and $\kappa=\{1,\cdots,N_c\}$ denote the pre-contingency and post-contingency network configuration, respectively; Constraint (3.13b) describes the voltage limits with $V_{\min,i}^{(\kappa)}$ and $V_{\max,i}^{(\kappa)}$ denoting the lower and the upper limits of voltage magnitude at bus i in normal operating case ($\kappa=0$) and in emergency cases ($\kappa\neq0$), respectively; Constraint (3.13c) denotes the thermal limits with $S_{ij,\max}^{(\kappa)}$ being the thermal limits of the line between bus i and bus j in the normal operating state ($\kappa = 0$) and in emergency cases ($\kappa \neq 0$), respectively; Constraints (3.13d) and (3.13e) are the generation capability constraints, with $P_{\min,i}$ and $P_{\max,i}$ denoting the lower and upper active generation power limits; $Q_{\min,i}$ and $Q_{\max,i}$ are the corresponding reactive power limits.

According to the definitions in [184], TTC is the maximum value of power that can be transferred without violating any limits (3.13b)-(3.13e) in both the normal operating state and in the emergency cases, which, in the above formulation, corresponds to the minimum λ among all the $\lambda^{(\kappa)}$ values:

$$\lambda^{\mathrm{TTC}} = \min\{\lambda^{(0)}, \lambda^{(1)}, \cdots, \lambda^{(N_c)}\} \tag{3.14}$$

Note that λ is regarded as a random variable when the uncertainties (i.e., $\boldsymbol{U}$) are integrated, which turns TTC into PTTC. As discussed in Section 3.1, once the mean and probability distributions of PTTC have been obtained, ATC can be calculated using (3.2). Conventionally, MC simulations are used to evaluate the PTTC by deterministic simulation tool (e.g., VSAT/DSATools in this thesis). However, to obtain accurate estimations of the probability distribution of PTTC, a large number of MC samples are required and fed into (3.13), resulting in enormous computational costs. Moreover, PTTC assessment faces significant challenges when random inputs follow arbitrary distributions (e.g., continuous, discrete, or mixed) and limited information (e.g., only raw data) is available. To address these challenges, a data-driven sparse PCE (DDSPCE) method is proposed to assess PTTC, considering the high penetration of RES, load variations, and unexpected branches. Specially, the DDSPCE method only requires a small number of sample evaluations to construct a surrogate model for the CPF-based PTTC formulation in (3.13).

3.3 The Proposed DDSPCE Method for PTTC Assessment

This section provides a comprehensive explanation of the data-driven sparse polynomial chaos expansion (DDSPCE) method used for assessing PTTC. Unlike the conventional generalized PCE method [6], which requires detailed probability distributions of all random inputs, the DDSPCE method constructs orthogonal polynomial bases for mixed random inputs (combining continuous and discrete) solely from data, specifically the moments. Moreover, to enhance computational efficiency while maintaining accurate estimation of the probabilistic characteristics of PTTC, a sparse PC scheme (such as LAR and a modified truncation scheme) is employed.

3.3.1 Data-Driven PCE for PTTC assessment

As presented in Chapter 2, a stochastic response $Y = G(\boldsymbol{\zeta})$ with finite second-order moments can be approximated by a PCE model. Let Y be the PTTC (i.e., λ^{TTC}) and the random vector $\boldsymbol{\zeta} = [\zeta_1, \cdots, \zeta_{\mathcal{M}}] \in \mathbb{R}^{\mathcal{M}}$ denote the random vector describing in 3.2.2, where $\zeta_j, j = \{1, \cdots, \mathcal{M}\}$ can be any random variable in $\boldsymbol{v}$, $\boldsymbol{r}$, $\boldsymbol{P_L}$, or $\boldsymbol{\rho}$ (i.e., wind speed, solar irradiance, load variation or unexpected branch outages). Besides, the following two assumptions are given:

Assumption 3.3.1. It is assumed that random vector $\boldsymbol{\zeta}$ has mutually independent counterparts with random variables $\zeta_j \in \mathbb{H}_{\boldsymbol{\zeta}}$ and response $Y \in \mathbb{H}_{\boldsymbol{\zeta}}$.

Therefore, the PTTC (i.e., λ^{TTC}), regarded as a random variable due to the stochastic characteristic of RESs, load variations, and unexpected branch outages, described by (3.13) can be approximated by a multi-dimensional PCE model (2.6) of order H [140]:

$$\lambda^{\mathrm{TTC}} = Y = G(\boldsymbol{\zeta}) \approx G^{\mathrm{pc}}(\boldsymbol{\zeta}) = \sum_{k=0}^{L-1} c_k \Psi_k(\zeta_1, \zeta_2, \cdots, \zeta_{\mathcal{M}}) \tag{3.15}$$

where c_k are the unknown PCE coefficients to be determined, and the multivariate polynomial bases $\Psi_k(\zeta_1, \zeta_2, \cdots, \zeta_{\mathcal{M}})$ are orthogonal with respect to the joint PDF $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ of $\boldsymbol{\zeta}$ as shown in (2.4). Conventionally, $\Psi_k(\zeta_1, \zeta_2, \cdots, \zeta_{\mathcal{M}})$ can be constructed through the tensor product of the univariate polynomial bases $\Phi_j(\zeta_j)$ using (2.7) and (2.8) directly if random inputs $\boldsymbol{\zeta}$ have mutually independent counterparts. However, in practical applications, it is quite common for random inputs to exhibit correlations. For instance, $\boldsymbol{U}$ in (3.12) is often dependent in physical space (e.g., wind speed $\boldsymbol{v}$ has temporal and spatial correlations [18, 133]). In this chapter, random inputs with linear correlation (spatial correlation) are considered. As introduced in ***Remark*** 2.1.3 Chapter 2, different decorrelation strategies are suggested for building polynomial bases when random inputs are correlated. This chapter employs the PCA technique [196] to eliminate the linear correlations between random inputs, i.e., the correlated input vector $\boldsymbol{U}$ is transformed to $\boldsymbol{\zeta}$ (decorrelated input vector) by: $\boldsymbol{U} = \mathcal{T}_{\mathrm{pca}}^{-1}(\boldsymbol{\zeta})$.

To build the PCE-based model, in the generalized PCE method, as introduced in Section 2.1.1 Chapter 2, the univariate orthogonal polynomial bases $\Phi_j(\zeta_j)$ selected according to Table 2.1 rely on the probability distribution of each individual continuous random variable ζ_j. However, as mentioned earlier, obtaining data sets instead of probabilistic distributions for random inputs (such as wind speed) is often more feasible in real-world power systems. Therefore, this thesis is to leverage the data-driven PCE method [140], which can construct orthonormal polynomial bases described in (2.7) solely from moments estimated from data. The detailed procedures for moment-based polynomials are provided in the following Section 3.3.2.

Remark 3.3.1. As a result, ζ_j is not limited to a predefined probability distribution and can be continuous, discrete, characterized by raw data sets, or even represented by a limited number of moments. Certainly, it can still be described by an arbitrary probability distribution if available.

3.3.2 Moment-Based Polynomials

Define the univariate orthogonal polynomial bases $\phi_j^{(\alpha_k^j)}(\zeta_j)$ in (2.7) for the j-th dimensional input ζ_j ($j = 1, \cdots, \mathcal{M}$) as follows:

$$\phi_j^{(l)}(\zeta_j) = \sum_{n=0}^{l} p_{n,j}^{(l)}(\zeta_j)^n \tag{3.16}$$

For simplicity, α_j^k is substituted by l for simplicity. where $l = \{1, \cdots, H\}$. $p_{n,j}$ is the unknown coefficient of the univariate polynomial basis $\phi_j^{(l)}(\zeta_j)$ in the n-th degree.

Given that the formulation of the polynomial basis $\phi_j^{(l)}(\zeta_j)$ remains consistent for each input variable ζ_j, the following polynomial basis formulation can be applied universally to any individual random input ζ_j. To construct the multivariate polynomial bases using (2.7), the univariate polynomial base $\phi_j^{(l)}(\zeta_j)$ constructed must satisfy the orthogonal condition:

$$\int_{\Omega_j} \phi_j^{(m)}(\zeta_j)\phi_j^{(l)}(\zeta_j)dF_{\zeta_j}(\zeta_j) = 0 \qquad \forall m \neq l \tag{3.17}$$

where $m, l = \{0, 1, \cdots, H\}$. To ensure the orthogonality of the polynomial bases, we begin by defining the coefficients of the leading terms for all polynomials to be 1 for simplicity, i.e.,

$$p_{l,j}^{(l)} = 1, \quad \forall l \tag{3.18}$$

When $l = 0$, by (3.18), we have $\phi_j^{(0)} = p_{0,j}^{(0)} = 1$. This procedure can be continued in a recursive way. Specifically, if we examine the orthogonality conditions of the polynomial $\phi_j^{(l)}(\zeta_j)$ with all lower-order polynomials, they can be described as follows:

$$
\begin{aligned}
&\int_{\zeta_j \in \Omega} p_{0,j}^{(0)} \left[\sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^n\right] dF_{\zeta_j}(\zeta_j) = 0 \\
&\int_{\zeta_j \in \Omega_j} \left[\sum_{n=0}^{1} p_{n,j}^{(1)} \zeta^n\right] \left[\sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^n\right] dF_{\zeta_j}(\zeta_j) = 0 \\
&\int_{\zeta_j \in \Omega_j} \left[\sum_{n=0}^{2} p_{n,j}^{(2)} \zeta_j^n\right] \left[\sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^n\right] dF_{\zeta_j}(\zeta_j) = 0 \\
&\vdots \\
&\int_{\zeta_j \in \Omega_j} \left[\sum_{n=0}^{l-1} p_{n,j}^{(l-1)} \zeta_j^n\right] \left[\sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^n\right] dF_{\zeta_j}(\zeta_j) = 0 \\
&p_l^{(l)} = 1
\end{aligned}
\tag{3.19}
$$

Then by substituting $p_{0,j}^{(0)} = 1$ into the first equation of (3.19), using the condition (3.18), we obtain $p_{1,j}^{(1)} = 1$. Continuing this process, substituting the first and second equations into the third equation, and so on, while maintaining the condition (3.18), we can represent the system of equations in (3.19) as follows:

$$
\begin{aligned}
&\int_{\zeta_j \in \Omega_j} \sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^n dF_{\zeta_j}(\zeta_j) = 0 \\
&\int_{\zeta_j \in \Omega} \sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^{n+1} dF_{\zeta_j}(\zeta_j) = 0 \\
&\int_{\zeta_j \in \Omega_j} \sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^{n+2} F_{\zeta_j}(\zeta_j) = 0 \\
&\vdots \\
&\int_{\zeta_j \in \Omega_j} \sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^{n+l-1} dF_{\zeta_j}(\zeta_j) = 0 \\
&p_l^{(l)} = 1
\end{aligned}
\tag{3.20}
$$

Remark 3.3.2. Recall that the n-th raw moment of ζ_j is defined as below:

$$
\mu_{n,j} = \int_{\zeta_j \in \Omega_j} \zeta_j^n dF_{\zeta_j}(\zeta_j) \tag{3.21}
$$

Integrate $\mu_{n,j}$, the n-th raw moment of ζ_j as defined in ***Remark*** (3.3.2), into (3.20), which can be reformulated as:

$$\begin{aligned}
&\sum_{n=0}^{l} p_{n,j}^{(l)} \mu_{n,j} = 0 \\
&\sum_{n=0}^{l} p_{n,j}^{(l)} \mu_{n+1,j} = 0 \\
&\sum_{n=0}^{l} p_{n,j}^{(l)} \mu_{n+2,j} 0 \\
&\quad \vdots \\
&\sum_{n=0}^{l} p_{n,j}^{(l)} \mu_{n+l-1,j} = 0 \\
&p_l^{(l)} = 1
\end{aligned} \tag{3.22}$$

As a result, the set of equations in (3.22) can be described by the moments of ζ_j and reconstructed in the matrix form:

$$\begin{bmatrix} \mu_{0,j} & \mu_{1,j} & \dots & \mu_{l,j} \\ \mu_{1,j} & \mu_{2,j} & \dots & \mu_{l+1,j} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{l-1,j} & \mu_{l,j} & \dots & \mu_{2l-1,j} \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} p_{0,j}^{(l)} \\ p_{1,j}^{(l)} \\ \vdots \\ p_{l-1,j}^{(l)} \\ p_{l,j}^{(l)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{3.23}$$

where the nth moment of the jth random input $\mu_{n,j}, n = \{0, 1, \cdots, 2l-1\}$ and $j = \{1, \cdots, \mathcal{M}\}$ can be approximated from the samples by

$$\mu_{n,j} = \frac{1}{M_p} \sum_{m=1}^{M_p} \zeta_{m,j}^n \tag{3.24}$$

where M_p represents the training sample size, and $\zeta_{m,j}$ denotes the sample points of an arbitrary input random variable ζ_j. When the raw data set or distributions of ζ_j are provided, the raw moment of ζ_j can be computed using (3.24) or(3.21).

The importance of moments $\mu_0, \mu_1, \dots, \mu_{2l-1}$ for the availability of any l-th order polynomial chaos expansion is evident from (3.23). These moments are necessary to establish a reliable framework for the l-th order polynomial chaos expansion. By solving (3.23) for the coefficients $p_{n,j}^{(l)}$, the orthogonal polynomial basis $\phi_j^{(l)}(\zeta_j)$ can be formulated using (3.16). It is worth noting that $\phi_j^{(l)}(\zeta_j)$ can be directly utilized for analysis; however, a normalized polynomial basis offers additional useful properties. As a result, a normalization procedure, as described in [140], is performed to ensure the basis is normalized and enhances its applicability.

$$\psi_j^{(l)}(\zeta_j) = \frac{1}{\|\phi_j^{(l)}(\zeta_j)\|} \sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^n, \quad l = \{0, \cdots, H\} \tag{3.25}$$

where $\|\phi_j^{(l)}(\zeta_j)\|$ is the norm of $\phi_j^{(l)}(\zeta_j)$. By integrating the n-th raw moment of ζ_j in ***Remark*** 3.3.2, $\|\phi_j^{(l)}(\zeta_j)\|$ can be represented by:

$$\begin{aligned}\int_{\zeta_j \in \Omega_j} \left[\phi_j^{(l)}(\zeta_j)\right]^2 dF_{\zeta_j}(\zeta_j) &= \int_{\zeta_j \in \Omega_j} \left[\sum_{n=0}^{l} p_{n,j}^{(l)} \zeta_j^k\right]^2 dF_{\zeta_j}(\zeta_j) \\ &= \sum_{n=0}^{l} \sum_{s=0}^{l} p_{n,j}^{(l)} p_{s,j}^{(l)} \mu_{n+s,j}\end{aligned} \tag{3.26}$$

Once the one-dimensional orthonormal polynomial basis is established, the determination of multi-dimensional orthonormal polynomials becomes straightforward by referring to equation (2.7). In this case, the notation $\phi_j^{(\alpha_j^k)}(\zeta_j)$ is replaced with $\psi_j^{(l)}(\zeta_j)$ to denote the orthonormal polynomials.

Remark 3.3.3. Given (3.26), it can be known that constructing a normalized l-th order orthogonal polynomial basis requires at least $2l$-th order raw moments of ζ_j. These raw moments are crucial for ensuring the orthonormality of the basis.

Remark 3.3.4. It is worth emphasizing that the construction of a univariate orthonormal basis $\psi_j^{(0)}(\zeta_j), \ldots, \psi_j^{(l)}(\zeta_j)$ ($l = 0, \ldots, H$) is only achievable *if and only if* the $2l$-th order raw moments $(\mu_{0,j}, \ldots, \mu_{2l,j})$ exist and the number of support points (distinct values of all sample points) of ζ_j is greater than the desired degree H of the basis, in cases where ζ_j is a discrete random variable or represented by a data set [140]. These requirements ensure the accuracy and validity of the orthonormal basis for subsequent analysis.

Remark 3.3.5. It is essential to understand that in practical power system applications, a low PCE order (e.g., $H = 2$), is often adequate for static-state assessment [86, 89, 93]. When data on random inputs (e.g., wind speed, solar irradiation, and load power) is available, the higher-order raw moments for each input typically exist. However, it is important to recognize that some distributions, particularly those with heavy tails (e.g., the Student's t-distribution), may lack higher-order moments. To address this challenge, distribution truncation strategies can be implemented [197].

In the calculation of PTTC, a large number of correlated mixed random variables $\boldsymbol{\zeta}$ with sample size M_p are considered, where M_p is generally much larger than the desired degree H. As such, the decorrelation strategy is applied and $\boldsymbol{\zeta}$ is transformed to $\boldsymbol{Z}$ through PCA typically satisfies the above condition for a specified degree H, where $\boldsymbol{\zeta}$ is replaced by $\boldsymbol{Z}$ in (2.6) (see ***Remark*** 2.1.3).

Once the orthogonal polynomial bases are built from the data, the next step to construct a surrogate PCE model (2.6) is to calculate the expansion coefficients c_k, $k = 1, \cdots, M$, which are presented in the following sections.

3.3.3 The Calculation of the PCE Coefficients with Sparse Adaptive Scheme

In this section, the hybrid LAR, which combines the OLS and LAR algorithms, is applied to calculate the PCE coefficients. In the hybrid LAR algorithm, a basis-adaptive scheme is achieved based on LAR, where a modified truncation scheme is utilized. The final PCE coefficients are determined by solving the least-square regression using OLS.

The Adaptive Sparse PCE Scheme: Least Angle Regression (LAR)

1) The Modified Truncation Scheme

The multi-dimensional orthonormal polynomials $\Psi_k(\boldsymbol{\zeta})$ are generated via a complete tensor product, resulting in a large increase in the number of polynomial terms as the number of random inputs grows. According to the sparsity-of-effects concept, the bases generated by interactions of low-order input variables are sufficient in most real issues [198]. To minimize computational expense, the hyperbolic (or q-norm) truncation approach is used [78], [198]:

$$\Psi_k(\zeta_1, \cdots, \zeta_{\mathcal{M}}) = \prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(\zeta_j) \tag{3.27}$$

$$\left(\sum_{j=1}^{\mathcal{M}} \left(\alpha_k^j\right)^q \right)^{\frac{1}{q}} \leq H, \quad k = 1, \cdots, \mathcal{M} \tag{3.28}$$

where $q \in (0, 1)$. k denotes the number of expansion terms. As such, a sparser PCE can be produced compared to the standard truncation in (2.7).

2) The LAR Algorithm

LAR, a powerful linear regression tool, is utilized for efficient variable selection. It aims to identify the most relevant predictors (e.g., the polynomial bases $\boldsymbol{\Psi_k}$ in (2.6)) for the model response Y (e.g., the PTTC) from a potentially extensive pool of candidates [199], [5]. The LAR method is employed by incorporating a penalty term $\lambda \|\boldsymbol{C}\|_1$ directly into the least-square minimization problem (2.11), leading to its modification below:

$$\hat{\boldsymbol{C}} = \arg\min \mathbb{E}\left[\sum_{s=1}^{M_p}\left[Y^{(s)} - \sum_{k=0}^{L-1} c_k \Psi_k(\boldsymbol{\zeta}^{(s)})\right]^2\right] + \lambda_{\text{lar}} \sum_{k=0}^{L-1} c_k \tag{3.29}$$

where λ_{lar} is a penalty factor. $\sum_{k=0}^{L-1} c_k = \|\boldsymbol{C}\|_1$ denotes the $\mathcal{L}_1$ norm. Besides, to reduce the computational burden and mitigate the risk of overfitting, the corrected leave-one-out cross-validation error (e_{cloo}) is employed as a stop criterion in the LAR algorithm [5]. Particularly, e_{cloo} is more sensitive to overfitting than e_{loo} (i.e., the leave-one-out cross-validation error), making it a more effective measure in this context. The e_{cloo} is computed using the following formula:

$$e_{\text{cloo}} = T(L, M_p) e_{\text{loo}} \tag{3.30}$$

with

$$\begin{cases} e_{\text{loo}} = \dfrac{\sum_{m=1}^{M_p}\left[\frac{Y^{(m)} - \hat{Y}^{(m)}}{1-h_i}\right]^2}{\sum_{m=1}^{M_p}\left[Y^{(m)} - \hat{\mu}_Y\right]^2} \\ T(L, M_p) = \dfrac{M_p}{M_p - L}\left(1 + \text{trace}\left[\left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\right]\right) \end{cases} \tag{3.31}$$

where h_i is the ith element of $\boldsymbol{h} = \text{diag}\left(\boldsymbol{\Psi}\left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^T\right)$; e_{loo} is the leave-one-out cross-validation error; $\hat{\mu}_Y = \frac{1}{M_p}\sum_{m=1}^{M_p} Y^{(m)}$ is the sample average of the response Y; $T(L, M_p)$ is the correction factor which will increase with the increase of the number of terms L and $T(L, M_p) \to 1$ when the size of samples $M_p \to \infty$. As demonstrated in [199], the LAR algorithm requires $\mathcal{O}(M_p L^2 + L^3)$ computations when $L < M_p$. The LAR algorithm is described in **Algorithm** 1 [5], [198]. Specially, this chapter applies the hybrid LAR, combining LAR and OLS, to determine the polynomial bases and calculate the PCE coefficients. The description of hybrid LAR and a flowchart are presented in Appendix D.

Algorithm 1 The LAR algorithm

1: Initialize $\{c_0, c_1, \cdots, c_{L-1}\} = 0$ and set the initial residual $\boldsymbol{r}_0 = \boldsymbol{Y}$.
2: Find the basis $\Psi_k(\boldsymbol{\zeta})$ which is the most relevant to the current residual.
3: Adjust all the coefficients $\{c_0, \cdots, c_{L-1}\}$ from 0 toward their least-square values of current active set $\{\boldsymbol{\Psi_k}\}$. This process continues until some other regressors $\{\boldsymbol{\Psi_k}\}$ achieve an equicorrelation with the residual.
4: Compute and record the estimated error e_{cloo} for the current iteration.
5: Update all the active coefficients c_k and adjust the predictors Ψ_k from the candidate set to the active set.
6: Continue this process until the number of predictors Ψ_k reaches $L_p = \min(L, M_p)$.

Remark 3.3.6. To efficiently reduce the time consumption of the LAR algorithm, especially in cases with high-dimensional random inputs or elevated PCE orders, an early stopping criterion is pragmatically implemented. Specifically, the algorithm terminates the addition of regressors either upon achieving a predefined accuracy threshold (e.g., 10^{-8} in the simulations) or when the e_{cloo} exceeds its minimum by at least 10% of the maximum iteration limit (e.g., $10\% \times L_p$). Moreover, this early stopping criterion is deactivated for relatively small training sample sizes (e.g., $M_p \leq 100$ in the simulations).

3.4 ATC Computation Approach

This section presents a detailed description of the proposed algorithm for ATC evaluation. The key point is on utilizing the developed DDSPCE method to estimate the probabilistic characteristics of PTTC. Subsequently, the ATC is assessed with a specified confidential level using (3.2). The overall procedures are as follows.

- **Step1.** Input the network data, contingency list, and details of the transaction of interest. Input M_p samples of $\mathcal{M}$ random inputs (e.g., wind speeds, solar irradiance, active load power, and the states of branches) $\boldsymbol{U}_p = (\boldsymbol{U}^{(1)}, \cdots, \boldsymbol{U}^{(M_p)}) \in \mathbb{R}^{\mathcal{M} \times M_p}$, obtained either through historical data set or generated from assumed probabilistic distributions.

- **Step 2.** Define the load-generation vector $\boldsymbol{b}$ (3.5) according to the transaction under study and evaluate PTTC $\boldsymbol{Y_p} = (\boldsymbol{Y}^{(1)}, \cdots, \boldsymbol{Y}^{(M_p)})$ associated with $\boldsymbol{U}_p$ by solving (3.12) using the Voltage Security Assessment Tool (VSAT) of DSATools.

- **Step 3.** Decorrelate the random variables $\boldsymbol{U}_p$ by applying the PCA technique, which transforms $\boldsymbol{U}_p$ into independent samples $\boldsymbol{\zeta}_p = \left(\boldsymbol{\zeta}^{(1)}, \cdots, \boldsymbol{\zeta}^{(M_p)}\right)$. The resulting dataset $[\boldsymbol{\zeta}_p, \boldsymbol{Y_p}]$ is then passed to **Step 4**.

- **Step 4.** Apply the moment-based method as introduced in section 3.3.2 to build the univariate orthonormal polynomial bases $\psi_j^{(l)}(\zeta_j)$ for every random input ζ_j. Specially, define the PCE order $H = (p_0, p_{\max})$. For each random variable ζ_j, $l = \{0, 1, \cdots, H\}$.

  4a) Calculate 0 to $2l$-th moments;

  4b) Determine the univariate polynomial coefficients: $p_{n,j}^{(l)}$, $n = 0, \cdots, l$ by solving (3.23);

  4c) Construct the orthogonal univariate polynomial basis $\phi_j^{(l)}(\zeta_j)$ according to (3.16);

  4d) Build the orthonormal polynomials $\psi_j^{(l)}(\zeta_j)$ by normalizing $\phi_j^{(l)}(\zeta_j)$ by (3.25).

- **Step 5.** Utilize the adaptive procedure described in Section 3.3.3 to construct the data-driven sparse PCE model (3.15) for assessing the PTTC. Set the PCE order $H = p_0$ ($p_0 \leq H \leq p_{\max}$), set $q = q_0$ ($q_0 \leq q \leq q_{\max}$), the PCE model (3.15) can be constructed as follows.

  5a) Truncate the multi-index l using (3.28) and build the corresponding multi-dimensional polynomial bases of degree H to construct the matrix $\boldsymbol{\Psi}$ in (2.11).

  5b) Implement the LAR algorithm to find the optimal sparse polynomial bases.

  5c) Compute the corrected leave-one-out error e_{cloo} according to (3.30) and (3.31). If the e_{cloo} reaches a target error or $e_{\text{cloo},q}^{(H)} \geq e_{\text{cloo},q}^{(H-1)} \geq e_{\text{cloo},q}^{(H-2)}$, store the polynomial bases Ψ_k with the lowest e_{cloo}. Otherwise, increase q and go to Step 5d). If q reaches $q_{\max}$, set $q = q_0$, $p = p + 1$ and return to Step 5a).

  5d) Compute the expansion coefficients c_k by (2.13) based on the polynomial bases $\boldsymbol{\Psi}(\boldsymbol{\zeta})$ with the smallest e_{cloo} and go to **Step 6**.

- **Step 6.** If the data-driven sparse PCE model has achieved the desired accuracy (e.g., $e_{\text{cloo}} < e_{\text{stop}}$), proceed to **Step 7**. Otherwise, increase the size of the training set by ΔM_p (e.g., $\boldsymbol{U_{\Delta p}}$) and calculate $\boldsymbol{Y_{\Delta p}}$ by solving (3.13). Update the variables as follows: $M_p \leftarrow M_p + \Delta p$, $\boldsymbol{U}_p \leftarrow (\boldsymbol{U}_p, \boldsymbol{U_{\Delta p}})$, $\boldsymbol{Y_p} \leftarrow (\boldsymbol{Y_p}, \boldsymbol{Y_{\Delta p}})$, and return to **Step 3**.

- **Step 7.** Once the PCE model (3.15) is constructed, acquire additional M_s sample points of $\boldsymbol{\zeta}$ from historical data or generated by their assumed probability distributions. Apply the PCA technique to decorrelate all samples to $\boldsymbol{\zeta}_s = \left(\boldsymbol{\zeta}^{(1)}, \cdots, \boldsymbol{\zeta}^{(M_s)}\right)$ and compute the PTTC $\boldsymbol{Y_s} = \left(Y^{(1)}, \cdots, Y^{(M_s)}\right)$ based on the established PCE model (3.15).

- **Step 8.** Compute various statistics of the PTTC, such as the mean value, standard deviation, PDF, and CDF.

- **Step 9.** Determine the TRM and the associated ATC value based on a specified confidential level $P_{cl}\%$, i.e., $\mathbb{P}(\mathrm{ATC}_{\mathrm{actual}} \geq (\mathbb{E}[\mathrm{PTTC}] - \mathrm{TRM})) = P_{cl}\%$ and generate the result report.

Remark 3.4.1. The number of M_s in **Step 7** is significantly larger than M_p (i.e., the number of training samples required to build DDSPCE-based model (2.6)) in **Step 2**, demonstrating that the computational effort required for PTTC evaluation through (3.13) is substantially reduced in DDSPCE compared to MCS. It is important to highlight that the majority of the computational cost in DDSPCE is concentrated in **Step 2**. As a result, the DDSPCE method exhibits much higher computational efficiency compared to MCS, particularly due to the fact that $M_p \ll M_s$.

Remark 3.4.2. In the DDSPCE method, the empirical training sample size (M_p in **Steps 3** to **5**), is approximately five times the dimension of random inputs ($5\mathcal{M}$). Therefore, initiating M_p at or near this value can potentially reduce the number of iterations required in the proposed algorithm's execution.

Remark 3.4.3. Control devices such as adjustable transformers and switchable shunts, as well as $N - K$ contingencies, can be easily integrated into the PTTC formulation and ATC assessment. Section 3.5.3 provides numerical examples that illustrate the incorporation of these elements.

3.5 Case Study I – The Modified IEEE 118-Bus System

In this section, the modified IEEE 118-bus system has been used to test the effectiveness of the proposed DDSPCE method for PTTC assessment. Specially, three different scenarios are considered. The first scenario includes only continuous random inputs which are wind speed $\boldsymbol{v}$, solar irradiance $\boldsymbol{r}$ and stochastic loads $\boldsymbol{P_L}$. In contrast to the first scenario, both continuous and discrete random inputs (i.e., line outages) are considered in the second scenario. The necessity of incorporating the discrete random variables in assessing PTTC will be verified, as these variables will significantly affect the stochastic characteristics of PTTC, leading to a reduction of the ATC value. The third scenario further takes into account $N-2$ contingency and adjustable transformers ULTC, respectively, to demonstrate their seamless incorporation into the formulation of PTTC and ATC assessment.

In this case, the probabilistic data used was generated from pre-assumed probability distributions (see `https://github.com/TxiaoWang/DDSPC-TTC.git`). Note that only data is used in the proposed DDSPCE method while the probability distribution information is not used. The proposed DDSPCE method was compared with the SPCE method [93], where the PCE model is built using the known probability distributions. MCS are performed to verify the accuracy and efficiency of the proposed DDSPCE method. For practical applications, the PTTC can be easily evaluated by the proposed DDSPCE method when there is sufficient data for the random inputs.

All simulations have been conducted on the MATLAB R2018b on a PC equipped with Intel Core i7-8700 (3.20GHz), 16GB RAM. The deterministic simulation tool used to calculate exact TTC values is the Voltage Security Assessment Tool (VSAT), a core toolset of DSATools; Toolbox UQLab is adopted to build the DDSPCE scheme [200].

3.5.1 Scenario 1: With Only Continuous Random Inputs

1) Simulation Setup

In this scenario, the IEEE 118-bus system [201], includes 19 generators, 35 synchronous condensers, 177 transmission lines, and 91 loads, which has a total load of 4242 MW and 1438 MVar. It incorporates 111 continuous random variables, including wind speeds $\boldsymbol{v}$, solar irradiance $\boldsymbol{r}$, and load power $\boldsymbol{P_L}$. Specifically, there are six wind farms connected to buses $\{$10, 25, 26, 49, 65, 66$\}$, six solar PV plants connected to buses $\{$12, 59, 61, 80, 89, 100$\}$, and 99 probabilistic loads. The power transfer under study is from the generators at buses $\{$87, 89, 111$\}$ to the loads at buses $\{$88, 90, 91, 92, 103$\}$. The contingency list includes five $N-1$ outages: $\{$L88-89, L7-12, L13-15, L49-54, L91-92$\}$.

2) TTC without Uncertainty

Firstly, the DSATools/VSAT solver is utilized to compute the TTC of the deterministic system, considering voltage limits, thermal limits, generation capacity, and stability limits. The obtained deterministic TTC is 139.9 MW. This value represents the TTC without uncertainty. For detailed information on the deterministic TTC in both normal and contingency cases, please refer to Table 3.1.

Table 3.1 The deterministic TTC in normal and contingency cases in the IEEE 118-bus system

Case No.	Outage Facility	Violation type	TTC(MW)
0	Base case	Max generation violation	299
1	L88-89	Voltage violation	139.9
2	L7-12	Max generation violation	299
3	L13-15	Max generation violation	299
4	L49-54	Max generation violation	299
5	L91-92	Max generation violation	299

3) The Probabilistic Characteristic of PTTC

Subsequently, the proposed DDSPCE method is employed to estimate the probabilistic characteristics of the PTTC. This includes determining key statistics such as the mean, standard deviation, PDF, and CDF of the PTTC. Specially, $M_p = \{278, 417, 556\}$ samples of $\boldsymbol{U}_p = (\boldsymbol{v}, \boldsymbol{r}, \boldsymbol{P_L})$ are applied to the deterministic tool (VSAT/DSATools) and the corresponding PTTC are evaluated. Based on the evaluated M_p sample-response pairs $[\boldsymbol{U}_p, \boldsymbol{Y_p}]$, the PCE model (3.15) is constructed in **Step 3** to **Step 5**. In order to illustrate the process of selecting the PCE order H, and the sample size M_p, Table 3.2 provides a comparison of the corrected leave-one-out error (e_{cloo}), the estimated mean value (μ), the standard deviation (σ) of PTTC, and the normalized estimation errors in percentage for different combinations. It can be observed from the table that as the value of H increases from 2 to 3, the e_{cloo} also increases. Based on this analysis, we conclude that $H = 2$ is the preferred choice for the order of the PCE model. Additionally, the corresponding preferred training sample size is determined to be $M_p = 556$. Hence, in the scenarios thereafter, $H = 2$ and $M_p = 556$.

Table 3.2 Comparison of estimation accuracy of the DDSPCE method for different sample sizes M_p and model orders H

M_p	$H=2$					$H=3$				
	e_{cloo}	μ	$\frac{\Delta\mu}{\mu_{\text{MCS}}}\%$	σ	$\frac{\Delta\sigma}{\sigma_{\text{MCS}}}\%$	e_{cloo}	μ	$\frac{\Delta\mu}{\mu_{\text{MCS}}}\%$	σ	$\frac{\Delta\sigma}{\sigma_{\text{MCS}}}\%$
278	0.0264	138.6281	-1.3652	30.9066	4.2371	0.1509	138.6281	-1.3652	26.3000	-11.2994
417	0.0172	139.4681	-0.7675	30.2735	2.1018	0.0554	139.4681	-0.7675	28.0600	-5.3635
556	0.0120	139.8183	-0.5183	29.7303	0.2698	0.0153	139.8183	-0.5183	28.8977	-2.5383

Remark 3.5.1. Additionally, the curves of e_{cloo} over iterations in the LAR algorithm with PCE order $H = 2$ and $H = 3$ are presented in Fig. 3.3. For example, when $H = 2$, the e_{cloo} drops sharply at the beginning and then enters a phase where it gradually increases, indicating that the model starts to overfit as more iterations are performed. Therefore, in the LAR algorithm, the optimal bases are adaptively selected with the minimum e_{cloo}. Refer to Section 3.3.3 and Appendix D for details of the algorithm.

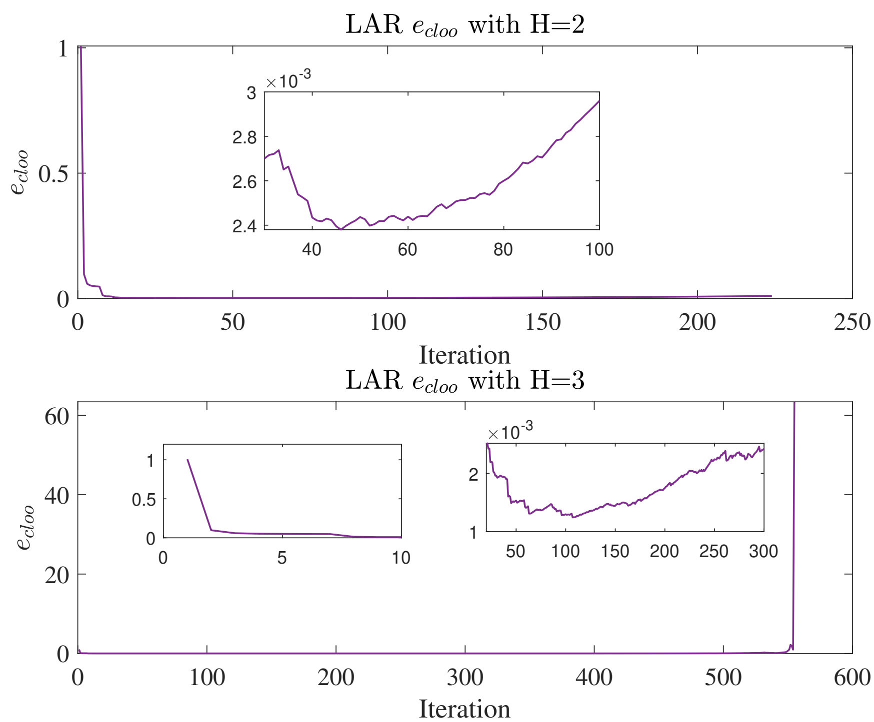

Fig. 3.3 The curves of the corrected leave-one-out error (e_{cloo}) for $H = 2$ and $H = 3$ with the inactivation of the early stopping criteria

The comparison of DDSPCE, SPCE [93], and MCS is presented in Table 5.1, which includes the mean value (μ), the standard deviation (σ), and their normalized errors in percentage (%). It is evident that the proposed DDSPCE method provides highly accurate estimation results compared to the benchmark MCS. Furthermore, Fig. 3.4 displays the comparisons of the estimated probability density function (PDF) and cumulative distribution function (CDF) of the PTTC obtained from $M_s = 10,000$ samples using the MCS, DDSPCE, and SPCE methods. The results from these three methods overlap significantly, indicating the good accuracy of the proposed DDSPCE method. Notably, unlike SPCE, the proposed DDSPCE method does not require any pre-assumed probability distributions of the random inputs.

Table 3.3 Comparison of the estimated statistics of the overall TTC by the MCS, DDSPCE, and SPCE methods

Index	MCS	DDSPCE (proposed)	SPCE [93]
μ	140.5468	139.8183	140.6716
σ	29.6503	29.7303	29.4201
$\frac{\Delta\mu}{\mu_{\mathrm{MCS}}}\%$	–	-0.5183	0.0888
$\frac{\Delta\sigma}{\sigma_{\mathrm{MCS}}}\%$	–	0.2698	-0.7764

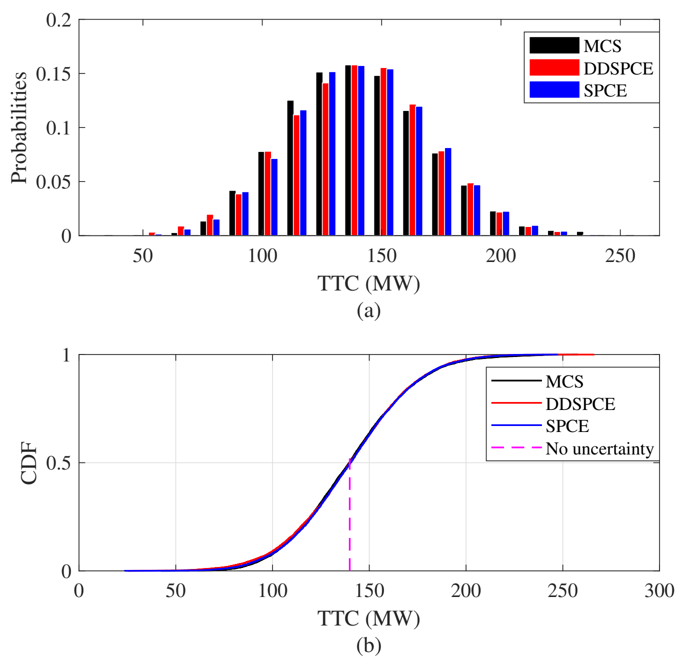

Fig. 3.4 The PDF and CDF of the PTTC calculated by the MCS, DDSPCE and the SPCE. They are almost overlapped. The TRM for 95% confidence level is 49.7043 MW and the corresponding ATC is 90.8425 MW.

4) Efficiency Comparison

Regarding computational efficiency, Table 3.4 provides a comparison of computation times between DDSPCE, SPCE, and MCS. The time for generating the data set $[\boldsymbol{\zeta}_p, \boldsymbol{Y_p}]$ $(M_p = 556)$ is denoted by t_{ed}, the time for constructing the PCE models (i.e., DDSPCE and SPCE) is represented by t_{sc}, and the time for computing the PTTC of $M_s = 10,000$ samples is denoted by t_{es}. It is evident that the DDSPCE method exhibits significantly reduced time consumption compared to MCS, requiring only approximately $\frac{1}{18}$ of the time required by MCS. Moreover, the construction time of the DDSPCE model (t_{sc}) is approximately five times shorter than that of the SPCE method.

Table 3.4 Comparison of Computational Time between DDSPCE, MCS, and SPCE

Method	$t_{\text{ed}}(s)$	$t_{\text{sc}}(s)$	$t_{\text{es}}(s)$	$t_{\text{total}}(s)$
MCS	–	–	173484.23	173484.23
DDSPCE	9368.61	1.09	0.33	9370.03
SPCE	9368.61	5.46	0.35	9374.42

5) ATC Computation

Once the statistical characteristics of PTTC have been obtained, acceptable TRM values can be determined by calculating the difference between the mean value of PTTC and the PTTC value at a specified confidence level. This determination is based on the CDF of PTTC, which allows for the calculation of ATC values. Table 3.5 presents the estimated TRM values corresponding to specific confidence levels, along with their corresponding ATC values. For instance, at a desired confidence level of 95%, where $\mathbb{P}(\text{ATC}_{\text{actual}} \geq (\mathbb{E}(\text{PTTC}) - \text{TRM})) = 0.95$, the calculated TRM is 49.7043 MW, resulting in an ATC value of 90.8425 MW.

Table 3.5 The estimated TRM and resulting ATC (MW) for different confidence levels based on the DDSPCE model

Confid. Level	$\mathbb{E}$(PTTC) (MW)	TRM (MW)	ATC (MW)
99.0%	140.5468	72.0892	68.4576
98.0%	140.5468	62.4110	78.1358
95.0%	140.5468	49.7043	90.8425
90.0%	140.5468	38.2358	102.3110
80.0%	140.5468	25.4402	115.1066

3.5.2 Scenario 2: With Mixed Random Inputs

1) Simulation Setup

In the second scenario, we extend the analysis to incorporate four line outages as discrete random variables into the IEEE 118-bus system. This allows us to evaluate the performance of the proposed method and examine the impacts of these discrete events on PTTC and ATC. The configuration for the continuous random variables, the transaction under study, and the $N-1$ contingency are the same as those in Scenario 1.

In this scenario, a total of 115 random variables are considered, including the four additional line outages: {L89-90, L90-91, L89-92, L92-94}. To simplify the analysis, we assume that the probabilities q_n of unavailability for each line are the same and independent. Specifically, we consider two cases: $q_n = 0.1$ and $q_n = 0.2$, representing different levels of line outage probabilities.

2) The Probabilistic Characteristics of PTTC

Similarly, the deterministic TTC of the system under the defined transaction direction, calculated using VSAT/DSATools, is found to be 139.9 MW. To evaluate the probabilistic characteristics of the PTTC in the presence of discrete events (line outages), we apply the proposed DDSPCE method for two cases: $q_n = 0.1$ and $q_n = 0.2$, representing different levels of line outage probabilities. The order of the PCE model H is set to 2 for both cases, ensuring accurate estimation with reasonable model complexity. To construct the DDSPCE model (3.15), a total of 556 simulations are required (M_p in **Step 3-5**). We generate $10,000$ samples (M_s in **Step 7**) to estimate the probabilistic characteristics of the PTTC using the established DDSPCE model.

The comparison between DDSPCE and MCS is presented in Table 3.6, where the case with $q_n = 0.0$ represents the continuous scenario (Scenario 1) without any line outages. The results demonstrate a substantial reduction in the mean value of PTTC (4.32% when $q_n = 0.1$, 9.14% when $q_n = 0.2$) and an increase in the variance (12.84% when $q_n = 0.1$, 25.65% when $q_n = 0.2$) of PTTC when incorporating the discrete random variables. These findings highlight the necessity and importance of considering discrete events in the assessment of PTTC and ATC to ensure the security and reliability of power grids, especially considering the complex and aging transmission networks we face today.

To assess the performance of the proposed method, Fig. 3.5 compares the results obtained by DDSPCE and MCS for the case of $q_n = 0.1$. The figure clearly illustrates that the proposed

Table 3.6 Comparison of the estimated statistics of the overall TTC (MW) by the MCS and DDSPCE methods for mixed case

q_n	MCS		DDSPCE			
	μ	σ	μ	$\frac{\Delta\mu}{\mu_{\mathrm{MCS}}}\%$	σ	$\frac{\Delta\sigma}{\sigma_{\mathrm{MCS}}}\%$
0.0	140.5468	29.6503	139.8183	-0.5183	29.7303	0.2698
0.1	134.4748	33.4574	133.8207	-0.4864	32.8810	-1.7228
0.2	127.6986	37.2558	127.1441	-0.4342	34.9234	-6.2605

DDSPCE method provides accurate estimates for the PTTC with significantly reduced computational time compared to MCS. In fact, the computational time required by DDSPCE is about $\frac{1}{18}$ of that required by MCS. Similar reliable results can be achieved for the case of $q_n = 0.2$ as well. These findings further highlight the computational efficiency and accuracy of the proposed method in handling mixed random inputs, making it a favorable choice for practical applications.

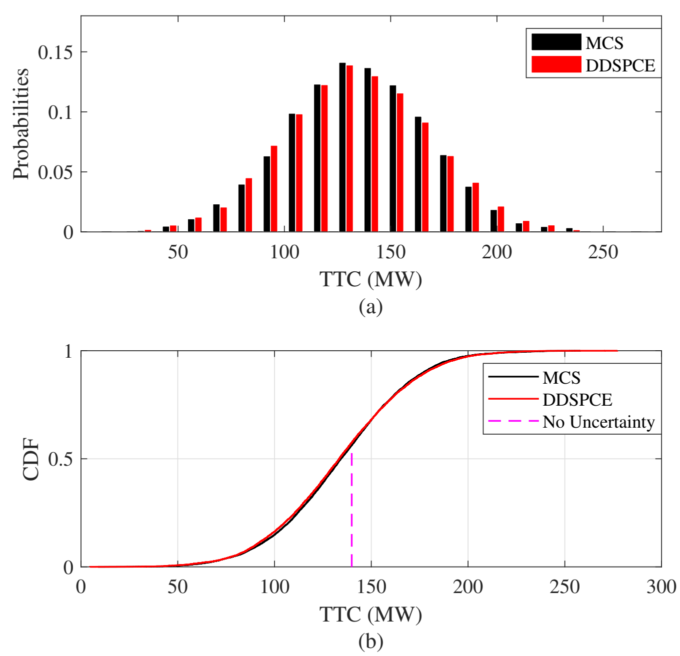

Fig. 3.5 The PDF and CDF of the PTTC calculated by the MCS and DDSPCE with probability $q_n = 0.1$. They are almost overlapped. The TRM for 95% confidence level is 56.4333 MW and the corresponding ATC is 78.0415 MW.

3) ATC Computation

Based on the probabilistic analysis of PTTC, the TRM and the corresponding ATC can be determined. Table 3.7 presents the calculated TRM values and resulting ATC values at a 95% confidence level. Notably, considering the presence of discrete random variables, the ATC level is significantly reduced, with a decrease of 14.09% when $q_n = 0.1$. This emphasizes the crucial importance of incorporating discrete events in ATC assessment, as it has a substantial impact on the overall system reliability and security.

Table 3.7 The estimated TRM and resulting ATC (MW) for confidence level at 95% based on the DDSPCE model

q_n	$\mathbb{E}$(PTTC) (MW)	TRM (MW)	ATC (MW)
0.0	140.5468	49.7043	90.8425
0.1	134.4748	56.4333	78.0415
0.2	127.6986	61.3488	66.3498

3.5.3 Scenario 3: Considering $N-2$ Contingency and Adjustable Transformer

1) Simulation Setup

The third scenario demonstrates the capability of incorporating $N-2$ contingencies and control devices, such as adjustable transformers (e.g., Under-Load Tap Changer (ULTC) transformers), in the proposed PTTC formulation and ATC assessment. In the subsequent case studies, the configuration for continuous random variables and the specific transaction being analyzed remains consistent with Scenario 1. For additional parameters and details, please refer to the previous GitHub link.

2) With $N-2$ Contingency Considered

In this case, the contingency list includes five $N-2$ outages, as shown in Table 3.8. The deterministic TTC under the transaction, taking into account the $N-2$ contingencies, is determined to be 88.8 MW. This value is significantly lower than the deterministic TTC of 139.9 MW observed in Scenario 1, where only $N-1$ contingency was considered.

Table 3.8 $N-2$ contingency list

Case No.	1	2	3	4	5
Outages	L88-89 L89-92	L7-12 L12-16	L13-15 L94-96	L49-54 L42-49	L91-92 L92-93

Next, the proposed DDSPCE model (2.6) is utilized to calculate the probabilistic characteristics of PTTC following the same procedures as in the previous scenarios (i.e., $H = 2$, $M_p = 556$, $M_s = 10000$). The estimated mean, standard deviation, and their normalized errors in $\%$, computed by the MCS, DDSPCE, and SPCE methods, are provided in Table 3.9. It is evident that when considering $N-2$ contingency, the DDSPCE model can accurately estimate the statistics of PTTC. Furthermore, the computational time required by the DDSPCE model is only about $\frac{1}{18}$ of the time required by MCS. It can be anticipated that the proposed PTTC formulation can also accommodate $N-K$ contingency scenarios ($K > 2$), if necessary, while the DDSPCE method can efficiently provide accurate estimations for the probabilistic characteristics of PTTC.

Table 3.9 Comparison of the estimated statistics of the overall TTC with $N-2$ contingency considered by the MCS, DDSPCE and SPCE methods

Method	μ	σ	$\frac{\Delta\mu}{\mu_{\text{MCS}}}\%$	$\frac{\Delta\sigma}{\sigma_{\text{MCS}}}\%$
MCS	90.2270	23.3703	–	–
DDSPCE	89.6840	23.4123	-0.6018	0.1797
SPCE	90.3824	22.8604	0.1722	-2.1818

3) With ULTC Transformer

In this case, the transformer between Bus 81 and Bus 80 (with a fixed ratio η of 0.935 in per unit) is replaced by an Under-Load Tap Changer (ULTC) transformer to maintain the voltage magnitude within a specified range. The ULTC transformer has a lower tap ratio η limit of 0.775 in per unit and an upper tap ratio limit of 1.185 in per unit. The deterministic TTC value, calculated using the DSATools/VSAT solver, is found to be 134.6 MW. This value represents a decrease of 5.3 MW compared to the deterministic TTC value of 139.9 MW in Scenario 1, where no ULTC transformer was considered (see Fig. 3.6 (b)). This decrease in TTC is expected as a trade-off to maintain the voltage level within the desired range.

Next, the proposed DDSPCE model is employed to calculate the statistical properties of PTTC, following the same procedures as in the previous scenarios (i.e., $H = 2$, $M_p = 556$, $M_s = 10,000$). The estimation results, as well as the comparisons between DDSPCE, SPCE, and MCS, are summarized in Table 3.10. The results demonstrate that even when incorporating the ULTC transformer, the DDSPCE method can still provide accurate estimations for the probabilistic characteristics of PTTC. Moreover, similar computational efficiency is observed in this case, with the DDSPCE method requiring approximately $\frac{1}{18}$ of the time consumed by MCS. Likewise, the determination of

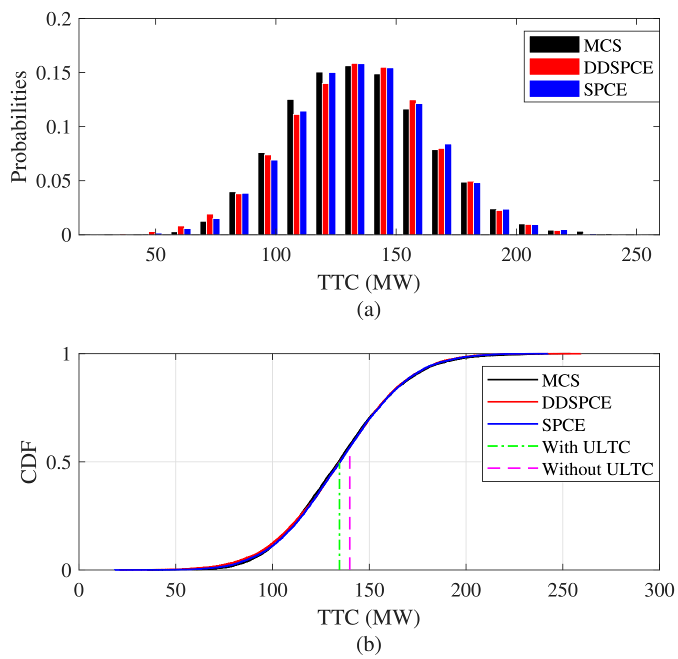

Fig. 3.6 The PDF and CDF of the PTTC calculated by the MCS, DDSPCE, and SPCE with ULTC transformer at Bus 81-Bus 80. They are almost overlapped. The TRM for a 95% confidence level is 49.4894 MW, and the corresponding ATC is 85.6214 MW. The deterministic TTC considering ULTC (the green dash-dot line) is slightly less than the deterministic TTC without ULTC (the magenta dashed line) obtained in Scenario 1.

the Total Reserve Margin (TRM) and the resulting Available Transfer Capability (ATC) can be derived from the calculated probabilistic characteristics of PTTC. Table 3.11 provides a comparison of the estimated TRM and the resulting ATC at a 95% confidence level, considering both scenarios with and without the ULTC. It is observed that the two cases exhibit similar TRM values when considering the same set of random inputs. However, due to the lower mean value of PTTC obtained

Table 3.10 Comparison of the estimated statistics of the overall TTC with ULTC considered by the MCS, DDSPCE, and SPCE methods

Method	μ	σ	$\frac{\Delta\mu}{\mu_{\mathrm{MCS}}}\%$	$\frac{\Delta\sigma}{\sigma_{\mathrm{MCS}}}\%$
MCS	135.1108	29.4460	–	–
DDSPCE	134.4281	29.5267	-0.5053	0.2741
SPCE	135.2722	29.2345	0.1195	-0.7183

after integrating the ULTC, a slightly lower ATC is obtained in this case. The incorporation of the ULTC transformer affects the TTC by ensuring voltage stability but may lead to a reduced ATC. These results highlight the significance of considering control devices such as ULTC transformers in ATC assessment, as it provides valuable information about the available transfer capability under operational constraints.

Table 3.11 Comparison of the estimated TRM and resulting ATC at 95% confidence level with and without ULTC by the DDSPCE model

Control device	$\mathbb{E}$(PTTC) (MW)	TRM (MW)	ATC (MW)
Without ULTC	140.5468	49.7043	90.8425
With ULTC	135.1108	49.4894	85.6214

* The results without ULTC are obtained directly from Scenario 1.

3.6 Case Study II – The Modified PEGASE 1354-Bus System

3.6.1 Simulation Setup

This section extends the proposed method to a large-scale power system, the Modified PEGASE 1354-Bus System, to test its performance. This system accurately depicts the magnitude and intricacy found within a segment of the European high-voltage transmission network. This network has 1,354 buses, 260 generators, and 1,991 branches, operating at voltage levels of 380 and 220 kV. The power demand of this network amounts to 73,060 MW, accompanied by a reactive power demand of 13,401 Mvar [202], [203]. In order to evaluate the influence of uncertainties on PTTC and ATC for this network, 6 wind generators, each with a capacity of 100MW, and 6 solar photovoltaic (PV) systems, each with a capacity of 100MW, have been integrated into the system. There are 236 total random inputs including 224 stochastic loads. The new transaction under study is provided in `https://github.com/TxiaoWang/DDSPC-TTC.git`.

Similarly, The probabilistic data utilized in this study was generated based on predetermined probability distributions (refer to the Github link for more details). It is important to note that the proposed DDSPCE method solely relies on the data itself and does not utilize probability distribution information. To compare its performance, the proposed DDSPCE method was contrasted with the SPCE method as described in [93], where the PCE model is constructed using known probability distributions. The MCS was conducted to validate the accuracy and efficiency of the proposed DDSPCE method. In practical applications, the proposed DDSPCE method can be employed to easily evaluate the PTTC when there is an adequate amount of data available for the random inputs.

3.6.2 The Probabilistic Characteristics of PTTC

Initially, the DSATools/VSAT solver is employed to calculate the TTC of the deterministic system, taking into account factors such as voltage limits, thermal limits, generation capacity, and stability limits. The resulting deterministic TTC is determined to be 481 MW. This value represents the TTC in the absence of any uncertainties. For a comprehensive breakdown of the deterministic TTC in both normal and contingency scenarios, please consult Table 3.12 for detailed information.

Table 3.12 The deterministic TTC in normal and contingency cases in the PEGASE 1354-bus system

Case No.	Outage Facility	Violation type	TTC(MW)
0	Base case	overload	1368.4
1	L8030-5837	overload	481
2	L1758-352	overload	1368.3
3	L5837-449	overload	1366.2
4	L5837-1817	overload	1304.2
5	L5837-2458	overload	698.7
6	T7056-9051	overload	1368.3
7	L7438-1526	overload	1368.3
8	L7438-5334	overload	1368.3

Subsequently, the proposed DDSPCE method is employed to determine the statistical properties of the PTTC using the same procedures as in the previous cases. The parameters used for this analysis include $H = 2$ (order of the PCE expansion), $M_p = 1180$ (number of polynomial terms), and $M_s = 50,000$ (number of samples). The estimation results, along with comparisons between DDSPCE, SPCE, and MCS, are summarized in Table 3.13 and Fig. 3.7. The results showcase that when extending to a large-scale system, the DDSPCE method can still provide accurate estimations for the probabilistic characteristics of PTTC. Moreover, high computational efficiency is observed in this case, with the DDSPCE method requiring approximately $\frac{1}{42}$ of the time consumed by MCS.

Table 3.13 Comparison of the estimated statistics of the overall TTC by the MCS, DDSPCE, and SPCE methods for the modified PEGASE 1354-bus system

Index	MCS	DDSPCE (proposed)	SPCE [93]
μ	482.9215	483.3158	482.8385
σ	65.7249	63.3901	65.1749
$\frac{\Delta\mu}{\mu_{\mathrm{MCS}}}\%$	–	0.0817	-0.0172
$\frac{\Delta\sigma}{\sigma_{\mathrm{MCS}}}\%$	–	-3.5525	-0.8300

3.6.3 ATC Computation

Once the probabilistic characteristics of PTTC have been assessed, TRM can be easily determined by specifying certain confidence levels and their corresponding ATC can be calculated. Table 3.14 showcases the calculated TRM values and the resulting ATC values at different confidence levels. This underscores the crucial importance of integrating discrete events in ATC assessments, as they have a substantial impact on the overall reliability and security of the system.

Table 3.14 The estimated TRM and resulting ATC (MW) for different confidence levels based on the DDSPCE model

Confid. Level	$\mathbb{E}$(PTTC) (MW)	TRM (MW)	ATC (MW)
99.0%	482.9215	149.6468	333.2747
98.0%	482.9215	131.5078	351.4137
95.0%	482.9215	105.1758	377.7465
90.0%	482.9215	81.7558	401.1657
80.0%	482.9215	53.2068	429.7147

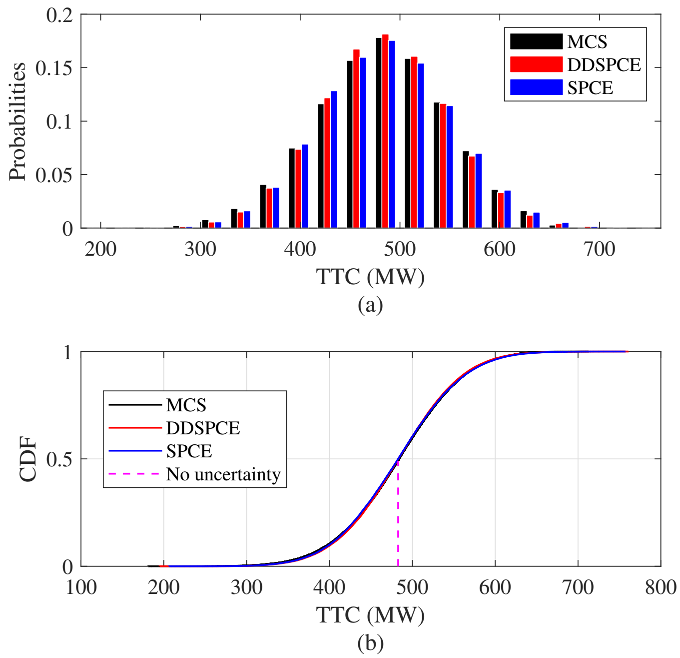

Fig. 3.7 The PDF and CDF of the PTTC calculated by the MCS, DDSPCE, and SPCE for the modified PEGASE 1354-bus system. They are almost overlapped. The TRM for a 95% confidence level is 105.1758 MW, and the corresponding ATC is 377.7456 MW.

3.7 Conclusions

In this chapter, we present a novel approach called the data-driven sparse Polynomial Chaos Expansion (DDSPCE) method for assessing the probabilistic characteristics of PTTC without the need for pre-assumed probability distributions of random inputs such as RES, load variation, and line outages. Based on the estimated probability distribution of PTTC, ATC values at certain confidence intervals are evaluated. The proposed DDSPCE method directly leverages data sets and can effectively handle a large number of mixed random inputs, including continuous and discrete variables.

Furthermore, the sparse PCE scheme is integrated to reduce computational effort while maintaining accuracy. The simulation results demonstrate that the proposed DDSPCE method accurately estimates the probabilistic characteristics of PTTC with high efficiency. Finally, we highlight the significance of incorporating discrete random variables in PTTC and ATC assessment, as they have a substantial impact on the statistics of PTTC and can significantly reduce the ATC level.

Chapter 4

A Data-Driven Sparse PCE Method for UQ in Economic Dispatch

This chapter delves into the impacts of stochastic RESs on the power system economic dispatch (ED). The previously proposed DDSPCE-based model is further extended to approximate the objective values for ED problems. This extension considers the randomness introduced by wind power generators, adding an additional layer of complexity. An integrated electric and gas system is utilized to validate the proposed model, incorporating real-world wind power data. This real-world dataset adds further challenges to the analysis and evaluation. The ideas and results presented in this chapter are primarily based on the published work [J2].

4.1 Introduction

As discussed in Section 1.2.2, the growing integration of RESs has underscored the need for accurate quantification of the uncertainties in power system ED. To tackle the challenge of uncertainty in the ED problem, several techniques have been proposed, including stochastic programming, specifically stochastic-optimization ED (SED) formulations (e.g., [106]), robust-optimization formulation (e.g., [98]), chance-constrained optimization (e.g., [204]), and more. However, solving the optimization problem based on the aforementioned formulations typically requires a large number of MC simulations, resulting in computational tractability issues. This becomes particularly challenging when uncertainties are significant, as existing approaches based on MC may fail to provide meaningful information [205], [206].

To alleviate the computational burden, various surrogate modeling-based UQ methods have been employed for SED problems. Safta et al. [37] utilized the PCE method to construct a surrogate model for the SED problem. Their approach demonstrated the capability to obtain accurate results efficiently using a reduced number of samples compared to MC-based methods. Li et al. [141] extended the work by adopting an SPCE method to mitigate the “curse of dimensionality” in the surrogate modeling of the SED problem. However, both the surrogate models proposed in [37], and [141] assume that the random inputs follow Gaussian distributions, which may not hold true in practical scenarios. In a more recent study, Hu et al. [207] proposed a Gaussian process emulator-based approach for solving the SED problem. Nonetheless, it has been acknowledged by Rajabi et al. [208] that the PCE method may outperform the Gaussian process emulator-based approach when the response’s probability distribution, such as the SED solution, exhibits multimodality. This chapter extends the DDSPCE method to solve the SED problem considering uncertainties modeling by real-world input data. The main contributions of this chapter are as follows.

- The proposed DDSPCE method constructs a surrogate model directly from raw data of random variables, without making any prior assumptions about the marginal distributions of the variables or the output responses.

- The proposed DDSPCE-based surrogate model enables accurate estimation of statistical information such as mean, variance, PDF, and CDF of the SED solution. Notably, the DDSPCE method achieves high computational efficiency, being approximately 33 times faster than Monte Carlo simulations on an integrated IEEE 118-bus power system and a 20-node gas system [209].

- The DDSPCE method maintains its accuracy even when faced with multimodal PDFs of the SED solution.

The remainder of this chapter is organized as follows: Section 4.2 provides the mathematical formulation of the SED problem. Section 4.3 elaborates on the proposed DDSPCE method for approximating the SED solutions. Section 4.5 presents the numerical study on an integrated IEEE 118-bus power system and a 20-node gas system. Section 4.6 gives the conclusions.

4.2 Mathematical Formulation of Stochastic-Optimization Economic Dispatch

A two-stage modeling approach commonly addresses the stochastic UC problem. In this approach, the problem is divided into two stages: the first stage involves determining the UC decisions in the day-ahead electric market. In contrast, the second stage focuses on optimizing the dispatch decisions [209]. This two-stage modeling framework allows for better decision-making under uncertainty and enables more efficient utilization of available resources. In this chapter, it is assumed that the first stage UC decisions for conventional generating units have already been determined using a day-ahead UC model, as described in previous studies [37], [207], [108]. These pre-determined UC decisions serve as input for the subsequent stages of the analysis. The goal of this chapter is to solve the multi-period ED problem with fixed UC decisions while considering the uncertainties from RES (e.g., wind power output). As discussed in [37], the stochastic renewable generation power output is modeled as random fields and is represented as functions of a vector of random variables $\boldsymbol{P_w}(v, t)$, which can be approximated by Karhunen-Loève expansions (KLE). Herein, v denotes the random sources (e.g., wind speed), and $\boldsymbol{P_w}$ is the corresponding generator active power output vector (e.g., wind generator power output in this chapter). However, as discussed in Chapter 3, the historical data may be easily assessed, thus, real-world wind power output data is utilized in this chapter.

The ED problem aims to minimize the operating cost $Q(\boldsymbol{P_g})$ by allocating the total demand among generating units, subject to physical and operational constraints [32]. Here, $\boldsymbol{P_g}$ represents the vector of the optimal power output of conventional generators. Traditionally, the economic dispatch problem is formulated as a deterministic optimization problem, where $Q(\boldsymbol{P_g})$ is considered to be a fixed value. However, when considering the uncertainties associated with the power output of renewable generation units, the operating cost becomes a random variable denoted as $Q(\boldsymbol{P_g}, \boldsymbol{P_w})$. Define the expected minimum cost as $\mathbb{E}[Q(\boldsymbol{P_g}, \boldsymbol{P_w})]$, which can be reformulated as [37]:

$$\min_{\boldsymbol{P_g}} \frac{1}{|\boldsymbol{\mathcal{S}}|} \sum_{s_{\mathrm{w}} \in \boldsymbol{\mathcal{S}}} Q(\boldsymbol{P_g^{s_{\mathrm{w}}}}, \boldsymbol{P}_w^{s_{\mathrm{w}}}) \tag{4.1}$$

where $\boldsymbol{\mathcal{S}}$ represents the sets of wind power samples (i.e., scenarios). The vectors $\boldsymbol{P_g^{s_{\mathrm{w}}}}$ and $\boldsymbol{P}_w^{s_{\mathrm{w}}}$ correspond to the power outputs from conventional generators and wind power generators, respectively, for a given sample s_{w}. $Q(\boldsymbol{P_g^{s_{\mathrm{w}}}}, \boldsymbol{P}_w^{s_{\mathrm{w}}})$ is the solution of the multi-period SED problem in

(4.2)-(4.3), which is obtained under the assumption of fixed UC decisions:

$$Q(\boldsymbol{P_g^{s_w}}, \boldsymbol{P_w^{s_w}}) = \min_{\boldsymbol{P_g}} \sum_{t \in T} \sum_{g \in G} C_g(P_g^{t,s_w}) \tag{4.2}$$

s.t.

$$\sum_{g \in G} P_g^{t,s_w} + \sum_{w \in W} P_w^{t,s_w} = \sum_{d \in D} P_d^t \quad \forall t \in T, s_w \in \boldsymbol{\mathcal{S}} \tag{4.3a}$$

$$\underline{P_l} \leq \sum_{g \in G} k_{lg} P_g^{t,s_w} + \sum_{w \in W} k_{lw} P_w^{t,s_w} - \sum_{d \in D} k_{ld} P_d^t \leq \overline{P_l} \quad \forall t \in T, s_w \in \boldsymbol{\mathcal{S}} \tag{4.3b}$$

$$P_g^{\min} x_g^t \leq P_g^{t,s_w} \leq P_g^{\max} x_g^t \quad \forall g \in G, t \in T, s_w \in \boldsymbol{\mathcal{S}} \tag{4.3c}$$

$$\begin{aligned} &- R_g^{RD} x_g^t - R_g^{SD}(x_g^{t-1} - x_g^t) - P_g^{\max}(1 - x_g^{t-1}) \\ &\leq P_g^{t,s_w} - P_g^{t-1,s_w} \leq R_g^{RU} x_g^{t-1} + R_g^{SU}(x_g^t - x_g^{t-1}) + P_g^{\max}(1 - x_g^t) \quad \forall g \in G, t \in T, s_w \in \boldsymbol{\mathcal{S}} \end{aligned} \tag{4.3d}$$

where (4.2) denotes the objective function, which aims to minimize the total operating cost in the multi-period SED problem. t is the specified time period considered from the set of time periods set $T = 1, \cdots, T_m$, representing, for example, a 24-hour period in the simulation study of this chapter. g denotes the generator index, and G represents the set of generators. The constraints (4.3) represent the operational and physical constraints based on the DC power flow model. The power balance constraint (4.3a) ensures that the total demand, $\sum_{d \in D} P_d^t$, at time t is met, where P_d^t is the d-th load demand. The power flow limits are defined in constraint (4.3b), where k_{lg}, k_{lw}, and k_{ld} are the sensitivity coefficients for the l-th transmission line with respect to the traditional generator g, wind generator w, and load d, respectively [210]. A detailed description of DC power flow can be found in Appendix E. The lower and upper bounds of thermal limits of transmission line l are represented by $\underline{P_l}$ and $\overline{P_l}$, respectively. The generation capacity limits are captured by constraint (4.3c), where x_g^t represents the pre-determined UC decision for generator g at time t. The ramping capability constraint of generator g is described by constraint (4.3d), which includes the ramping down rate R_g^{RD}, ramping up rate R_g^{RU}, shut-down ramp rate R_g^{SD}, and start-up ramp rate R_g^{SU}.

Remark 4.2.1. It should be noted that in this chapter, the gas system is integrated into the power system for the case study. Additional constraints related to the gas network are considered (refer to Appendix E for details). However, the essence of the SED problem remains unchanged. For more comprehensive information, please refer to [209].

The SED problem inherently involves complex constraints and optimization. To evaluate the cost associated with load variation, wind generation, and other stochastic factors, MC simulations are commonly employed. However, these simulations require a large number of realizations and can be computationally demanding, even with scenario reduction techniques. This chapter extends the proposed DDSPCE-based surrogate model to serve as a surrogate model, which is capable of estimating the expected minimum cost, $\mathbb{E}[Q(\boldsymbol{P_g}, \boldsymbol{P_w})]$, with significantly fewer samples compared to traditional MC-based approaches. By leveraging wind power data, the proposed method can effectively estimate various statistical measures, including the PDF and CDF of the minimum cost. By applying the DDSPCE-based surrogate model, an accurate cost estimation while mitigating the computational burden associated with extensive MC simulations can be achieved.

4.3 A DDSPCE-based Model for SED

This section extends the DDSPCE-based surrogate model to capture the relationship between the input random variables $\boldsymbol{P_w}$ (representing wind power) and the minimum cost $Q(\boldsymbol{P_g}, \boldsymbol{P_w})$. By employing a linear combination of multivariate orthogonal polynomials, the proposed DDSPCE-based model provides an accurate approximation of the minimum cost while significantly reducing the computational effort required. By utilizing the DDSPCE-based surrogate model, we can achieve a comprehensive understanding of how wind power variability impacts the ED problem.

This approach facilitates efficient and reliable decision-making in power system operation by considering the uncertainties associated with wind power generation. The DDSPCE-based surrogate model serves as a valuable tool for effectively analyzing and managing the effects of wind power uncertainties in power system planning and operation.

4.3.1 The PCE Representation for SED

Consider a model $Y = G(\boldsymbol{\zeta})$ with $\boldsymbol{\zeta} = \{\zeta_1, \cdots, \zeta_{\mathcal{M}}\}$ being a random input vector. As a result, Y is also a random variable due to the randomness of inputs $\boldsymbol{\zeta}$. In this chapter, the random inputs under study are wind generator power output $\boldsymbol{P_w}$, while the minimum cost Q discussed in Section 4.2 is the system response. As introduced in Chapter 2, a stochastic response $Y = G(\boldsymbol{\zeta})$ with finite second-order moment can be approximated by the orthogonal polynomial bases of $\boldsymbol{\zeta}$ (4.4):

$$Q(\boldsymbol{\zeta}) = Y = G(\boldsymbol{\zeta}) \approx G^{\mathrm{pc}}(\boldsymbol{\zeta}) = \sum_{k=0}^{L-1} c_k \Psi_k(\boldsymbol{\zeta}) \tag{4.4}$$

where c_k are unknown PCE coefficients to be calculated; the multivariate orthogonal polynomial bases $\Psi(\boldsymbol{\zeta})$, as discussed in Chapter 2, are required to be orthogonal with respect to the joint PDF of $\boldsymbol{\zeta}$. It should be noted that the series in (4.4) converges to Q when $L \to \infty$ in the sense of $\mathcal{L}^2$-norm. Typically, $\Psi(\boldsymbol{\zeta})$ is constructed using (2.7) and (2.8) via appropriate truncation schemes for practical applications.

In this chapter, the proposed DDSPCE-based model is applied to solve the SED problem. Figure 4.1 illustrates the essence of the proposed DDSPCE method and its connection to the original SED problem (4.2)-(4.3). Instead of conducting extensive MC simulations on the SED problem with a large number of scenarios, the proposed approach involves evaluating $Q(\boldsymbol{P_g}, \boldsymbol{P_w})$ using a small number of $\boldsymbol{P_w}$ samples. These input-response pairs are then utilized to construct a DDSPCE-based surrogate model (4.4), which is a purely algebraic equation. This surrogate model is capable of quickly estimating the production cost $Q(\boldsymbol{P_g}, \boldsymbol{P_w})$ for a large number of $\boldsymbol{P_w}$ scenarios by substituting the corresponding values into the established DDSPCE-based surrogate model (4.4). Compared to the original SED problem (4.2)-(4.3), the DDSPCE-based surrogate model offers significantly faster evaluation, resulting in substantial computational savings. Building the DDSPCE-based surrogate model involves two main tasks: constructing the multidimensional polynomial basis $\Psi_k(\boldsymbol{\zeta})$ using the available $\boldsymbol{P_w}$ data and computing the coefficients c_k in (4.4).

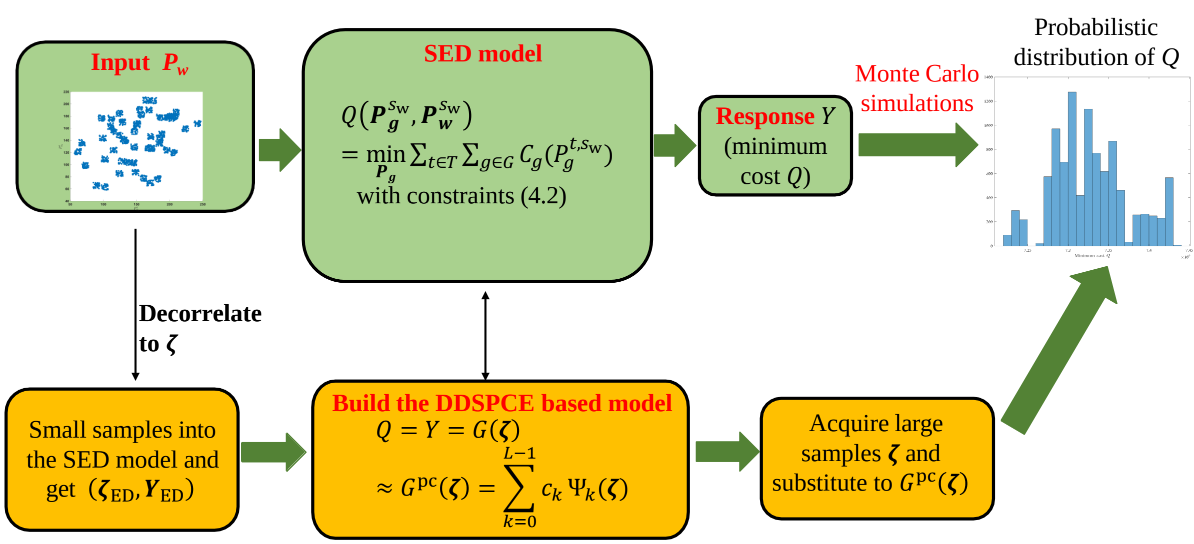

Fig. 4.1 Relation between the SED and DDSPCE-based surrogate model

4.3.2 The Construction of Univariate Orthonormal Polynomial Bases

To build the PCE model, the first step is to construct the multivariate polynomial bases $\Psi_k(\boldsymbol{\zeta})$, as discussed in Section 2.1.2, which is generally constructed via the tensor product of univariate polynomial bases $\phi_j^{(\alpha_j^k)}$ of ζ_j using (2.7). As such, the focus to build $\Psi_k(\boldsymbol{\zeta})$ is turned to determine $\phi_j^{(\alpha_j^k)}$. In the SED problem, considering that the distributions of random inputs, such as the wind power output $\boldsymbol{P_w}$ may be inaccurate or unknown in advance, therefore, the moment-based method (introduced in Section 3.3.2) is exploited to construct the univariate polynomial bases $\phi_j^{(\alpha_j^k)}$. In this method, by utilizing the data directly, without any preassumed probability distribution models, the coefficients of univariate polynomial basis can be determined by the coefficients matrix equation (3.23). Once the univariate polynomial bases $\phi_j(\zeta_j^{(\alpha_j^k)})$ have been constructed, the normalization procedure can be applied by using (3.25) to enhance its applicability.

Remark 4.3.1. In general, there are correlations in $\boldsymbol{P_w}$, which need to be addressed through decorrelation techniques such as PCA (see Appendix A), i.e., decorrelated random inputs $\boldsymbol{\zeta} = \mathcal{T}_{\text{pca}}(\boldsymbol{P_w})$. As such, multivariate orthogonal polynomial bases $\Psi_k(\zeta_1, \cdots, \zeta_{\mathcal{M}})$ can be directly constructed using the tensor product of univariate polynomial bases (i.e., (3.27)) together with the q-norm truncation (3.28). It is important to point out that when dealing with random inputs characterized by thick-tailed distributions or nonlinear correlations, the use of PCA may yield biased results. To address this issue, alternative methods such as vine copula and Rosenblatt transform (see Appendix B) can be employed to model the dependence structure and decouple the input data components [38], [211].

4.3.3 The Calculation of PCE coefficients c_k

Once the polynomial bases $\Psi_k(\boldsymbol{\zeta})$ have been constructed, the orthogonal matching pursuit (OMP) method is applied to calculate the PCE coefficients c_k. The OMP algorithm is an iterative approach used to select regressors that exhibit the highest correlation with the current approximation residual. In each iteration, the algorithm adds these regressors to the active set of basis functions and updates the coefficients c_k for all active regressors using the OLS method (see Section 2.1.3). To determine the model order H and the sparse candidate basis, the OMP algorithm employs the leave-one-out cross-validation error e_{loo} error estimator, as described in (3.31). A detailed explanation of the OMP procedure can be found in **Algorithm** 2 [198]. By leveraging the OMP algorithm, a sparse PCE-based surrogate model can be achieved, which significantly reduces computational effort while ensuring accuracy.

Algorithm 2 The OMP algorithm [198]

1: Initialize $\{c_0, c_1, \cdots, c_{L-1}\} = 0$ and set the initial residual $\boldsymbol{r}_0 = \boldsymbol{Y}$. Define the candidate and active sets as $\Psi_{\mathcal{C}}^{(0)} = \Psi_k$ and $\Psi_{\mathcal{A}}^{(0)} = \varnothing$. $m \geq 1$ indicates the current iteration.
2: Find the basis $\Psi_k^{(m)}(\boldsymbol{\zeta})$ which is the most relevant to the current residual.
3: Include the polynomial $\Psi_k^{(m)}$ into the current active set $\Psi_{\mathcal{A}}^{(m)}$.
4: Adjust and update the PCE coefficients $c_k^{(m)}$ using OLS in (2.13) based on the active set $\Psi_{\mathcal{A}}^{(m)}$.
5: Update the residual r_m = $\boldsymbol{Y} - \Psi_{\mathcal{A}}^{(m)} \boldsymbol{C}^{(m)}$.
6: Update and save the current error e_{eloo} calculated using (3.31).
7: Continue this process until the number of predictors Ψ_k reaches $L_p = \min(L, M_p)$.

4.4 Evaluation of the Minimum Cost Q of SED

This section provides a detailed description of the proposed algorithm for solving the SED problem, particularly for the minimum cost Q. Specially, the developed DDSPCE is used to estimate the probabilistic characteristics of the minimum cost Q.

- **Step 1.** Input the network data, and the M_p samples of random inputs (e.g., wind generator power output $\boldsymbol{P_{w_p}} \in \mathbb{R}^{M_p \times \mathcal{M}}$.

- **Step 2.** Solve the stochastic-optimization ED in (4.2) and (4.3) to obtain the response $\boldsymbol{Y_p} = (Y^{(1)}, \cdots, Y^{(M_p)})$ (e.g., the minimum cost Q).

- **Step 3.** Decorrelate random inputs $\boldsymbol{P_{w_p}}$ using the PCA technique (see ***Remark*** 2.1.3 and Appendix A), which converts $\boldsymbol{P_{w_p}}$ to independent samples $\boldsymbol{\zeta_p} = \left(\boldsymbol{\zeta}^{(1)}, \cdots, \boldsymbol{\zeta}^{(M_p)}\right)$. Pass the data set $[\boldsymbol{\zeta_p}, \boldsymbol{Y_p}]$ to **Step 4**.

- **Step 4.** Apply the moment-based method discussed in Section 3.3.2 to construct the univariate orthonormal polynomial bases $\phi_j^{(\alpha_j^k)}$ for each ζ_j. Specially, define the PCE order $H = (p_0, p_{\max})$. For each random variable ζ_j, $(\alpha_j^k = \{0, 1, \cdots, H\}$.

- **Step 5.** Apply the algorithms described in Section 4.3.3 to build the DDSPCE-based model (4.4).

  5a) Truncate the index α_j^k using (3.28) and construct the corresponding multivariate polynomial base of degree H to build the matrix $\boldsymbol{\Psi}$ in (2.11).

  5b) Apply the OMP algorithm to find the optimal sparse polynomial base and compute the PCE coefficients c_k by (2.13).

5c) Compute the leave-one-out error e_{loo} according to (3.31). Go to **Step 6**.

- **Step 6.** If the DDSPCE-based model has reached the prescribed accuracy (e.g., $e_{\text{loo}} < e_{stop}$), go to **Step 7**. Otherwise, enlarge the size of training set by ΔM_p (e.g. $\boldsymbol{P_{w,\Delta p}}$) and calculate $\boldsymbol{Y_{\Delta p}}$ by solving (4.2) and (4.3), then let $M_p \leftarrow M_p + \Delta p$, $\boldsymbol{P_{w_p}} \leftarrow (\boldsymbol{P_{w_p}}, \boldsymbol{P_{w_p,\Delta p}}), \boldsymbol{Y_p} \leftarrow (\boldsymbol{Y_p}, \boldsymbol{Y_{\Delta p}})$ and go back to **Step 3**.

- **Step 7.** After constructing the PCE model, obtain M_s sample points of $\boldsymbol{P_w}$ from raw data. Use the PCA technique to convert these samples into $\boldsymbol{\zeta_s} = \left(\boldsymbol{\zeta}^{(1)}, \cdots, \boldsymbol{\zeta}^{(M_s)}\right)$. Then, determine the PTTC values $\boldsymbol{Y_s} = \left(Y^{(1)}, \cdots, Y^{(M_s)}\right)$ utilizing the defined PCE model (4.4).

- **Step 8.** Determine the statistical properties of the minimum cost Q, such as the mean, standard deviation, PDF, and CDF.

4.5 Case Study - The Integrated Electricity and Gas System

4.5.1 Simulation Setup

This section evaluates the performance of the proposed DDSPCE method on an integrated electricity and gas system (IEGS), specifically the IEEE 118-bus system combined with a 20-node gas system. To incorporate wind power generation, 5 wind farms are integrated into the system located at buses $\{2, 33, 51, 81, 108\}$. The wind power data utilized in the analysis is obtained from the NREL's Western Wind Data Set [212]. The simulation is conducted over a 24-hour period, encompassing time periods $T = 1, \cdots, 24$. Consequently, the wind generator output $\boldsymbol{P_w}$ constitutes a random vector of dimension 120, comprising the wind power data for each time period and wind farm. For detailed information regarding the configuration of the IEGS, including the power and gas network data, wind power data, and load profile, please refer to: `https://github.com/TxiaoWang/DDSPCE-based-Stochastic-ED.git`.

4.5.2 The Probabilistic Characteristics of the Minimum Cost Q

To evaluate the performance of the proposed DDSPCE-based surrogate model, this chapter compares the probabilistic characteristics of the minimum cost Q estimated from the DDSPCE-based surrogate model with those obtained using the SPCE method [93] and the benchmark $10,000$-sample MCS. Specially, 1100 training samples (i.e., $M_p = 1100$) are used to build the DDSPCE model, and the PCE order H is set as 2. We compare the mean μ, standard deviation σ, PDF, and CDF of Q. It can be clearly observed from Table 4.1 and Fig. 4.2 that the DDSPCE and the SPCE methods can provide good estimations for the probabilistic characteristics (including mean, variance, PDF, and CDF) of the minimum cost Q.

Table 4.1 Comparison of the estimated statistics of the minimum cost Q by the MCS, the DDSPCE, and the SPCE methods with $M_p = 1100$.

Index	MCS	DDSPCE	SPCE
μ	7.3276×10^6	7.3276×10^6	7.3277×10^6
σ	4.7069×10^4	4.7159×10^4	4.8600×10^4
$\frac{\Delta\mu}{\mu_{\mathrm{MC}}}\%$	$--$	$\mathbf{2.0739 \times 10^{-5}}$	1.8503×10^{-3}
$\frac{\Delta\sigma}{\sigma_{\mathrm{MC}}}\%$	$--$	$\mathbf{1.9218 \times 10^{-1}}$	3.2529

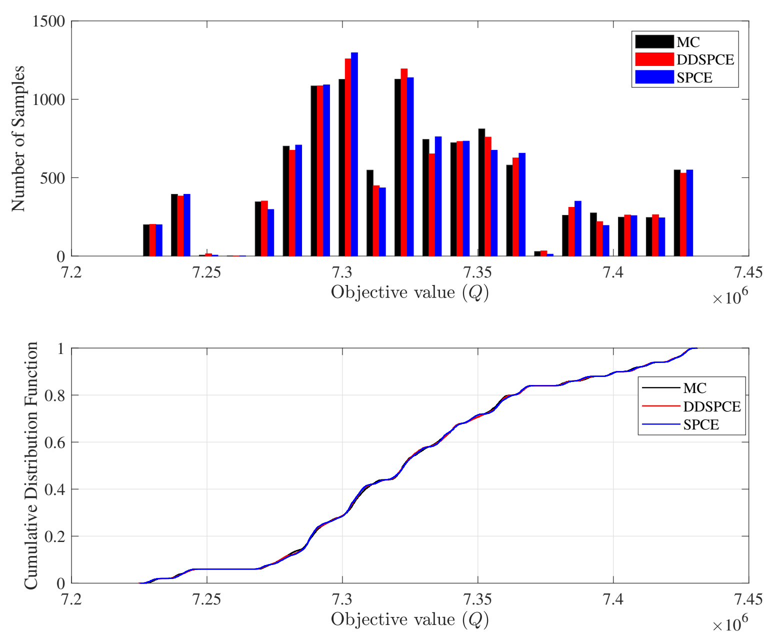

Fig. 4.2 Comparison of the PDF and CDF of the minimum cost Q by the MCS, the DDSPCE, and the SPCE methods with $M_p = 1100$.

Remark 4.5.1. It is worth noting that the SPCE-based model in this simulation is constructed based on the PDF inferred from the available data. In contrast, the proposed DDSPCE-based surrogate model is built directly from the data without relying on prior assumptions or inferred PDFs. This allows us to assess the performance of the DDSPCE-based surrogate model in capturing the probabilistic characteristics of Q accurately and efficiently.

4.5.3 Efficiency Test

The DDSPCE-based surrogate model (4.4) is constructed using a relatively small number of samples, specifically $1,100$ samples for training ($M_p = 1100$). The advantage of the DDSPCE method is its significantly reduced computational time compared to MCS. In fact, the time consumption of the DDSPCE method is only about $\frac{1}{9}$ of the time required by MCS.

However, suppose only the mean and variance of the minimum cost are of interest, and the detailed PDF and CDF are unnecessary. In that case, accurate estimations can be achieved with as few as 300 samples ($M_p \approx 2.5\mathcal{M}$), which represents approximately $\frac{1}{33}$ of the time needed for MCS. Table 4.2 shows the estimated statistics of the minimum cost Q from the proposed DDSPCE, the SPCE, and the MCS methods with $M_p = 300$. The detailed computational time comparison can be seen in Table 4.3. This demonstrates the efficiency of the DDSPCE-based surrogate model in providing accurate results with a significantly reduced computational burden. In comparison, the SPCE method requires additional time for the PDF inferring procedure, resulting in a total time consumption of $8.83s$ more than the proposed DDSPCE method. Overall, the DDSPCE method offers a computationally efficient solution for estimating the mean, variance, and detailed PDF of the minimum cost, providing a significant improvement over traditional MCS and even outperforming the SPCE method in terms of computational efficiency.

Table 4.2 Comparison of the estimated statistics of Q by the MC simulations, the DDSPCE, and the SPCE methods with $\mathcal{N} = 300$.

Index	MC	DDSPCE	SPCE
μ	7.3276×10^6	7.3276×10^6	7.3277×10^6
σ	4.7069×10^4	4.7228×10^4	5.0379×10^4
$\frac{\Delta\mu}{\mu_{\mathrm{MC}}}\%$	$--$	$\mathbf{4.8945 \times 10^{-4}}$	1.7092×10^{-3}
$\frac{\Delta\sigma}{\sigma_{\mathrm{MC}}}\%$	$--$	$\mathbf{3.3939 \times 10^{-1}}$	7.0318

Remark 4.5.2. Based on my simulation experience, it can be observed that a sample size of approximately $2.5\mathcal{M}$ is generally sufficient to achieve accurate estimations of the mean and variance. However, when it comes to accurately estimating the PDF of a *unimodal* system response, a larger

Table 4.3 Comparison of Computational Time by the MC simulations, the DDSPCE, and the SPCE with $M_p = 300$ for the SED problem

Method	$t_{\text{ed}}(s)$	$t_{\text{sc}}(s)$	$t_{\text{es}}(s)$	$t_{\text{total}}(s)$
MC	–	–	5280	5280
DDSPCE	158.88	1.06	0.33	160.27
SPCE	158.88	9.91	0.31	169.10

* t_{ed}: time for evaluating the training samples; t_{sc}: time for constructing the PCE models; t_{es}: time for evaluating Q of $10,000$ samples.

sample size of around $5\mathcal{M}$ is typically required. For *multimodal* system responses, even more samples are needed to ensure accurate PDF estimation, with approximately $9\mathcal{M}$ samples being necessary for the context of this study. These findings highlight the importance of adapting the sample size based on the specific characteristics of the system under investigation.

4.6 Conclusions

This chapter applies the DDSPCE-based surrogate model for the SED problem. The DDSPCE-based surrogate model allows for the approximation of the probability characteristics of the SED solution, such as the mean, variance, PDF, and CDF, without assuming a pre-defined probability distribution for the random inputs. Simulation results on an Integrated Electricity and Gas System (IEGS), specifically the IEEE 118-bus system integrated with a 20-node gas system, validate the effectiveness of the proposed DDSPCE method in accurately and efficiently estimating the probabilistic properties of the SED solution. Notably, the proposed method demonstrates its capability to handle multimodal probability distributions of the SED solution, further highlighting its accuracy and efficiency in handling more generalized scenarios compared to previous methods such as the one described in Chapter 3 [213].

Moreover, this chapter introduces the use of OMP as a coefficient-finding technique within the DDSPCE method. This approach offers faster convergence compared to the method presented in Chapter 3 [213]. Additionally, unlike the situation in Chapter 3 [213] where both the random inputs and the model response (Total Transfer Capability) were assumed to be unimodal, this chapter considers multimodal random inputs derived from real-world data and the model response, which is the minimum cost Q. The successful handling of multimodal situations further demonstrates the accuracy and efficiency of the DDSPCE method in more general contexts.

Chapter 5

The PCE-based Global Sensitivity Analysis for Uncertainty Control in Power System Static Security and Economic Dispatch

Chapter 3 and Chapter 4 show the impacts of volatile RESs on the transfer capability and the ED cost (e.g., increasing penetration of RESs may result in a reduction of ATC). This chapter focuses on measuring the importance of the system's uncertain inputs to the system output. This chapter aims to design effective uncertainty control measures to mitigate these impacts (e.g., enhance the transfer capability). To achieve this goal, the PCE theory in Chapter 2 is exploited and extended, where two PCE-based methods are applied to global sensitivity analysis (GSA) for correlated random inputs in two power system applications. The ideas and results of this chapter are based on the submitted work [J3].

5.1 Introduction

Modern power systems are subjected to various types of random sources. Specially, the uncertainty level of power systems has increased with the growing penetration of volatile renewable generations (e.g., wind power and solar PV power) due to their output intermittency and variability. Therefore, sensitivity analysis of uncertainties as discussed in Section 1.2.3 has become a vital tool for designing effective uncertainty control measures in power systems.

Sensitivity analysis of uncertainties includes three fundamental steps, i.e., uncertainty modeling, uncertainty propagation, and sensitivity analysis. In the previous two chapters [J1, J2], novel data-driven PCE methods have been proposed for UQ in the probabilistic transfer capability assessment and the ED problem. This chapter focuses on sensitivity analysis, specially, the GSA, for power systems, which aims to quantify the importance of an individual or a group of random variables on the system model responses (e.g., voltage variation, probabilistic load margin, PTTC, objective values of ED).

The GSA methods mainly contain the Morris method [214], Sobol' indices [215], ANCOVA indices [216] and Kucherenko indices [217]. Among them, variance-based sensitivity analysis methods, particularly, the Sobol' indices are widely adopted for sensitivity analysis, which are usually conducted through the MC simulations, yet these methods are impractical due to their high computational cost. Spectral methods, as referenced in [218,219], have been employed for GSA. However, these methods might be suitable only for low-dimensional problems, such as the Fourier amplitude sensitivity test [218], or for independent random inputs, like the random balance design [219].

To tackle these issues, surrogate model-based sensitivity analysis has been explored to measure the importance of an individual or a group of random variables on the system model responses efficiently to design effective controls. Ye et al. [129] proposed a kriging-based model to calculate Sobol' indices and to quantify the impact of independent stochastic power injections on voltage variations, which does not consider the correlation between random inputs was not considered. Another popular surrogate model, PCE, has been widely applied to cope with uncertainties in power systems for response estimations, e.g., Muhlpfordt et al. [88] have solved the stochastic optimal power flow using the PCE method, and Xu et al. [38] proposed a data-driven nonparametric method, which combined the Bayesian inference and PCE method, to assess the probabilistic load margin. In literature, many works tended to build a PCE model from independent random inputs (e.g., [213, 220]) in order to obtain closed forms of mean and variance of model response, and Sobol' indices. Liu et al. [221] (in the context of structural reliability) applied popular decorrelation techniques, e.g., the Nataf transform or the Rosenblatt transform, in constructing PCE-based models for global sensitivity indices of correlated random variables. Particularly, Ni et al. developed a model based on PCE in [131] to calculate global sensitivity indices for correlated random inputs in power flow solutions, but without discussing the PCE-based model was built with or without decorrelation, even though both approaches might lead to errors in the derived GSA.

Yet, limited investigations or comparisons regarding the use of PCE-based models in estimating global sensitivity indices for dependent random inputs in power systems, which nevertheless are crucial for effective uncertainty control as discussed in Section 1.2.3 (Chapter 1). In previous studies, Mara et al. [222] derived two sensitivity indices using ANOVA (ANalysis of VAriance) techniques for dependent inputs. However, these methods may not be practical for high-dimensional problems. Two alternative PCE-based methods have been proposed to address correlated random inputs in GSA using ANCOVA (ANalysis of COVAriance). The first method constructs a PCE model from independent random inputs, which ensures the orthogonality of polynomial bases and the convergence of the PCE model in the L_2 norm [221,223]. This method addresses correlated random inputs through decorrelations, employing techniques like the Nataf transform or the Rosenblatt transform (e.g., within civil engineering contexts [223]). However, such decorrelation may lead to inaccuracies in estimating ANCOVA indices. The second method, as indicated in [151], overlooks dependencies during the PCE model construction, which might offer satisfactory accuracy, though a theoretical justification was not presented.

This chapter compares three different PCE-based models (without decorrelation and with decorrelation using different nonlinear transforms) in ANCOVA indices-based GSA for uncertainty control in power system security and ED considering correlated inputs. The contributions of this chapter are as follows.

- This chapter compares and demonstrates the accuracy of ANCOVA-based global sensitivity indices estimated from three different PCE-based models in two power system applications.
- The proposed methods are time-inexpensive compared with the conventional MC simulations for GSA in power systems.
- Simulation results demonstrate that the PCE model, constructed by directly employing correlated random inputs (neglecting the dependency), produces the most accurate ANCOVA indices for both applications.
- Leveraging the acquired sensitivity information, efficacious mitigation strategies are devised to minimize the variance of system responses and enhance system efficacy.

The remainder of this chapter is organized as follows. Section 5.2 introduces the covariance-based global sensitivity indices for dependent random inputs. The PCE method and PCE-based ANCOVA indices calculation are elaborated in Section 5.3. Applications of the proposed PCE-based sensitivity analysis methods in ATC enhancement and ED problem are provided in Section 5.4 and Section 5.5, respectively. Conclusions are discussed in Section 5.7.

5.2 The GSA for Dependent Random Inputs

This section focuses on GSA for power systems with correlated random inputs (e.g., spatial or temporal correlations may exist among different wind speeds). Based on this, the dominant random inputs that affect the variance of model response most significantly will be identified. To this end, a covariance-based method referred to as ANCOVA is introduced to decompose the variance of stochastic response as a summation of contributions of each random input. ANCOVA is generalized from Sobol' indices by Xu et al. [224] and Li et al. [216] for correlated inputs. The Sobol' indices are widely used in global sensitivity analysis for independent random inputs, which allow representing the model response of the sum of contributions of each random input in an interpretable and understandable way [215, 225]. The Sobol' decomposition is derived via a High Dimensional Model Representation (HDMR) of the stochastic model, first formulated in [215] only for independent input variables. To tackle this issue, Xu et al. [224] extended the Sobol' decomposition to a more generic case considering dependent random inputs. To clearly present the relationship between correlated inputs with the stochastic response, the HDMR decomposition for a generic power system model is provided in Section 5.2.1 and the corresponding ANCOVA-based global sensitivity indices are derived in Section 5.2.2.

5.2.1 The HDMR Decomposition

Consider a stochastic response model $Y = G(\boldsymbol{\zeta})$ with an input vector $\boldsymbol{\zeta} = \{\zeta_1, \cdots, \zeta_{\mathcal{M}}\} \in \mathbb{R}^{\mathcal{M}}$, where $\boldsymbol{\zeta}$ could be volatile renewables (e.g., wind speed and solar radiations), load variations, unexpected line outages, etc. Assuming input variables $\boldsymbol{\zeta}$ and the response Y are equipped with finite second-order moments, i.e., $\mathbb{E}[\zeta_j^2] < +\infty$ for $j \in \{1, \cdots, \mathcal{M}\}$ and $\mathbb{E}[Y^2] < +\infty$. No assumptions are required for the dependence structure of the inputs. Based on the original idea of Sobol' decomposition, the stochastic model $Y = G(\boldsymbol{\zeta})$ with correlated input variables can be decomposed via HDMR [216, 224]:

$$\begin{aligned} Y = G(\boldsymbol{\zeta}) = G_0 + \sum_{1 \le j \le \mathcal{M}} G_j(\zeta_j) + \sum_{1 \le j < d \le \mathcal{M}} G_{j,d}(\zeta_j, \zeta_d) + \cdots \\ + \sum_{1 \le j_1 < \cdots < j_m \le \mathcal{M}} G_{j_1, \cdots, j_m}(\zeta_{j_1}, \cdots, \zeta_{j_m}) + \cdots G_{1, \cdots, \mathcal{M}}(\zeta_1, \cdots, \zeta_{\mathcal{M}}) \end{aligned} \tag{5.1}$$

where G_0 is a constant, generally, the mean value of Y, i.e., $\mathbb{E}[Y]$; $G_j(\zeta_j)$ represents the effect of a single random input ζ_j on Y; $G_{j,d}(\zeta_j, \zeta_d)$ describes the effect of the combined random inputs ζ_j and

ζ_d to Y, etc. For a simple example, $Y = \zeta_1 + \zeta_1\zeta_2 + \zeta_2 + \zeta_2^2$, the functions representing the effects of each input variable on Y are: $G_1(\zeta_1) = \zeta_1$, $G_2(\zeta_2) = \zeta_2 + \zeta_2^2$ and the function representing the interaction of random inputs is $G_{1,2}(\zeta_1, \zeta_2) = \zeta_1\zeta_2$.

Let $\mathcal{A} = \{1, \cdots, \mathcal{M}\}$ and $\boldsymbol{\beta}$ be non empty subsets of $\mathcal{A}$, then (5.1) can be rewritten as

$$Y = G_0 + \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}}) \tag{5.2}$$

where $G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})$ represent the combined contribution of random inputs $\boldsymbol{\zeta}_{\boldsymbol{\beta}}$ on Y.

Remark 5.2.1. Note that when random inputs $\boldsymbol{\zeta}$ are independent, the decomposition in (5.1) is unique if G_0 is a constant and the integral of each component function $G_{j_1,\cdots,j_m}$ with respect to any of their own variables are zero [215], i.e., $\int_{\Omega_k} G_{j_1,\cdots,j_m}(\zeta_{j_1}, \cdots, \zeta_{j_m})d\zeta_{j_k} = 0$, with $1 \leq j_1 < \cdots < j_m \leq \mathcal{M}$ and $j_k \in \{j_1, \cdots, j_m\}$.

5.2.2 The ANCOVA-based Global Sensitivity Indices

Based on the model decomposition in (5.2), let us write the variance of model response Y as:

$$\begin{aligned}\mathrm{Var}[Y] &= \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right] = \mathbb{E}\left[(Y - G_0)\left(G_0 + \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}}) - G_0\right)\right] \\ &= \mathrm{Cov}\left[Y, G_0 + \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right] = \mathrm{Cov}\left[Y, \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right]\end{aligned} \tag{5.3}$$

Note that the second to the last equality is based on the definition of covariance: $\mathrm{Cov}[X_1, X_2] := \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])]$. According to the distributive property of covariance, (5.3) can be further derived as:

$$\mathrm{Var}[Y] = \mathrm{Cov}\left[Y, \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right] = \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} \mathrm{Cov}\left[Y, G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right] \tag{5.4}$$

Substituting the expression of Y in (5.2) to (5.4), we have:

$$\begin{aligned}
\mathrm{Var}[Y] &= \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} \mathrm{Cov}\left[G_0 + \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}}), G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right] \\
&= \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} \mathrm{Cov}\left[G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}}) + \sum_{v\subseteq\mathcal{A},\boldsymbol{v}\neq\emptyset,\boldsymbol{v}\neq\boldsymbol{\beta}} G_{\boldsymbol{v}}(\boldsymbol{\zeta}_{\boldsymbol{v}}), G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right] \\
&= \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset} \mathrm{Var}\left[G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right] + \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset}\left[\mathrm{Cov}\left[G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}}), \sum_{\boldsymbol{v}\subseteq\mathcal{A},\boldsymbol{v}\neq\emptyset,\boldsymbol{v}\neq\boldsymbol{\beta}} G_v(\boldsymbol{\zeta}_{\boldsymbol{v}})\right]\right] \\
&= \sum_{\boldsymbol{\beta}\subseteq\mathcal{A},\boldsymbol{\beta}\neq\emptyset}\left[\mathrm{Var}\left[G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}})\right] + \mathrm{Cov}\left[G_{\boldsymbol{\beta}}(\boldsymbol{\zeta}_{\boldsymbol{\beta}}), \sum_{\boldsymbol{v}\subseteq\mathcal{A},v\neq\emptyset,\boldsymbol{v}\neq\boldsymbol{\beta}} G_{\boldsymbol{v}}(\boldsymbol{\zeta}_{\boldsymbol{v}})\right]\right]
\end{aligned} \tag{5.5}$$

The above technique to decompose $\mathrm{Var}[Y]$ is referred to as the ANCOVA decomposition, also named Structural and Correlated Sensitivity Analysis [216, 220, 224]. The decomposition in (5.4) indicates that the variance of Y can be represented by the summation of the covariance between $G_{\boldsymbol{\beta}}(\boldsymbol{\beta}_{\boldsymbol{\beta}})$ and the response Y for all $\boldsymbol{\beta} \subseteq \mathcal{A}, \boldsymbol{\beta} \neq \emptyset$. Therefore, according to the above decomposition, the importance of ζ_j to Y is defined as the covariance between $G_j(\zeta_j)$ and Y and the ANCOVA index for an individual random input $\zeta_j, j \in \boldsymbol{\beta}$ can be defined as $S_j = \frac{\mathrm{Cov}[Y, G_j(Z_j)]}{\mathrm{Var[Y]}}$, which depicts the total effect of ζ_j on the variation of response Y. Due to the correlation between random input samples, the covariance between different terms does not equal zero (i.e., $\mathrm{Cov}[G_j(\zeta_j), G_{\boldsymbol{v}}(\boldsymbol{\zeta}_{\boldsymbol{v}})] \neq 0$ for $\boldsymbol{v} \subseteq \mathcal{A}.\boldsymbol{v} \neq j, \boldsymbol{v} \neq \emptyset$). Using (5.5), S_j can be further divided into the uncorrelated and correlated effects of ζ_j on $\mathrm{Var}[Y]$.

$$S_j = \frac{\mathrm{Cov}\left[Y, G_j(\zeta_j)\right]}{\mathrm{Var[Y]}} = S_j^{(\mathrm{U})} + S_j^{(\mathrm{C})} \tag{5.6}$$

with

$$\begin{aligned}
S_j^{(\mathrm{U})} &= \frac{\mathrm{Var}[G_j(\zeta_j)]}{\mathrm{Var}[Y]} \\
S_j^{(\mathrm{C})} &= \frac{\mathrm{Cov}\left[G_j(\zeta_j), \sum_{\substack{\boldsymbol{v}\subseteq\mathcal{A} \\ \boldsymbol{v}\neq\{j\}}} G_v(\boldsymbol{\zeta}_{\boldsymbol{v}})\right]}{\mathrm{Var}[Y]}
\end{aligned} \tag{5.7}$$

where $S_j^{(\mathrm{U})}$ is the uncorrelated effect of ζ_j on $\mathrm{Var}[Y]$ (i.e., the physical role of ζ_j on the model response variance $\mathrm{Var}[Y]$); $S_j^{(\mathrm{C})}$ is the correlated effect of ζ_j to $\mathrm{Var}[Y]$ (i.e., the effect of the correlation between ζ_j and other $\zeta_{\boldsymbol{v}}$ on $\mathrm{Var}[Y]$). For instance, $\zeta_j, j = 1, 2$ represent the wind speeds at two distinct wind farm locations. If the objective is to analyze the influence of variations in these wind speeds on the model response variance Y, such as the TTC, then $S_1^{(U)}$ delineates the impact of ζ_1 alone on TTC Y. Conversely, $S_1^{(C)}$ indicates the effect of the correlation between the two wind speeds on Y. The metric S_1 encapsulates the cumulative influence of the variation of wind speed ζ_1 on Y. Consequently, any ζ_j with the largest S_j value is deemed as the dominant random input. Reducing the variance of these dominant inputs (by smoothing them) can most effectively reduce $\mathrm{Var}[Y]$. Specially, $S^{(C)}$ can be further interpreted by two parts:

$$S_j^{(\mathrm{C})} = \frac{\mathrm{Cov}\left[G_j(\zeta_j), \sum_{\substack{\boldsymbol{w}\subseteq\mathcal{A} \\ \{j\}\in\boldsymbol{w}}} G_{\boldsymbol{w}}(\boldsymbol{\zeta}_{\boldsymbol{w}})\right]}{\mathrm{Var}[Y]} + \frac{\mathrm{Cov}\left[G_j(\zeta_j), \sum_{\substack{\boldsymbol{u}\subseteq\mathcal{A} \\ \{j\}\notin\boldsymbol{u}}} G_{\boldsymbol{u}}(\boldsymbol{\zeta}_{\boldsymbol{u}})\right]}{\mathrm{Var}[Y]} \tag{5.8}$$

where $G_{\boldsymbol{w}}$ depends on ζ_j and other input variables in $\mathcal{A}$ and the first part of (5.8) indicates the interactive and correlative effect of ζ_j with other input variables. $G_{\boldsymbol{u}}$ depends on inputs in $\mathcal{A}$ excluding ζ_j, and the second part of (5.8) indicates only the correlative effect of ζ_j with other inputs.

S_j is used to characterize the importance of Z_j on $\mathrm{Var}[Y]$, which can be positive, negative, or zero. Random inputs ζ_j with the largest S_j are considered dominant because smoothing them out (i.e., reducing their variance to zero) can most effectively reduce $\mathrm{Var}[Y]$. A high S_j might arise from the structural contributions of ζ_j to $\mathrm{Var}[Y]$ or its significant correlations with other variables. By definition, $S_j^{(U)} \geq 0$ for all j. Correlations between random inputs can either increase, decrease, or not affect a variable's contribution. Consequently, $S_j^{(\mathrm{C})}$ can be positive, negative, or zero. A notably small $|S_j^{(\mathrm{C})}|$ suggests $S_j^{(U)} \approx S_j$, indicating a minimal impact of the correlation on ζ_j's contribution to $\mathrm{Var}[Y]$. Conversely, a significantly high $|S_j^{(\mathrm{C})}|$, implying $S_j^{(C)} \approx S_j$, denotes a substantial correlation impact on the contribution of ζ_j to $\mathrm{Var}[Y]$ [138].

The essence of designing an effective uncertainty control measure lies in accurate estimations of S_j. Once S_j are calculated and ranked, the critical random inputs ζ_j can be identified with the largest S_j values. Based on this, the control measure is designed to reduce the variance most effectively by smoothing out the critical random inputs ζ_j.

In order to calculate the ANCOVA indices S_j, MC simulations, the most simple and widely used method, can be carried out to identify Y and the terms $G_j(\zeta_j), G_{j,d}(\zeta_j, \zeta_d), \cdots$ in (5.1) as discussed in [220]. However, a massive number of MC samples are required to guarantee the accurate estimations of Y and the terms (5.1), which is computationally expensive, leading to its impractical applications. To address this issue, we introduce a PCE-based surrogate model to calculate the ANCOVA indices efficiently in Section 5.3.

5.3 The Proposed PCE-based Global Sensitivity Analysis Method

In this section, we will introduce two different methods for constructing the PCE-based models as surrogate models of $Y = G(\boldsymbol{\zeta})$. Based on the built PCE models, the ANCOVA indices can be estimated efficiently. The PCE-based ANCOVA indices will be elaborated in Section 5.3.1.

5.3.1 The PCE-based ANCOVA indices

Given a random input vector $\boldsymbol{\zeta}$and the stochastic response $Y = G(\boldsymbol{\zeta})$ with finite second-order moments can be approximated by PCE (2.6) as introduced in Chapter 2 [78]. Let $\boldsymbol{\alpha} = \{\alpha_k^1, \cdots, \alpha_k^{\mathcal{M}}\}$ be the multi-indices for the multivariate orthogonal polynomials basis $\Psi_{\boldsymbol{k}}$. Using the multi-index notation, the PCE-based model (2.6) can be rewritten as:

$$Y = G(\boldsymbol{\zeta}) \approx G^{\mathrm{pc}}(\boldsymbol{\zeta}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^{\mathcal{M}}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) = \boldsymbol{C}^T \boldsymbol{\Psi}(\boldsymbol{\zeta}) \tag{5.9}$$

where c_0 is the constant term with the multi-index $\boldsymbol{\alpha} = \{\alpha_0^1 = 0, \cdots, \alpha_0^{\mathcal{M}} = 0\}$. $c_{\boldsymbol{\alpha}} =$ are the L unknown coefficients to be determined. $\boldsymbol{C} = \{c_{\boldsymbol{\alpha}_0}, \cdots, c_{\boldsymbol{\alpha}_{L-1}}\}^T$ and the polynomials $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) = \{\Psi_{\boldsymbol{\alpha}_0}(\boldsymbol{\zeta}), \cdots, \Psi_{\boldsymbol{\alpha}_{L-1}}(\boldsymbol{\zeta})\}^T$. Particularly, $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ required to satisfy the orthogonal condition that $\int_{\Omega} \Psi_k(\boldsymbol{\zeta}) \Psi_m(\boldsymbol{\zeta}) f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) = 0, k \neq m$, with Ω being the support of $\boldsymbol{\zeta}$ and $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ the joint PDF of $\boldsymbol{\zeta}$, such that the $\mathcal{L}^2$ convergence (2.5) is satisfied.

Similarly, write the PCE-based model (5.9) in terms of the HDMR (5.1), and it can be obtained that:

$$\begin{aligned} \hat{Y} = G^{\mathrm{pc}}(\boldsymbol{\zeta}) = G_0^{\mathrm{pc}} + \sum_{1 \le j \le \mathcal{M}} G_j^{\mathrm{pc}}(\zeta_j) + \sum_{1 \le j < d \le \mathcal{M}} G_{j,d}^{\mathrm{pc}}(\zeta_j, \zeta_d) \\ + \cdots + G_{1,\cdots,\mathcal{M}}^{\mathrm{pc}}(\zeta_1, \cdots, \zeta_{\mathcal{M}}) \end{aligned} \tag{5.10}$$

where G_0^{pc} is the constant term. Substituting (2.7) into (5.9), and comparing (5.9) and (5.10), we have

$$
\begin{aligned}
G_0^{\mathrm{pc}} &= c_0 \\
G_j^{\mathrm{pc}}(\zeta_j) &= \sum_{\alpha_k^j=1}^{H} c_{\alpha_k^j} \Psi_{\alpha_k^j}(\zeta_j) = \sum_{\alpha_k^j=1}^{H} c_{\alpha_k^j} \phi_j^{(\alpha_k^j)}(\zeta_j) \\
G_{j,d}^{\mathrm{pc}}(\zeta_j,\zeta_d) &= \sum_{\alpha_k^j=1}^{H} \sum_{\alpha_k^d=1}^{H} c_{\boldsymbol{\alpha}_k^{j,d}} \Psi_{\boldsymbol{\alpha}_k^{j,d}}(\zeta_j,\zeta_d) = \sum_{\alpha_k^j=1}^{H} \sum_{\alpha_k^d=1}^{H} c_{\boldsymbol{\alpha}_k^{j,d}} \phi_j^{(\alpha_k^j)}(\zeta_j) \phi_d^{(\alpha_k^d)}(\zeta_d) \\
&\cdots
\end{aligned}
\tag{5.11}
$$

where c_0 is the constant term, and H denotes the order of PCE-based models, which can be determined by the stopping criteria in (3.30)-(3.31). $G_j^{\mathrm{pc}}(\zeta_j)$ includes the terms $\Psi_{\alpha_k^j}$ only depend on a single input variable ζ_j and $G_{j,d}^{\mathrm{pc}}(\zeta_j,\zeta_d)$ includes the terms $\Psi_{\boldsymbol{\alpha}_k^{j,d}}$ relying on inputs ζ_j and ζ_d, etc.

According to the decomposition in (5.10) and (5.11), the ANCOVA indices (5.3) can be calculated through the built PCE-based model (5.9). Given M_s samples of $\boldsymbol{\zeta}^{(s)}, s = \{1, \cdots, M_s\}$, then the model response $\hat{Y}^{(s)}$ and each term in (5.10) can be evaluated efficiently from the PCE-based model (5.9), i.e., $\hat{Y}^{(s)} = G^{\mathrm{pc}}(\boldsymbol{\zeta}^{(s)})$, $\widehat{G}_j(\boldsymbol{\zeta}_j^{(s)}) = G_j^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)})$, $\widehat{G}_{j,d}(\boldsymbol{\zeta}_j^{(s)}, \boldsymbol{\zeta}_d^{(s)}) = \widehat{G}_{j,d}^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)}, \boldsymbol{\zeta}_d^{(s)})$, etc. Then the sample mean $\widehat{\mathbb{E}}[Y]$ and the sample variance $\widehat{\mathrm{Var}}[Y]$ can be calculated by:

$$
\begin{aligned}
\widehat{\mathbb{E}}[Y] &= \frac{1}{M_s} \sum_{s=1}^{M_s} G^{\mathrm{pc}}(\boldsymbol{\zeta}^{(s)}) \\
\widehat{\mathrm{Var}}[Y] &= \frac{1}{M_s - 1} \sum_{s=1}^{M_{\mathrm{S}}} \left[G^{\mathrm{pc}}(\boldsymbol{\zeta}^{(s)}) - \widehat{\mathbb{E}}[Y] \right]^2
\end{aligned}
\tag{5.12}
$$

Furthermore, the sample covariance between Y and $G_j(\zeta_j)$ can be estimated by:

$$
\widehat{\mathrm{Cov}}\left[Y, G_j(\zeta_j)\right] = \frac{1}{M_s - 1} \sum_{s=1}^{M_{\mathrm{S}}} \left[\hat{Y}^{(s)} - \widehat{\mathbb{E}}[Y] \right] \left[G_j^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)}) - \widehat{\mathbb{E}}[G_j^{\mathrm{pc}}(\zeta_j)] \right] \tag{5.13}
$$

where the sample mean of $G_j^{\mathrm{pc}}(\zeta_j)$ is:

$$
\widehat{\mathbb{E}}[G_j^{\mathrm{pc}}(\zeta_j)] = \frac{1}{M_s} \sum_{s=1}^{M_s} G^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)}) \tag{5.14}
$$

Lastly, the sample variance of $G_j^{\mathrm{pc}}(\zeta_j)$ is:

$$\widehat{\mathrm{Var}}[G_j^{\mathrm{pc}}(\zeta_j)] = \frac{1}{M_s - 1} \sum_{s=1}^{M_{\mathrm{S}}} \left[G_j^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)}) - \widehat{\mathbb{E}}[G_j^{\mathrm{pc}}(\zeta_j)] \right]^2 \tag{5.15}$$

Then, based on (5.12)-(5.15), the ANCOVA indices in (5.6) are determined:

$$\begin{aligned} S_j &= \frac{\widehat{\mathrm{Cov}}\left[\hat{Y}, G_j^{\mathrm{pc}}(\zeta_j)\right]}{\widehat{\mathrm{Var}}[Y]} = \frac{\frac{1}{M_s-1}\sum_{s=1}^{M_{\mathrm{S}}}\left[\hat{Y}^{(s)} - \widehat{\mathbb{E}}[Y]\right]\left[G_j^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)}) - \widehat{\mathbb{E}}[G_j^{\mathrm{pc}}(\zeta_j)]\right]}{\frac{1}{M_s-1}\sum_{s=1}^{M_{\mathrm{S}}}\left[G^{\mathrm{pc}}(\boldsymbol{\zeta}^{(s)}) - \widehat{\mathbb{E}}[Y]\right]^2} \\ S_j^{(\mathrm{U})} &= \frac{\widehat{\mathrm{Var}}[G_j^{\mathrm{pc}}(\zeta_j)]}{\widehat{\mathrm{Var}}[Y]} = \frac{\frac{1}{M_s-1}\sum_{s=1}^{M_{\mathrm{S}}}\left[G_j^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)}) - \widehat{\mathbb{E}}[G_j^{\mathrm{pc}}(\zeta_j)]\right]^2}{\frac{1}{M_s-1}\sum_{s=1}^{M_{\mathrm{S}}}\left[G^{\mathrm{pc}}(\boldsymbol{\zeta}^{(s)}) - \widehat{\mathbb{E}}[Y]\right]^2} \\ S_j^{(\mathrm{C})} &= S_j - S_j^{(\mathrm{U})} \end{aligned} \tag{5.16}$$

Remark 5.3.1. Note that random inputs $\boldsymbol{\zeta}$ have mutually independent counterparts which is a special case for the ANCOVA index calculation. In this case, $\mathbb{E}[Y]$ and $\mathrm{Var}[Y]$ can be easily calculated using (2.14). Besides, Cov $\left[G_j(\zeta_j), \sum_{\boldsymbol{u}\subseteq\mathcal{A},\{j\}\notin\boldsymbol{u}} G_{\boldsymbol{u}}(\boldsymbol{\zeta}_{\boldsymbol{u}})\right] = 0$, which indicates there is no correlative contribution in the variation of Y, and the ANCOVA index is identical to Sobol' index [215].

Calculation of S_j for Independent Random Inputs $\boldsymbol{\zeta}$

To compute S_j according to (5.16), definitely, the first step is to establish the PCE model (5.9). When the random inputs $\boldsymbol{\zeta}$ are mutually *independent*, the polynomial $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ can be formulated through the tensor product of univariate orthogonal polynomial bases $\phi_j^{(\alpha_k^j)}(\zeta_j)$, which guarantees the orthogonality of $\Psi_k(\boldsymbol{\zeta})$. Specifically, this can be constructed using (2.7) with $\Psi_{\boldsymbol{\alpha}}(\zeta_1, \cdots, \zeta_{\mathcal{M}}) = \prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(\zeta_j)$. The polynomial $\phi_j^{(\alpha_k^j)}(\zeta_j)$ can be initially derived from raw data or a presumed probabilistic model of ζ_j, with α_k^j denoting the corresponding degree of ϕ_j. Notably, this correspondence employs the moment-based method (refer to Section 3.3.2) to construct univariate polynomial bases ϕ_j. Once $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ is established, the coefficients $c_{\boldsymbol{\alpha}}$ can be determined using the regression methods (e.g., the OLS (see Section 2.1.3), the LAR (refer to Section 3.3.3), and the hybrid LAR (see Appendix D)), using M_p number of sample evaluations (M_p is a small number).

Calculation of S_j for Correlated Random Inputs $\boldsymbol{\zeta}$

As discussed above, it is clear that the key point to calculating the ANCOVA index is to build an accurate PCE model to approximate the response Y. If random inputs $\boldsymbol{\zeta}$ are independent, it is straightforward to construct a PCE model. However, it is common that random inputs $\boldsymbol{\zeta}$ (e.g., wind speed and solar radiation) are correlated in practical power system applications [226–228], which poses challenges to the PCE construction for the ANCOVA index calculation. If $\boldsymbol{\zeta}$ are correlated, the multivariate polynomial basis $\Psi_{\alpha}(\boldsymbol{\zeta})$, as stated in [145], cannot be constructed purely through the tensor product of the univariate polynomial basis $\phi_j^{(\alpha_k^j)}(\zeta_j)$ since the joint PDF $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ is not the product of the marginal PDF of ζ_j and the orthogonal condition (2.4) does not hold. In previous work, two methods have been suggested for estimating ANCOVA indices, yet without theoretical proof: 1) building a PCE model by ignoring the input correlations, as shown in [151], a reasonably accurate approximation of Y from the PCE model can be achieved; 2) building a PCE model with decorrelated random inputs as claimed in [220, 229] claimed that a PCE model built with independent random inputs still holds for correlated inputs following the same marginal distributions. Therefore, two alternative PCE-based methods are investigated and compared in this chapter.
i) Method 1: the *first* method, denoted as *PCE_correlate*, as suggested by [151], the PCE model (5.9) is constructed directly through a set of sample-response pairs $[\boldsymbol{\zeta}_p, \boldsymbol{Y}_p]$ by ignoring the input dependencies and building multivariate polynomials basis $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ orthogonal with respect to $\prod_{j=1}^{\mathcal{M}} f_{\zeta_j}(\zeta_j)$, where $f_{\zeta_j}(\zeta_j)$ is the marginal PDF of ζ_j.
ii) Method 2: the *second* method, as suggested by [220, 229], building a PCE model from decorrelated inputs, which includes: 1) decorrelating correlated samples modeling the dependencies of inputs by Gaussian (or vine) copula and then converting correlated samples $\boldsymbol{\zeta}_p$ into independent samples $\boldsymbol{Z}_p$; 2) constructing the PCE model (5.9) using the sample-response pairs $[\boldsymbol{Z}_p, \boldsymbol{Y}_p]$. For convenience, this method applies the Nataf and the Rosenblatt transform for decorrelation, and are denoted by *PCE_NT* and *PCE_RT*, respectively.

Remark 5.3.2. Note that the Nataf transform is employed for $\boldsymbol{\zeta}$ with a Gaussian copula. Conversely, the Rosenblatt transform is utilized when $\boldsymbol{\zeta}$ exhibits a more intricate correlation, such as nonlinear or tail dependence [223]. See Appendix B for the details of the two transforms.

Once the PCE-based model (5.9) is constructed using the above two methods, acquire a correlated sample set $\boldsymbol{\zeta}^{(s)}, s = \{1, \cdots, M_s\}$ with sample size M_s ($M_s \gg M_p$), and evaluate the response $\hat{Y}$ and each term in (5.10) by substituting $\boldsymbol{\zeta}^{(s)}$ into the constructed PCE-based models efficiently. Then the ANCOVA indices S_j for each random input ζ_j can be estimated by (5.16) with the help of

(5.12)-(5.15). Next, rank S_j and identify the dominant inputs which are with the highest S_j values. After that, the control measure is designed by smoothing the dominant random inputs (i.e., reducing the variance of dominant inputs to zero, e.g., by energy storage systems(ESSs)) to reduce $\mathrm{Var}[Y]$ in the most effective way.

Remark 5.3.3. Note that to guarantee the number of samples used for the response Y estimation and ANCOVA indices calculation is sufficient, M_{S} is typically set as a large number (e.g., $M_s = 10,000$).

Procedures for the PCE-based methods in ANCOVA indices estimation and uncertainty control for power systems

The detailed steps of the implementation of the PCE-based methods in ANCOVA indices estimation and uncertainty control for power systems are presented below.

Method 1 *PCE_correlate*-based ANCOVA indices estimation and uncertainty control

Step 1. Input network data, and generate a set of input samples $\boldsymbol{\zeta}_p \in \mathbb{R}^{M_p \times \mathcal{M}}$ samples of $\mathcal{M}$ random inputs $\boldsymbol{\zeta}$ (e.g., wind speeds).

Step 2. Fed input samples $\boldsymbol{\zeta}_p$ into the deterministic power system analysis tools to calculate the system responses $\boldsymbol{Y}_p \in \mathbb{R}^{M_p}$. Pass the sample-response pairs $[\boldsymbol{\zeta}_p, \boldsymbol{Y}_p]$ to **Step 3**.

Step 3. Construct the *PCE_correlate*-based model in (5.9):

a) Build the univariate polynomials $\phi_j^{(\alpha_k^j)}(\zeta_j)$ using the moment-based method (3.23);

b) Construct the multivariate polynomials $\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ through the tensor product of $\phi_j^{(\alpha_k^j)}(\zeta_j)$ using (2.7);

c) Calculate the unknown coefficients $\boldsymbol{c}_{\boldsymbol{\alpha}}$ using hybrid LAR (refer to Appendix D) based on the M_p sample pairs.

Step 4. Derive the terms $G_j^{\mathrm{pc}}(\zeta_j)$, $G_{j,d}^{\mathrm{pc}}(\zeta_j, Z_d)$, ..., in the HDMR (5.11) from the *PCE_correlate* model constructed in **Step 3**.

Step 5. Acquire a large number of M_{S} input samples $\boldsymbol{\zeta}^{(s)}$, evaluate $\hat{Y}^{(s)} = G^{\mathrm{pc}}(\boldsymbol{\zeta}^{(s)})$ based on the constructed PCE model, and $\widehat{\mathbb{E}}[Y]$, $\widehat{\mathrm{Var}}[Y]$, $G_j(\zeta_j)$, $\widehat{\mathbb{E}}[G_j^{\mathrm{pc}}(\zeta_j)]$, $\widehat{\mathrm{Var}}[G_j^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)})]$, and $\widehat{\mathrm{Cov}}\left[Y, G_j^{\mathrm{pc}}(\zeta_j)\right]$ using (5.12)-(5.15).

Step 6. Calculate the ANCOVA indices S_j based on the (5.16). Identify the critical random inputs with the highest S_j values.

Step 7. Uncertainty control: design control measures by smoothing the critical random inputs identified in **Step 6**, i.e., reducing the variance of the critical random inputs to zero.

Method 2 *PCE_NT* or *PCE_RT*-based ANCOVA indices estimation and uncertainty control

1: **Step 1.** Input network data, and generate a set of input samples $\boldsymbol{\zeta}_p \in \mathbb{R}^{M_p \times \mathcal{M}}$ samples of $\mathcal{M}$ random inputs $\boldsymbol{\zeta}$ (e.g., wind speeds);

2: **Step 2.** Fed input samples $\boldsymbol{\zeta}_p$ into the deterministic power system analysis tools to calculate the system responses $\boldsymbol{Y}_p \in \mathbb{R}^{M_p}$. Pass the sample-response pairs $[\boldsymbol{\zeta}_p, \boldsymbol{Y}_p]$ to **Step 3**;

3: **Step 3.** Construct the *PCE_NT* and *PCE_RT*-based models in (5.9):

   a) Decorrelate the input samples $\boldsymbol{\zeta}_p$ to $\boldsymbol{Z}_p$ using the Nataf or Rosenblatt transform; Pass data set $[\boldsymbol{Z}_p, \boldsymbol{Y}_p]$ to **Step 3 b)**;

   b) Build the univariate polynomials $\phi_j^{(\alpha_k^j)}(Z_j)$ using the moment-based method (3.23);

   c) Construct the multivariate polynomials $\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{Z})$ through the tensor product of $\phi_j^{(\alpha_k^j)}(Z_j)$ using (2.7);

   d) Calculate the unknown coefficients $\boldsymbol{c}_{\boldsymbol{\alpha}}$ using hybrid LAR (refer to Appendix D) based on the M_p sample pairs obtained in **Step 3 a)**;

4: **Step 4.** Derive the terms $G_j^{\mathrm{pc}}(\zeta_j)$, $G_{j,d}^{\mathrm{pc}}(\zeta_j, Z_d)$, ..., in the HDMR (5.11) from the *PCE_NT* and *PCE_RT* model constructed in **Step 3**.

5: **Step 5.** Acquire a large number of M_{S} input samples $\boldsymbol{\zeta}^{(s)}$, evaluate $\hat{Y}^{(s)} = G^{\mathrm{pc}}(\boldsymbol{\zeta}^{(s)})$ based on the constructed PCE model, and $\widehat{\mathbb{E}}[Y]$, $\widehat{\mathrm{Var}}[Y]$, $G_j(\zeta_j)$, $\widehat{\mathbb{E}}[G_j^{\mathrm{pc}}(\zeta_j)]$, $\widehat{\mathrm{Var}}[G_j^{\mathrm{pc}}(\boldsymbol{\zeta}_j^{(s)})]$, and $\widehat{\mathrm{Cov}}\left[Y, G_j^{\mathrm{pc}}(\zeta_j)\right]$ using (5.12)-(5.15).

6: **Step 6.** Calculate the ANCOVA indices S_j based on the (5.16). Identify the critical random inputs with the highest S_j values.

7: **Step 7.** Uncertainty control: design control measures by smoothing the critical random inputs identified in **Step 6**, i.e., reducing the variance of the critical random inputs to zero.

Remark 5.3.4. In these two methods, **Step 1,2,4,5,6,7** are the same and the only difference is in **Step 3**, where for **Method 2**, correlated samples $\boldsymbol{\zeta}_p$ are converted to independent samples first before building the PCE model. Besides, the main time consumption of the proposed two PCE-based methods is in **Step 2** for generating the sample-response pairs $[\boldsymbol{\zeta}_p, \boldsymbol{Y}_p]$.

5.4 Case Study I - Available Transfer Capability Enhancement

This section applies the *PCE_correlate* (**Method 1**), *PCE_NT* and *PCE_RT* (**Method 2**) to estimate the ANCOVA indices for the power system ATC enhancement. This application first estimates the ANCOVA indices in the PTTC assessment, aiming to find the critical random inputs ζ_j dominating the variance of PTTC. Next, the control measure is designed by smoothing the dominant random inputs ζ_j (i.e., reduce the variance of ζ_j to zero) to reduce the variance of PTTC in an effective way, thus enhancing ATC, which is further demonstrated by the MC simulations.

5.4.1 System Configurations

In this case, simulations are performed on the modified IEEE 24-bus reliability test system to test the proposed two PCE-based methods. There are six random variables $\zeta_j, j = \{1, \cdots, 6\}$ including three wind speeds and three solar radiations, i.e., $\boldsymbol{\zeta} = [\boldsymbol{v}, \boldsymbol{r}]$ with 3 wind speeds $\boldsymbol{v} = [v_1, v_2, v_3]$ and 3 solar radiations $\boldsymbol{r} = [r_1, r_2, r_3]$. To be specific, there are three wind farms added into bus {1,2,15} and three solar PV plants added into bus {16,18,21}. The response Y considered in this case is the PTTC defined from generators at bus 7 to loads at bus {3,4,9}.

In this case, the probabilistic data of random inputs are generated from assumed probabilistic distribution (see Appendix C) and detailed random parameter configurations can be found in Appendix F. Note that only the data information is utilized in *PCE_correlate*, while the information of the probability distributions is also used in *PCE_NT* and *PCE_RT* for decorrelation.

5.4.2 The PCE-based Models Accuracy Test

Firstly, sixty sample response pairs ($M_p = 60$) are generated ($[\boldsymbol{\zeta}_p, \boldsymbol{Y}p]$ for **Method 1** and $[\boldsymbol{Z}p, \boldsymbol{Y}_p]$ for **Method 2**) to construct the three PCE-based models, as detailed in **Steps 1-3** of both **Method 1** and **Method 2**. The order H for these models is selected as 2 according to (3.30)-(3.31). Then, $M_s = 10,000$ samples of correlated random inputs are evaluated using the three PCE-based models, as described in **Step 5**. The estimated statistics from these three PCE-based models are then compared with the benchmark results from LHS-based MC simulations, and the findings are summarized in Table 5.1. As illustrated in Fig. 5.1 and Fig. 5.2, the *PCE_correlate* model (depicted in blue) provides results that align most closely with the benchmark LHS-based MC simulations (shown in black), in the estimation of the PDF and CDF of PTTC.

Table 5.1 Comparisons of the estimated statistics of the overall TTC by the MC simulations, *PCE_correlate*, *PCE_NT* and *PCE_RT*

Index	MC	*PCE_correlate*	*PCE_NT*	*PCE_RT*
$\mathbb{E}[\mathrm{PTTC}]$	283.2862	283.2451	283.1519	283.1759
$\hat{\sigma}[\mathrm{PTTC}]$	0.9702	0.9494	1.1192	1.1336
$\frac{\Delta\mathbb{E}[\mathrm{PTTC}]}{\mathbb{E}[Y]}\%$	–	-1.4503×10^{-2}	-4.7417×10^{-2}	-3.8930×10^{-2}
$\frac{\hat{\sigma}[\mathrm{PTTC}]}{\hat{\sigma}[Y]}\%$	–	-2.1434	15.3560	16.8375

* $\hat{\sigma}[\cdot]$ denotes the estimated standard deviation.

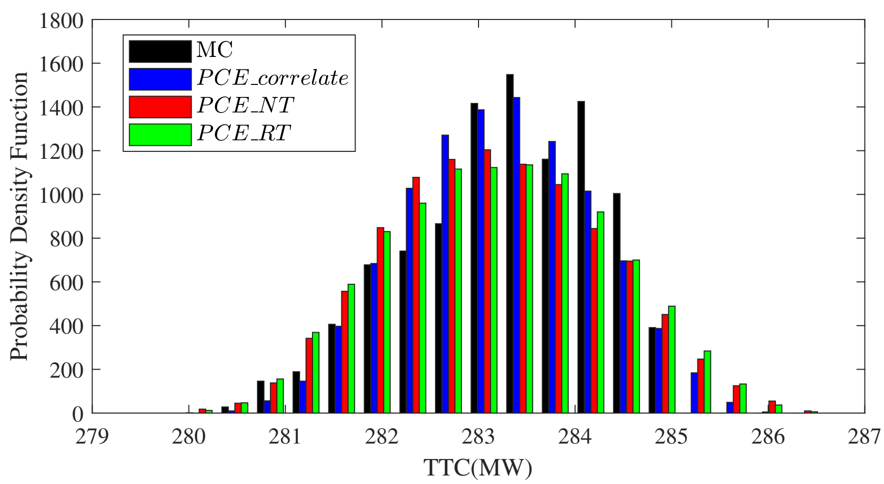

Fig. 5.1 The PDFs of PTTC from *PCE_correlate*, *PCE_NT*, *PCE_RT* and the MC simulations.

5.4.3 The ANCOVA Indices Estimation

Subsequently, the ANCOVA indices S_j are calculated in **Step 6**, and the results and presented in Table 5.2. It can be seen that the three PCE-based models show different ANCOVA estimations. S_j estimated from *PCE_correlate* shows that ζ_4, ζ_5 and ζ_6 share similar contributions on the variance

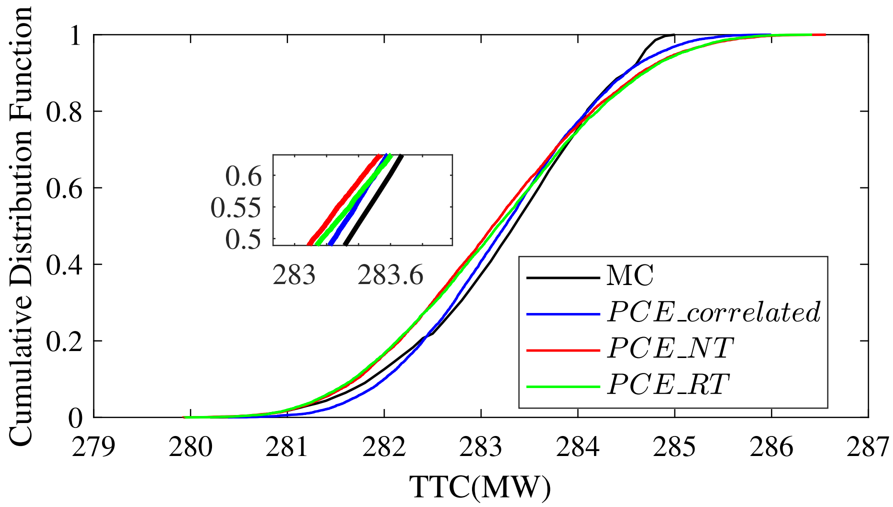

Fig. 5.2 The CDFs of PTTC from *PCE_correlate*, *PCE_NT*, *PCE_RT* and the MC simulations.

of PTTC, which gives the first three dominant random inputs. ζ_4 and ζ_6 are the dominant inputs according to S_j estimated from *PCE_NT* and *PCE_RT*, respectively.

Table 5.2 ANCOVA indices S_j for PTTC from the three PCE-based models

Input	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6
PCE_correlate	0.0257	0.0000	0.1887	**0.2610**	**0.2512**	**0.2735**
PCE_NT	0.0437	0.0000	0.1548	**0.4499**	0.1896	0.1620
PCE_RT	0.0179	0.0000	0.1205	0.2586	0.1523	**0.3419**

5.4.4 The PTTC variation Control

From the estimated ANCOVA indices presented in Section 5.4.3, the dominant random inputs are identified from the three PCE-based models. Particularly, $S_2 = 0$ signifies that fluctuations in ζ_2 remain inconsequential to the variance of PTTC. To validate these findings, smooth out ζ_4, ζ_5, and ζ_6 (e.g., reduce their variance to zero) in **Step 7** one by one to observe how $\mathrm{Var}[\mathrm{PTTC}]$ changes across the PCE-based models and MC simulations. As evidenced by Table 5.3, the *PCE_correlate* model aligns most closely with the MC simulations. Clearly, smoothing ζ_4, ζ_5, and ζ_6 exerts similar impacts on $\mathrm{Var}[\mathrm{PTTC}]$. However, smoothing ζ_6 is slightly more effective in reducing $\mathrm{Var}[\mathrm{PTTC}]$ (e.g., leading to an approximate 20% reduction in the standard deviation of PTTC, i.e., $\hat{\sigma}[Y]$), thus enhancing the ATC. Such outcomes validate the accuracy of the sensitivity indices estimated by

PCE_correlate, as presented in Table 5.2.

Table 5.3 Case I: Statistics of PTTC before and after smoothing ζ_j by the benchmark MC simulations and the three PCE-based models

Methods	ζ_4		ζ_5		ζ_6		Before
	$\widehat{\sigma}[Y]$	$\Delta\sigma_{\mathrm{rr}}\%$	$\widehat{\sigma}[Y]$	$\Delta\sigma_{\mathrm{rr}}\%$	$\widehat{\sigma}[Y]$	$\Delta\sigma_{\mathrm{rr}}\%$	$\widehat{\sigma}[Y]$
PCE_correlate	0.7706	$\mathbf{-2.11}$	0.7785	$\mathbf{-1.73}$	0.7572	$\mathbf{-2.17}$	0.9494
PCE_NT	0.7692	-2.29	0.9605	21.24	0.9764	26.15	1.1192
PCE_RT	0.9147	16.20	1.0052	26.89	0.7491	-3.22	1.1336
MC	0.7872	–	0.7922	–	0.7740	–	0.9702

* $\Delta\sigma_{\mathrm{rr}} = (\hat{\sigma}[Y_{\mathrm{pc}}] - \hat{\sigma}[Y_{\mathrm{mc}}])/\hat{\sigma}[Y_{\mathrm{mc}}]$ denotes the normalized standard deviation estimation error by the three PCE models.
* The reduction of estimated standard deviation is calculated by: $\frac{\hat{\sigma}[\text{after}]-\hat{\sigma}[\text{before}]}{\hat{\sigma}[\text{before}]}\%$, e.g., For the MC simulations $\frac{\hat{\sigma}_{\mathrm{mc}}[\text{after}]-\hat{\sigma}_{\mathrm{mc}}[\text{before}]}{\hat{\sigma}_{\mathrm{mc}}[\text{before}]}\% \approx 20\%$.

5.4.5 Efficiency Comparison

Regarding efficiency, It is worth noting that most of the computational cost of PCE-based models lies in generating the M_p sample pairs $[\boldsymbol{\zeta}_p, \boldsymbol{Y}_p]$, i.e., $t_{\mathrm{ed}} \gg t_{\mathrm{sc}}$ and $t_{\mathrm{ed}} \gg t_{\mathrm{ancova}}$. As can be seen from Table 5.4, the three PCE-based models share similar computational time, and *PCE_correlate* requires 421.4s for the ANCOVA indices estimation and is slightly faster ($\approx$ 2s) than the other two PCE-based models.

Table 5.4 Comparison of computational time by the *PCE_correlate*, *PCE_NT*, *PCE_RT* with the size training sample pairs $M_p = 60$ for Case I.

Method	$t_{\mathrm{ed}}(s)$	$t_{\mathrm{sc}}(s)$	$t_{\mathrm{ancova}}(s)$	$t_{\mathrm{total}}(s)$
PCE_correlate	420.6080	0.6180	0.1810	421.4070
PCE_NT	420.6080	2.8504	0.1085	423.5669
PCE_RT	420.6080	2.6499	0.1172	423.3751

* t_{ed}: time for evaluating the training samples; t_{sc}: time for constructing the PCE models; t_{ancova}: time for evaluating the sensitivity indices S_j of $10,000$ samples.

5.5 Case Study II - Economic Dispatch Problem

To verify the proposed methods in a more complicated system, the second case presents the ANCOVA indices estimation for the economic dispatch (ED) problem considered in Chapter 4. In contrast to the first case, the second case uses real-world data from NREL's Western Wind Data Set [212], which exhibits unknown distribution types and potentially complicated correlations. To make matters more challenging, Y (the ED cost) is multimodal. The UQLab toolbox is adopted to build the PCE-based models and the ANCOVA indices calculation [137] [138].

5.5.1 System Configurations

This case tests the proposed PCE-based methods on the IEEE 118-bus system integrated with a 20-node gas system, namely, an integrated electricity and gas system (IEGS). 5 wind farms are added into the system at bus $\{2, 33, 51, 81, 108\}$ using the NREL's Western Wind Data Set [212]. The time period considered is 24 hours. Thus, there are 120 random inputs (24 time periods for each wind farm) and the response Y is the ED cost. The detailed configurations can be found in Section 4.5, Chapter 4.

5.5.2 The PCE-based Models Accuracy Test

Firstly, the three PCE-based models are constructed using $M_p = 1100$ sample-response pairs, as detailed in **Steps 1-3**. Similarly, the order H for all three models is chosen as 2, based on the hybrid LAR algorithm (refer to Section 3.3.3 and Appendix D). Then, $M_s = 10,000$ correlated input samples are fed to each model during **Step 5**. The estimated statistics for model accuracy comparison, are presented in Table 5.5. Additionally, the PDF and CDF derived from the three PCE-based models and MC simulations are illustrated in Fig. 5.3 and Fig. 5.4, respectively. A comparative analysis of these estimated statistics and probability distributions underscores that the *PCE_correlate* model aligns most closely with the benchmark MC simulations.

5.5.3 The ANCOVA Indices Estimation

Subsequently, the ANCOVA indices S_j are computed and ordered by significance in **Step 7**. The top 40 dominant random inputs identified by the three PCE-based models constitute a third of all inputs, which are depicted in Fig. 5.5. Notably, there are distinct differences in the top 40 dominant random inputs across the three PCE-based models.

Table 5.5 Comparison of the estimated statistics of the ED cost by the MC simulations, *PCE_correlate*, *PCE_NT* and *PCE_RT*

Index	MC	*PCE_correlate*	*PCE_NT*	*PCE_RT*
$\mathbb{E}[Y]$	7.3276×10^6	7.3275×10^6	7.3275×10^6	7.3379×10^6
$\hat{\sigma}[Y]$	4.7069×10^4	4.7090×10^4	5.9115×10^4	4.7920×10^4
$\frac{\Delta\mathbb{E}[Y]}{\mathbb{E}[Y]}\%$	–	-3.6935×10^{-4}	0.1349	0.1412
$\frac{\Delta\hat{\sigma}[Y]}{\hat{\sigma}[Y]}\%$	–	0.0449	25.5927	1.8083

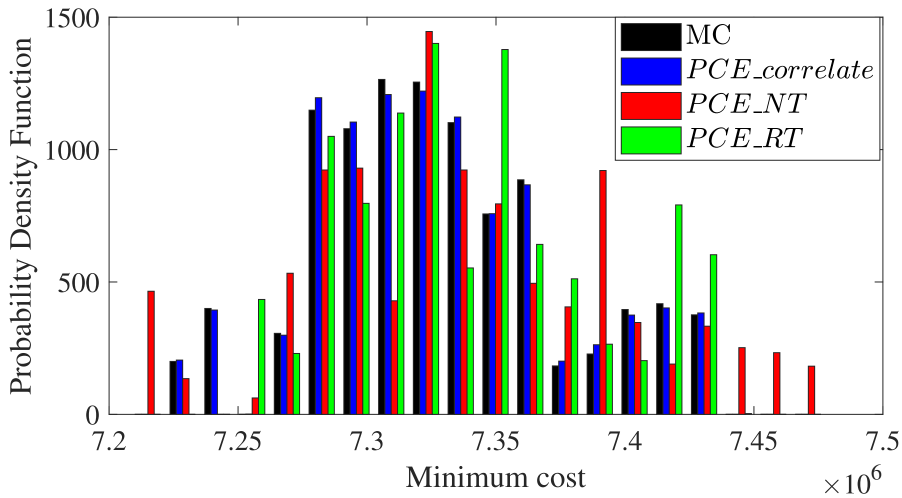

Fig. 5.3 The PDFs of the ED cost from *PCE_correlate*, *PCE_NT*, *PCE_RT* and the MC simulations.

5.5.4 Impacts of Smoothing Dominant Random Inputs on the ED cost

To validate the estimated ANCOVA indices, the top 40 inputs identified by each of the three PCE-based models are individually smoothed (i.e., meaning their variances are set to zero), and the results are then compared with MC simulations as outlined in **Step 7**. The statistical comparisons of the ED cost, both before and after smoothing the dominant inputs, are presented in Table 5.6. Additionally, Fig. 5.6 presents the CDFs before and after smoothing the top 40 dominant inputs through the MC simulations. Table 5.6 and Fig. 5.6 demonstrate that the *PCE_correlate* model is the most effective in reducing the variance of Y, achieving an approximate reduction of 95.9% in the standard deviation of Y. The *PCE_RT* model also produces commendably accurate results when benchmarked against MC simulations.

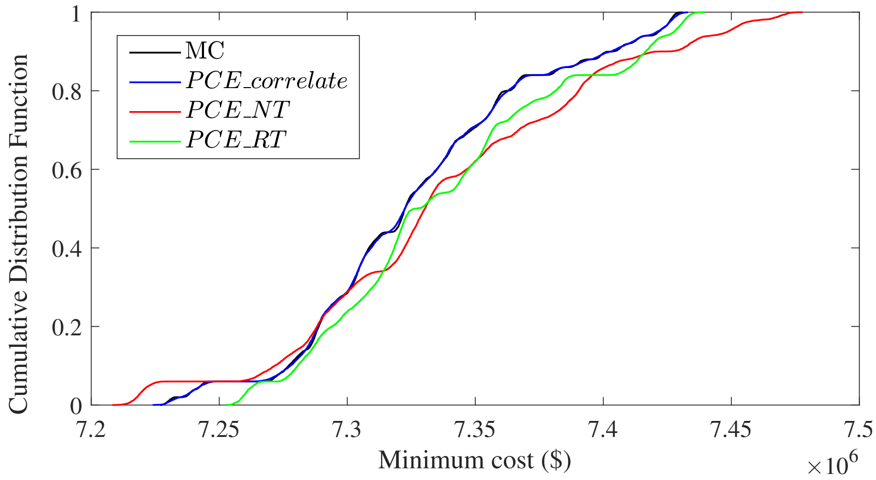

Fig. 5.4 The CDFs of the ED cost from the three PCE-based models and the MC simulations.

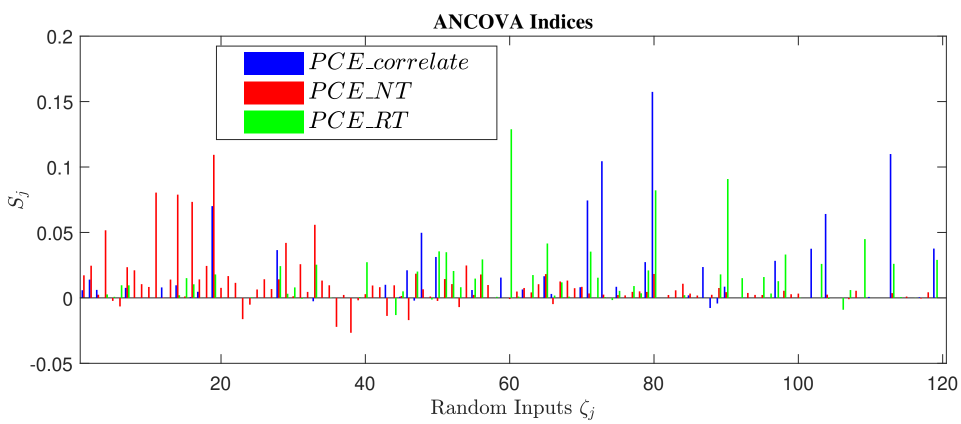

Fig. 5.5 S_j estimated by the three PCE-based models for the ED cost

5.5.5 Efficiency Comparison

Regarding efficiency, as shown in Table 5.7, *PCE_RT* requires 714.9s for ANCOVA indices estimation, which is much slower than *PCE_correlate* ($\approx$ 88s) and *PCE_NT* ($\approx$ 70s). For this case, due to the high dimension of the random inputs, the PCE-based models with decorrelated inputs requiring additional transform (the NT or the RT) share much more computation time.

Table 5.6 Case II: Comparison of the standard deviation of the ED cost before and after smoothing the three sets of top 40 dominant inputs by the benchmark MC simulations, and the three PCE-based models.

Methods	Set 1: Top 40s			Set 2: Top 40s			Set 3: Top 40s			Before
	$\hat{\sigma}[Y]$	$\Delta\sigma[Y]\%$	$\Delta\sigma_{\text{re}}\%$	$\hat{\sigma}[Y]$	$\Delta\sigma[Y]\%$	$\Delta\sigma_{\text{re}}\%$	$\hat{\sigma}[Y]$	$\Delta\sigma[Y]\%$	$\Delta\sigma_{\text{re}}\%$	$\hat{\sigma}[Y]$
PCE_correlate	1.4925×10^3	$\mathbf{-96.83}$	$\mathbf{-0.0093}$	6.0660×10^3	$\mathbf{-87.12}$	$\mathbf{-0.0020}$	1.0652×10^3	$\mathbf{-97.74}$	$\mathbf{-0.0190}$	4.7090×10^4
PCE_NT	1.0467×10^4	-82.98	-0.1422	1.1919×10^3	-97.98	0.1270	1.0586×10^4	-82.09	-0.1442	5.9115×10^4
PCE_RT	6.9201×10^2	-98.56	0.1194	5.1244×10^2	-98.93	0.1379	9.2717×10^2	-98.07	0.0224	4.7920×10^4
MC	1.9115×10^3	$\mathbf{-95.94}$	–	6.1449×10^3	$\mathbf{-86.94}$	–	1.9205×10^3	$\mathbf{-95.92}$	–	4.7069×10^4

* Set 1, Set 2, and Set 3 are the three sets of top 40 inputs from *PCE_correlate*, *PCE_NT* and *PCE_RT*, respectively. $\Delta\sigma[Y] = (\hat{\sigma}[Y_{\text{after}}] - \hat{\sigma}[Y_{\text{before}}])/\hat{\sigma}[Y_{\text{before}}]$; $\hat{\sigma}[Y_{\text{before}}]$ and $\hat{\sigma}[Y_{\text{after}}]$ denote the estimated standard deviation of Y before and after smoothing, respectively; $\Delta\sigma_{\text{re}} = (\Delta\sigma[Y_{\text{pc}}] - \Delta\sigma[Y_{\text{mc}}])/\Delta\sigma[Y_{\text{mc}}]$, describing how close the standard deviation reduction by the PCE-based model is to the one by the benchmark MC simulation.

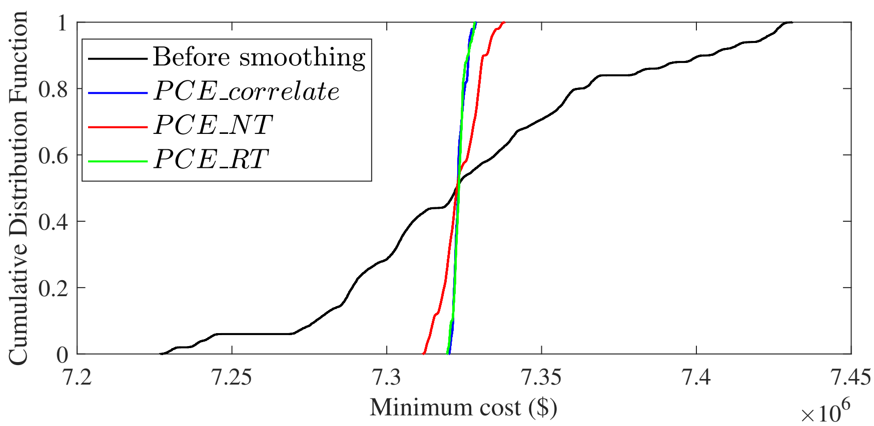

Fig. 5.6 The CDFs of the ED cost after smoothing the three sets of top 40 inputs by MC simulations.

Table 5.7 Comparison of computational time by the *PCE_correlate*, *PCE_NT* , *PCE_RT* with the size training sample pairs $M_p = 1100$ for case 2.

Method	$t_{\text{ed}}(s)$	$t_{\text{sc}}(s)$	$t_{\text{ancova}}(s)$	$t_{\text{total}}(s)$
PCE_correlate	582.5600	31.0738	13.2071	626.8409
PCE_NT	582.5600	48.2677	13.6878	644.5155
PCE_RT	582.5600	123.6416	8.6962	714.8978

* t_{ed}: time for evaluating the training samples; t_{sc}: time for constructing the PCE models; t_{ancova}: time for evaluating the sensitivity indices S_j of $10,000$ samples.

5.6 Discussions of the Results

In this section, the performance differences between the *first* and *second* methods are discussed.

The *first method:* For dependent random inputs, ensuring the orthogonality of polynomial bases requires the multivariate orthogonal polynomial bases $\boldsymbol{\Psi_\alpha}$ to be constructed as (see Appendix G)

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) = \left(\prod_{j=1}^{\mathcal{M}} \frac{f_{\zeta_j}(\zeta_j)}{f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})}\right)^{\frac{1}{2}} \prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(\zeta_j)$$

where $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ and $f_{\zeta_j}(\zeta_j)$ denote the joint PDF of $\boldsymbol{\zeta}$ and marginal PDF of ζ_j, respectively. This ensures the convergence of the PCE model to the response Y in the L_2 norm. However, *PCE_correlate* omits the density term $\left[\prod_{j=1}^{\mathcal{M}} \frac{f_{\zeta_j}(\zeta_j)}{f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})}\right]^{\frac{1}{2}}$ and uses only $\prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(\zeta_j)$ as the bases $\boldsymbol{\Psi_\alpha}$. This is because obtaining an accurate joint PDF $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ can be challenging, and the density term might be highly nonlinear. Despite this omission, the calculation of $c_{\boldsymbol{\alpha}}$ (as in **Step 3 c)** of **Method 1**) may compensate for the errors, resulting in a model with satisfactory accuracy, i.e., $Y \approx G^{\mathrm{pc}}(\boldsymbol{\zeta})$.

The *second* method: The methods *PCE_NT* and *PCE_RT* initiate by transforming the random inputs $\boldsymbol{\zeta}_p$ into $\boldsymbol{Z}_p$ using: $\boldsymbol{\zeta}_p = \mathcal{T}^{-1}(\boldsymbol{Z}_p)$. Following this transform, PCE-based models are formulated using the sample pairs $[\boldsymbol{Z}_p, \mathbf{Y}_p]$, which can be expressed as:

$$Y = G(\boldsymbol{\zeta}) \approx G^{\mathrm{pc}}(\boldsymbol{Z}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^{\mathcal{M}}} c_{\boldsymbol{k}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{Z}).$$

However, during the computation of ANCOVA indices in **Step 6** of **Method 2**, the PCE-based models replace $\boldsymbol{Z}$ with $\boldsymbol{\zeta}$. This implies that the model is approximated as $Y \approx G^{\mathrm{pc}}(\boldsymbol{\zeta})$, even though the PCE-based model formulated in **Step 3** is designed to ensure $Y \approx G^{\mathrm{pc}}(\boldsymbol{Z}) = G^{\mathrm{pc}}(\mathcal{T}(\boldsymbol{\zeta}))$. The transform $\mathcal{T}$ inherently introduces errors, and these cannot be compensated since the PCE-based models are solidified by **Step 3**. Consequently, the performance of the *second* method may not achieve the same level of performance as the *first* method.

5.7 Conclusions

This chapter compares three different PCE-based models for estimating ANCOVA indices-based GSA of two power system applications considering correlated random inputs. Simulation results demonstrate that *PCE_correlate*, ignoring the input dependencies, which provides the most accurate response and ANCOVA indices estimations in the two power system applications, compared with PCE models built after decorrelation using the Nataf or the Rosenblatt transform, i.e., *PCE_NT* or *PCE_RT*. Based on the ANCOVA indices determined by *PCE_correlate*, effective control measures for uncertainty management can be developed to reduce the system response variance and improve overall system performance.

Chapter 6

Summary and Conclusions

6.1 Thesis Summary

This thesis primarily contributed to the development of data-driven sparse PCE-based models, designed to quantify uncertainties in the probabilistic assessment of power system static security and ED. Through the development of data-driven PCE-based methods, this thesis facilitates the modeling of uncertainty, thereby enabling a comprehensive probabilistic static security assessment, ED, and a GSA-based approach for uncertainty control in power systems.

This thesis first focused on quantifying the impacts of uncertainties on ATC. Uncertainties are characterized as mixed random inputs (e.g., continuous random inputs include RESs and stochastic loads, and discrete random inputs like unexpected equipment outages). A subsequent introduction of a probabilistic CPF-based mathematical framework paved the way for PTTC assessment. After that, the DDSPCE method was devised for estimating the probability distributions of PTTC, based on which, ATC at a certain confidence interval was determined. A statistically-equivalent surrogate model was developed based on the DDSPCE method, to enhance the computational efficiency of the traditional PTTC calculation method, MC simulations. The proposed DDSPCE method, which directly utilized raw data, could handle a large number of mixed random inputs without the need for predefined probability distributions. The integration of the sparse PCE scheme further optimizes computational burden. Simulation studies on the IEEE 118-bus system and the IEEE 1354-bus system demonstrated the efficiency and efficacy of the proposed DDSPCE method in determining the probabilistic characteristics (e.g., mean, variance, PDF, and CDF) of PTTC. Furthermore, the integration of discrete random variables (e.g., equipment outages) in PTTC and ATC assessments was validated by their pronounced impacts on PTTC statistics and ATC levels.

Later on, the DDSPCE method was employed to study the impacts of uncertainties, particularly wind power, on ED. The surrogate model, built based on the DDSPCE method, accurately approximates the statistic information (e.g., mean, variance, PDF, and CDF) of the objective values of the ED problem. This method exploited data directly, without knowing any presupposed probability distributions. Simulation results, particularly on the integrated IEEE 118-bus power system and 20-node gas system, demonstrated the efficiency of the proposed method 33 times faster than MC simulations, even when confronted with a multimodal PDF of the SED solution.

Driven by the impacts of uncertainty on probabilistic transfer capability and ED in power system operations, this thesis also aspired to mitigate the impacts of uncertainties brought by volatile RES, to enhance system security and economic efficiency. To this end, the PCE methods were developed to perform GSA using the ANCOVA indices, such that the dominant random inputs were identified to assist in designing uncertainty control strategies. Specially, different PCE-based models for estimating ANCOVA indices of correlated inputs are designed and compared within power system transfer capability enhancement and ED. Simulation results showed that *PCE₋ correlate*, which ignores input factor dependencies, provided the most accurate ANCOVA indices compared with PCE-based models built after decorrelation using the NT or the RT. Effective uncertainty control measures were designed by leveraging the ANCOVA indices calculated by *PCE₋ correlate* to mitigate the impacts of uncertainties. This reduced the variance of system response and enhanced the system performance. These findings offered invaluable insights for uncertainty management and control design in real-world power system operations.

6.2 Conclusions

The primary research outcomes and contributions of this thesis are encapsulated as follows:

1. The introduction of a DDSPCE method, utilizing available random input data, is adept at accurately quantifying the impacts of uncertainties (e.g., arising from RESs, load fluctuations, and unexpected equipment outages) on ATC. The proposed method, without requiring any preassumed probabilistic distributions for random inputs, can efficiently and accurately estimate the probabilistic characteristics (mean, variance, PDF, and CDF) of PTTC, based on which, ATC with a certain confidence interval can be determined. An integrated sparse scheme further enhances its computational efficiency and accuracy. Additionally, the simulation results highlight the significance of integrating discrete variables (unexpected equipment

outages) in the PTTC and ATC evaluation. Control devices (e.g., adjustable transformers) and $N-K$ contingency can be readily incorporated into the PTTC formulation and ATC assessment.

2. This thesis develops a DDSPCE-based surrogate model for addressing the uncertainties of the ED during daily power system operations. This surrogate model developed directly from the raw data set of random variables, operates without preassumed distributions of random inputs. The DDSPCE-based surrogate model stands out in its accurate estimation of the statistical information (mean, variance, PDF, and CDF) of the ED solutions (e.g., objective function). Moreover, this proposed model can manage a vast number of random inputs with high efficiency. Extensive simulation studies conducted on an IEGS, utilizing real-world wind power data, highlight the efficacy and effectiveness of the proposed method in quantifying the impacts of uncertainties on ED solutions (e.g., the objective function). Notably, this method proves high accuracy even in scenarios where the ED solutions exhibit multimodality properties. These empirical findings accentuate the remarkable efficacy and efficiency of the proposed DDSPCE method in addressing a broad range of complex scenarios.

3. This thesis undertakes a detailed comparative analysis, evaluating different PCE-based models (both with and without decorrelation random inputs and incorporating diverse nonlinear transforms) for GSA in uncertainty control for power system security and ED. The derived ANCOVA-based global sensitivity indices from *PCE_correlate* enable to allocation of the dominant random inputs, based on which, the effective uncertainty control measures can be designed to mitigate the uncertainty impacts. The results drawn from this analysis provide important guidance for uncertainty management and control in enhancing power system security and economic efficiency.

6.3 Recommendations for Future Work

Based on the proposed methods, the contributions of this thesis, and its conclusions, the subsequent research avenues are suggested for future exploration:

- This thesis studied the impacts of unforeseen equipment failures (with certain probabilities), such as line and generator outages, on power system transfer capability. It is essential to also consider equipment outages resulting from real extreme events (e.g., typhoons, snowstorms, and wildfires). Subsequent research will delve into the impacts of extreme weather conditions on power system security. The proposed DDSPCE method will be further employed, leveraging real-world data from extreme events (e.g., considering the probability of line outages related to different wind speeds during windstorms).

- This thesis proposes a data-driven PCE-based surrogate model to quantify the uncertainties on ATC and PTTC in power systems, particularly focusing on the relationship between uncertainties and TTC for individual tasks. As discussed in [230], it is essential to consider causal relevance when assessing TTC across different transmission tasks to optimize the efficacy of data-driven approaches. Consequently, future research endeavors will delve deeper into the interplay between various transmission tasks when evaluating both ATC and TTC.

- This thesis proposed the DDSPCE method to estimate the probabilistic characteristics of PTTC and ED solutions (e.g., objective functions). The proposed PCE-based method, however, presumed that system response distributions are smooth. In contrast, certain response distributions (e.g., like the generator outputs in the ED problem), might exhibit nonsmooth characteristics. Additionally, in practical scenarios, only an extremely small number of evaluations might be available for training models. To address these challenges, future research will explore the integration of PCE methods with other metamodels (e.g., GPR).

- The quality of data used to build DDSPCE-based models is important to the performance of response (e.g., in PTTC or ED solutions) estimations. Nevertheless, as highlighted in [112], the performance of the proposed model may be degraded if outliers of the training dataset are not considered. To this end, future research will consider and address the outliers in the training dataset to improve the robustness of the proposed method.

- Regarding GSA-based uncertainty control, further analytical investigation will be carried out to comprehensively analyze and compare the performance of the three PCE-based models in GSA while considering different global sensitivity indices.

- This thesis designed uncertainty control measures based on the estimated ANCOVA global sensitivity indices, which aim to mitigate the impacts of uncertainties by smoothing the dominant random inputs (e.g., wind power fluctuations) utilizing ESSs. Future work involves designing ESSs control and optimization policies to improve system security [231].

Appendix A

Principal Component Analysis (PCA)

This appendix introduces the singular value decomposition (SVD)-based method for PCA [149, 232]. PCA is a statistical procedure that uses orthogonal transformations to convert correlated variables into a set of linearly uncorrelated variables called principal components. The SVD-based algorithm is one of the mathematical techniques to achieve this. Given random inputs $\boldsymbol{\xi} \in \mathbb{R}^{n\times m}$, it is not a prerequisite for the variables in $\boldsymbol{\xi}$ to exhibit a multivariate Gaussian distribution for the formulation of principal components [233]. However, it assumes that $\boldsymbol{\xi}$ are with mean $\mathbf{0}$. Besides, the different scales of realizations of $\boldsymbol{\xi}$ may affect the principal components. To this end, a standardization procedure is conducted before applying the SVD-based method for PCA:

$$\boldsymbol{X} = \frac{\boldsymbol{\xi} - \boldsymbol{\mu}_{\boldsymbol{\xi}}}{\boldsymbol{\sigma}_{\boldsymbol{\xi}}} \tag{A.1}$$

where $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ is the data after standardization. A SVD of $\boldsymbol{X}$ is:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{A.2}$$

where $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{V} \in \mathbb{R}^{m\times m}$ are orthogonal. The columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are formed by eigenvectors of $\boldsymbol{X}\boldsymbol{X}^T$ and $\boldsymbol{X}^T\boldsymbol{X}$, respectively. $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times m}$ is a rectangular diagonal matrix with non-negative diagonal entries σ_i be the singular values of $\boldsymbol{X}$, i.e., the square roots of nonzero eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$. These singular values are in descending order with the largest σ_1 in the first diagonal entry of $\boldsymbol{\Sigma}$ (i.e., the position $(1, 1)$).

The above SVD procedure can be linked to PCA by:

$$\boldsymbol{\zeta} = \boldsymbol{X}\boldsymbol{V} \tag{A.3}$$

where $\boldsymbol{V}$ contains the principal component coefficients, with each column of $\boldsymbol{V}$ representing the coefficients for one principal component, which are the eigenvectors of the covariance matrix of $\boldsymbol{X}$, and the rows are in descending order of component variance. Therefore, after PCA, $\boldsymbol{X}$ is decorrelated to $\boldsymbol{\zeta} \in \mathbb{R}^{n \times m}$. Note that, in this thesis, the dimension of random inputs $\boldsymbol{X}$ after PCA remains the same, while PCA can also be used for dimension reduction. Readers may refer to [149, 232] for more details about SVD and PCA.

Appendix B

Copula Theory and Transforms

Sklar's theorem establishes a fundamental relationship between joint distributions and marginal distributions of random variables. It states that if random variables $\boldsymbol{\xi} \in \mathbb{R}^n$ have a joint distribution denoted as $F_{\xi_1 \ldots \xi_n}(\xi_1, \ldots, \xi_n)$ and corresponding marginal distributions $F_{\xi_1}(\xi_1), \cdots, F_{\xi_n}(\xi_n)$, there exists a unique n-dimensional copula C such that [234]:

$$F_{\xi_1 \ldots \xi_n}(\xi_1, \ldots, \xi_n) = C\left(F_{\xi_1}(\xi_1), \ldots, F_{\xi_n}(\xi_n)\right) \tag{B.1}$$

Note that if $F_{\xi_i}, i \in \mathbb{R}^n$ are continuous, copula C is unique. Let $F_{\xi_i}(\xi_i) = U_i, i \in \mathbb{R}^n$, then the joint PDF of $\boldsymbol{\xi}$ can be obtained by:

$$f_{\boldsymbol{\xi}}(\xi) = \prod_{i=1}^{n} f_{\xi_i}(\xi_i) \cdot c(F_{\xi_1}(\xi_1), \cdots, F_{\xi_n}(\xi_n)) \tag{B.2}$$

with c be the copula density function being defined by $c(U_1, \cdots, U_n) = \frac{\partial^n C(U_1, \cdots, n)}{\partial U_1, \cdots, \partial U_n}$.

B.1 Gaussian Copula and Nataf Transform

The Nataf transform is a powerful tool to convert correlated random inputs $\boldsymbol{\xi}$ modeled by Gaussian copula dependence structure, to independent standard Gaussian random inputs $\boldsymbol{Z}$. Considering random inputs $\boldsymbol{\xi}$ with the correlation being modeled by Gaussian copula, which is characterized by its correlation matrix $\boldsymbol{R}$. The Nataf transform denoted as $T_{\mathrm{NT}}(\boldsymbol{\xi})$, is expressed as [93]:

$$\boldsymbol{Z} = \mathcal{T}_{\mathrm{NT}}(\boldsymbol{\xi}) = T_3 \circ T_2 \circ T_1(\boldsymbol{\xi}) \tag{B.3}$$

where the transform (B.3) convert correlated inputs $\boldsymbol{\xi}$ into standard Gaussian inputs $\boldsymbol{Z}$. Specially, the trio of transform T_1, T_2, and T_3 can be represented by:

$$
\begin{aligned}
T_1 : \boldsymbol{\xi} \mapsto \boldsymbol{U} &= [F_{\xi_1}(\xi_1), \ldots, F_{\xi_n}(\xi_n)]^T \\
T_2 : \boldsymbol{U} \mapsto \boldsymbol{V} &= [\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_n)]^T \\
T_3 : \boldsymbol{V} \mapsto \boldsymbol{Z} &= \boldsymbol{\Gamma}^{-1}\boldsymbol{V}
\end{aligned}
$$

where $\boldsymbol{U}$ is a sample derived from a Gaussian copula with the linear correlation matrix $\boldsymbol{R}$. Φ denotes the marginal CDF of a single standard Gaussian random input $U_i, i \in \mathbb{R}^n$. $\boldsymbol{\Gamma}$ is the distinct lower triangular matrix resulting from the Cholesky decomposition of $\boldsymbol{R}$, such that $\boldsymbol{R} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$. The non-diagonal element R_{ij} can be deduced using:

$$
\rho_{ij} = \iint_{\mathbb{R}^2} \left(\frac{\xi_i - \mu_i}{\sigma_i}\right)\left(\frac{\xi_j - \mu_j}{\sigma_j}\right) \Phi_{i,j}(V_i, V_j, R_{ij})\, dV_i\, dV_j
$$

with μ_i and σ_i being the mean and variance of ξ_i. $\Phi_{i,j}(V_i, V_j, R_{ij})$ be the joint PDF of random inputs ξ_i and ξ_j following standard Gaussian distribution with zero mean and correlation matrix $\boldsymbol{R}$.

B.2 Vine Copula and Rosenblatt Transform

Given random inputs $\boldsymbol{\xi}$ with highly nonlinear correlation or thick-tailed dependence, vine copula can be used to model the dependence structure. In this thesis, C-vine copula is applied [91]:

$$
c(\boldsymbol{\xi}) = \prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{j,j+i|\{1,\ldots,j-1\}}\left(\xi_{j|\{1,\ldots,j-1\}}, \xi_{j+i|\{1,\ldots,j-1\}}\right) \tag{B.4}
$$

where $c_{j,j+i|\{1,\ldots,j-1\}}\left(\xi_{j|\{1,\ldots,j-1\}}, \xi_{j+i|\{1,\ldots,j-1\}}\right)$ denotes the pair copula between ξ_j and ξ_{j+1} conditioned on $\xi_1, \cdots, \xi_{j-1}$. The Rosenblatt transform is a powerful tool, mapping $\boldsymbol{\xi}$ with copula structure in (B.4) into independent random inputs $\boldsymbol{Z}$ by [91]:

$$
\boldsymbol{Z} = \mathcal{T}_{\mathrm{RT}}(\boldsymbol{\xi}) = T_{\mathrm{RT}_1} \circ T_{\mathrm{PIT}}(\boldsymbol{\xi}) \tag{B.5}
$$

with

$$
T_{\mathrm{PIT}} : \boldsymbol{U} = [F_{\xi_1}(\xi_1), \ldots, F_{\xi_n}(\xi_n)] \tag{B.6}
$$

$$T_{\mathrm{RT}_1}:\begin{cases} Z_1 = F_{\xi_1}\left(\xi_1\right) \\ Z_2 = F_{\xi_2 \mid \xi_1}\left(\xi_2 \mid \xi_1\right) \\ \vdots \\ Z_n = F_{\xi_n \mid \xi_1, \ldots, \xi_{n-1}}\left(\xi_n \mid \xi_1, \ldots, \xi_{n-1}\right) \end{cases}. \tag{B.7}$$

where (B.6) indicates the probability integral transform and $F_{j|1,\cdots,n-1}$ are the CDFs of conditioned random inputs. Note that if the dependence structure of random inputs $\boldsymbol{\xi}$ is modeled by Gaussian copula, the Rosenblatt transform simplifies to the Nataf transform.

Readers may refer to [91, 93, 234] for more details about copula and the nonlinear transforms.

Appendix C

Uncertainty Modeling – Typical Probability Distributions

C.1 Uncertainty Modeling – Wind Speed

Typically, the wind speed in many locations around the world can be modeled by Weibull distribution, where v follows the following probability density function (PDF) [8, 18, 133]:

$$f_w(v) = \frac{\gamma_w}{c_w}\left(\frac{v}{c_w}\right)^{\gamma_w - 1} \exp\left[-\left(\frac{v}{c_w}\right)^{\gamma_w}\right] \tag{C.1}$$

where v is the wind speed, γ_w is the equivalent shape parameter and c_w is the scale parameter, respectively.

C.2 Uncertainty Modeling – Solar Irradiance

As discussed in [190], solar irradiance can be represented by a Beta distribution, with the following PDF:

$$f_R(r) = \frac{\Gamma(\alpha_R + \beta_R)}{\Gamma_R(\alpha_R)\,\Gamma_R(\beta_R)}\left(\frac{r}{r_{\max}}\right)^{\alpha_R - 1}\left(1 - \frac{r}{r_{\max}}\right)^{\beta_R - 1} \tag{C.2}$$

where α_R and β_R are the shape parameters of the distribution, Γ_R denotes the Gamma function, r and $r_{\max}(\mathrm{W/m^2})$ are the respective actual and maximum solar irradiance.

C.3 Uncertainty Modeling – Load Variation

As discussed in [193, 194], the load uncertainty can be modeled by a Gaussian distribution, where the active load power P_L follows the PDF given below:

$$f\left(P_L\right) = \frac{1}{\left(\sqrt{2\pi}\sigma_P\right)} \text{exp}\left(-\frac{\left(P_L - \mu_P\right)^2}{2\sigma_P^2}\right) \tag{C.3}$$

where μ_P and σ_P denote the mean value and the forecasting error (i.e., the standard deviation) of P_L, respectively, which can be provided by the load forecaster or historical data. Typically, σ_P is assumed to be $2\% - 10\%$ of the mean value of P_L. Generally, only the active power is predicted by the load forecaster, whereas the reactive power is determined under the assumption of constant power factor [153].

Appendix D

The Hybrid LAR Algorithm

The hybrid LAR is a modified version of the LAR in **Algorithm** 1 [5], which is applied to achieve a basis adaptive (both q-norm and degree adaptive). At the end of each LAR iteration, L_p predictors $\Psi_k(\boldsymbol{\zeta})$ are obtained. Instead of using the coefficients c_k calculated by the LAR algorithm, the preferred coefficients c_k are recalculated by OLS based on the current predictors $\Psi_k(\boldsymbol{\zeta})$.The sparse PCE scheme based on hybrid LAR with the proposed stop criteria (i.e., e_{cloo}) offers several notable advantages over the conventional leave-one-out error e_{loo}-based stop criteria [198]:

- Improved computational efficiency: The sparse PCE scheme requires running the LAR procedure only once for all M_p samples, whereas the conventional approach requires $M_p + 1$ iterations. This significantly reduces the computational burden.
- Enhanced resistance to overfitting: The proposed stop criteria effectively address the overfitting problem commonly encountered in PCE. By utilizing the corrected leave-one-out cross-validation error, the sparse PCE scheme provides better control over model complexity and prevents overfitting, resulting in more accurate and reliable predictions.
- Robustness with small training sample size: The proposed stop criteria exhibit excellent performance even with a limited number of training samples. This is particularly valuable in scenarios where obtaining a large dataset for training the PCE model may be challenging or costly.

Overall, the sparse PCE scheme with the suggested stop criteria offers a more computationally efficient and robust approach to probabilistic modeling, mitigating overfitting issues and accommodating small training sample sizes. A flowchart describing the hybrid LAR algorithm is provided in Fig. D.1.

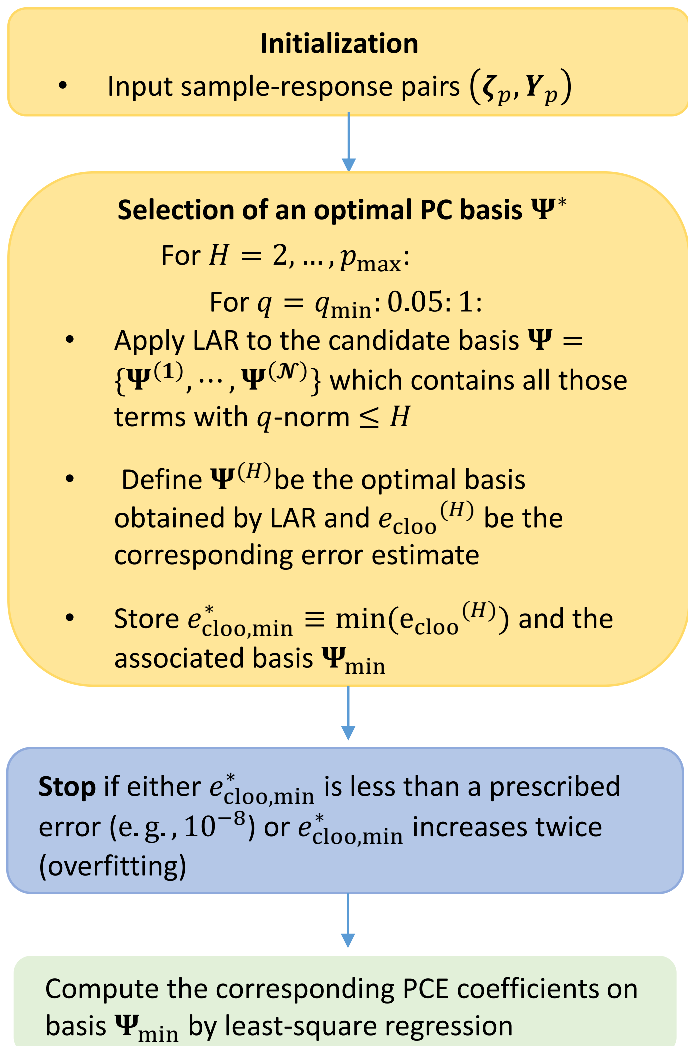

Fig. D.1 The hybrid LAR flowchart [5].

Appendix E

The Stochastic-Optimization ED formulations for IEGS

The mathematical formulation of the stochastic-optimization ED (SED) problem for IEGS is as follows.

$$Q(\boldsymbol{P_g}^{s_\mathrm{w}}, \boldsymbol{P_w}^{s_\mathrm{w}}) = \min_{\boldsymbol{P_g}} \sum_{t \in T} \left(\sum_{g \in G} C_g(P_g^{t,s_\mathrm{w}}) + \sum_{s \in S} C_s(g_s^{t,s_\mathrm{w}}) \right) \tag{E.1}$$

where (E.1) is the objective function, i.e., to minimize the total production cost; t is the specific time period in the time periods set $T = \{1, \cdots, T_m\}$ (e.g., 24-hour period in the simulation study); g is the generator index and G is the generator set. Power network constraints are as follows.

$$\text{s.t.}$$

$$\sum_{g \in G} P_g^{t,s_\mathrm{w}} + \sum_{w \in W} P_w^{t,s_\mathrm{w}} = \sum_{d \in D} P_d^t \quad \forall t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.2a}$$

$$\underline{P_l} \le \sum_{g \in G} k_{lg} P_g^{t,s_\mathrm{w}} + \sum_{w \in W} k_{lw} P_w^{t,s_\mathrm{w}} - \sum_{d \in D} k_{ld} P_d^t \le \overline{P_l} \quad \forall t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.2b}$$

$$P_g^{\min} x_g^t \le P_g^{t,s_\mathrm{w}} \le P_g^{\max} x_g^t \quad \forall g \in G, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.2c}$$

$$\begin{aligned} & - R_g^{RD} x_g^t - R_g^{SD}(x_g^{t-1} - x_g^t) - P_g^{\max}(1 - x_g^{t-1}) \\ & \le P_g^{t,s_\mathrm{w}} - P_g^{t-1,s_\mathrm{w}} \le R_g^{RU} x_g^{t-1} + R_g^{SU}(x_g^t - x_g^{t-1}) \\ & + P_g^{\max}(1 - x_g^t) \quad \forall g \in G, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \end{aligned} \tag{E.2d}$$

The operational and physical constraints are given in (E.2a)-(E.2d) based on the direct current (DC) power flow model. Equation (E.2a) is the power balance constraint, where P_d^t is the d-th load demand at time t. Constraint (E.2b) denotes the power flow limits, where k_{lg}, k_{lw} and k_{ld} are the sensitivity coefficients for the l-th transmission line with respect to the traditional generator g, wind generator w and load d, respectively [209]. $\underline{P_l}$ and $\overline{P_l}$ are thermal limits of transmission line l. Constraint (E.2c) denotes the generation capacity limits with x_g^t being the pre-determined UC decision for generator g at time t. Constraint (E.2d) describes the ramping capability constraint of generator g, where R_g^{RD} and R_g^{RU} denote the ramping down and up rate; R_g^{SD} and R_g^{SU} denote the shut-down and start-up ramp rate. $C_s(\boldsymbol{.})$ is the cost of gas well.
The gas network constraints given in (E.3).

$$G_s^{\min} \leq g_s^{t,s_\mathrm{w}} \leq G_s^{\max} \quad \forall s \in S, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.3a}$$

$$G_a^{\min} \leq \pi_a^{t,s_\mathrm{w}} \leq G_a^{\max} \quad \forall a \in G_a, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.3b}$$

$$g_b^{t,s_\mathrm{w}} = W_b \sqrt{((\pi_{e(b)}^{t,s_\mathrm{w}})^2 - (\pi_{a(b)}^{t,s_\mathrm{w}})^2)} \quad \forall b \in G_b, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.3c}$$

$$\pi_{e(c)}^{t,s_\mathrm{w}} \leq \alpha_c \pi_{a(c)}^{t,s_\mathrm{w}} \quad \forall c \in G_c, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.3d}$$

$$0 \leq g_b^{t,s_\mathrm{w}} \leq G_b \quad \forall b \in G_b, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.3e}$$

$$0 \leq g_c^{t,s_\mathrm{w}} \leq G_c \quad \forall c \in G_c, t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \tag{E.3f}$$

$$\begin{aligned} &\sum_{s \in G_s} g_{s(a)}^{t,s_\mathrm{w}} + \sum_{b_1 \in G_b} g_{b_1(a)}^{t,s_\mathrm{w}} - \sum_{b_2 \in G_b} g_{b_2(a)}^{t,s_\mathrm{w}} + \sum_{c_1 \in G_c} g_{c_1(a)}^{t,s_\mathrm{w}} \\ &- \sum_{c_2 \in G_c} g_{c_2(a)}^{t,s_\mathrm{w}} = \sum_{d \in G_d} G_{d(a)}^t + \sum_{g \in G} \Theta_g P_{g(a)}^{t,s_\mathrm{w}} \quad t \in T, s_\mathrm{w} \in \boldsymbol{\mathcal{S}} \end{aligned} \tag{E.3g}$$

where (E.3a) denotes the output capacity limits of a gas well; constraint (E.3b) denotes the nodal pressure range; constraint (E.3c) describes the gas flow g_b^{t,s_w} in gas passive pipeline b; constraint (E.3d) is the simplified gas compressor model; constraints (E.3e)-(E.3f) denote the gas flow transmission capacity limits in a gas passive pipeline and gas compressor, respectively; constraint (E.3g) denotes the gas nodal balance. The detailed notations can be found in [209]. Note that in formulation (4.3)-(E.3), constraint (E.2b) is based on the direct current (DC) power flow model, which states the power transmission line capacity. Detailed explanations are given below. The DC power flow is with the assumption that relates the real power injection to bus voltage angles δ by neglecting line resistances and voltage magnitudes at all buses are assumed to be 1 per unit [235]. The flow of active power P_{ij} is given by:

$$P_{ij} = B'_{ij}\delta_{ij} \qquad \text{or} \qquad \delta_{ij} = A'_{ij}P_{ij} \tag{E.4}$$

where $B'_{ij} = -\frac{1}{x_{ij}}$ for $i \neq$ reference bus and $j \neq$ reference bus, which is the element of the bus susceptance matrix B' and A' is the power system transmission network incidence matrix. Let a_i and a_j be the ith and jth rows of A'. k_{ln} indicates the sensitivity factor of line l flow with respect to unit n can be represented by:

$$k_{ln} = k_{ij} = \frac{a_{in} - a_{jn}}{x_{ij}} \tag{E.5}$$

Then we separate system buses as generation and load buses, the transmission line capacity constraints can be expressed by:

$$P_l = \sum_{g\in G} k_{lg}P_g^{t,s_{\mathrm{w}}} + \sum_{w\in W} k_{lw}P_w^{t,s_{\mathrm{w}}} - \sum_{d\in D} k_{ld}P_d^t \tag{E.6}$$

where k_{lg}, k_{lw} and k_{ld} are the sensitivity coefficients for line l flow with respect to the power output of generators g, wind farms w, and power loads d.

Remark E.0.1. It should be noted that the coupling effect of the gas system and its impact on the uncertainty analysis of the power system have not been addressed within the scope of this thesis. These aspects present potential avenues for future research endeavors.

Appendix F

Random Inputs Parameter Configurations: Chapter 5 Case Study I - Available Transfer Capability Enhancement

This appendix presents the random parameter configurations of the case study in Section 5.4 Chapter 5.

Parameter Configurations of Wind Speed Distributions

It is assumed that the probability distributions and corresponding parameters of all random inputs are available. The probabilistic data of all random inputs are generated from assumed probability distributions. Specially, the wind speed empirical distribution is characterized by a Weibull distribution [7, 8]. A detailed description of wind speed distribution can be found in Appendix C. Table F.1 provides the parameters of wind speed. Once the wind speed data is obtained, the wind turbine generator's real output power is calculated through the wind speed-power curve through (3.7).

Table F.1 Wind speed and wind turbine parameters [7, 8]

Bus	c_w	γ	P_r	V_r	V_{in}	V_{out}	PF
1	11.5762	2.7022	150.00	13.50	3.50	25.00	0.95
2	11.5762	2.7022	150.00	13.50	3.50	25.00	0.95
15	11.5762	2.7022	150.00	13.50	3.50	25.00	0.95

Parameter Configurations of Solar Radiation Distributions

This section presents the parameter configurations of solar radiations. it is assumed that solar radiation is characterized by Beta distribution [9], with a detailed description in Appendix C. Table F.2 provides the parameters of the solar radiation. Once the solar radiation data is obtained, the solar generator's real output power is calculated through the radiation-power curve through (3.8).

Table F.2 Solar radiation and solar PV parameters [9]

Bus	α_R	β_R	r_{min}	r_{max}	P_r	R_c	R_{std}	PF
16	1.110	0.730	0.0	1000.0	150.00	150.0	1000.0	1.0
18	1.110	0.730	0.0	1000.0	150.00	150.0	1000.0	1.0
21	1.110	0.730	0.0	1000.0	150.00	150.0	1000.0	1.0

Correlation Matrix of Random Inputs

The case study in Section 5.4 Chapter 5 assumes linear correlations between random inputs with the following correlation matrix $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0.2 & 0.3 & 0 & 0 & 0 \\ 0.2 & 1 & 0.1 & 0 & 0 & 0 \\ 0.3 & 0.1 & 1 & -0.2 & 0 & 0 \\ 0 & 0 & -0.2 & 1 & 0.28 & 0.35 \\ 0 & 0 & 0 & 0.28 & 1 & 0.1 \\ 0 & 0 & 0 & 0.35 & 0.1 & 1 \end{bmatrix}$$

Appendix G

Construction of Multivariate Orthogonal Polynomial Basis – Dependent Random Inputs

For dependent random inputs, Soize et al. [145] (see *Lemma 1*) suggested another method for constructing the PCE model considering dependent random inputs with the following set of functions for the multivariate orthogonal polynomial bases:

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) = \left(\prod_{j=1}^{\mathcal{M}} \frac{f_{\zeta_j}(\zeta_j)}{f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})} \right)^{\frac{1}{2}} \prod_{j=1}^{\mathcal{M}} \phi_j^{(\alpha_k^j)}(\zeta_j) \tag{G.1}$$

where $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ and $f_{\zeta_j}(\zeta_j)$ denote the joint PDF of $\boldsymbol{\zeta}$ and marginal PDF of ζ_j, respectively. Instead of directly forming the multivariate polynomial bases $\Psi_k(\boldsymbol{\zeta})$ through the tensor product of univariate polynomial bases using (2.7), a density term $\left(\prod_{j=1}^{\mathcal{M}} \frac{f_{\zeta_j}(\zeta_j)}{f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})} \right)^{\frac{1}{2}}$ was further added in constructing $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$, where the orthogonality of $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ has been demonstrated [145].

References

[1] International Renewable Energy Agency (IRENA), “Renewable capacity statistics 2023,” 2023.

[2] International Renewable Energy Agency (IRENA), “Renewable capacity highlights 2023,” 2023.

[3] Energy Information Administration (EIA) of the U.S. Department of Energy, “U.s. electricity generation by energy source.” Accessed: 2023-08-09.

[4] T. T. C. T. Force, “Available transfer capability definitions and determination,” tech. rep., North American Electric Reliability Council, 1996.

[5] G. Blatman and B. Sudret, “Adaptive sparse polynomial chaos expansion based on least angle regression,” *Journal of computational Physics*, vol. 230, no. 6, pp. 2345–2367, 2011.

[6] D. Xiu and G. E. Karniadakis, “The wiener–askey polynomial chaos for stochastic differential equations,” *SIAM journal on scientific computing*, vol. 24, no. 2, pp. 619–644, 2002.

[7] X. Wang, H.-D. Chiang, J. Wang, H. Liu, and T. Wang, “Long-term stability analysis of power systems with wind power based on stochastic differential equations: Model development and foundations,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1534–1542, 2015.

[8] R. Karki, P. Hu, and R. Billinton, “A simplified wind power generation model for reliability evaluation,” *IEEE transactions on Energy conversion*, vol. 21, no. 2, pp. 533–540, 2006.

[9] Z. M. Salameh, B. S. Borowy, and A. R. Amin, “Photovoltaic module-site matching based on the capacity factors,” *IEEE transactions on Energy conversion*, vol. 10, no. 2, pp. 326–332, 1995.

[10] K. C. Seto, G. Churkina, A. Hsu, M. Keller, P. W. Newman, B. Qin, and A. Ramaswami, “From low- to net-zero carbon cities: The next global agenda,” *Annual Review of Environment and Resources*, vol. 46, no. 1, pp. 377–415, 2021.

[11] L. C. Kunkel, H. L. Breetz, and J. K. Abbott, “100% renewable electricity policies in us cities: A mixed methods analysis of adoption and implementation,” *Energy Policy*, vol. 167, p. 113053, 2022.

[12] G. Papaefthymiou and K. Dragoon, “Towards 100% renewable energy systems: Uncapping power system flexibility,” *Energy Policy*, vol. 92, pp. 69–82, 2016.

[13] M. O’Malley, “Editorial: Towards 100Transactions on Power Systems, vol. 37, no. 4, pp. 3187–3189, 2022.

[14] A. Hailemariam, K. Ivanovski, and R. Dzhumashev, “Does r&d investment in renewable energy technologies reduce greenhouse gas emissions?,” *Applied Energy*, vol. 327, p. 120056, 2022.

[15] R. Cergibozan, “Renewable energy sources as a solution for energy security risk: Empirical evidence from oecd countries,” *Renewable Energy*, vol. 183, pp. 617–626, 2022.

[16] M. Nazemi and P. Dehghanian, “Powering through wildfires: An integrated solution for enhanced safety and resilience in power grids,” *IEEE Transactions on Industry Applications*, vol. 58, no. 3, pp. 4192–4202, 2022.

[17] W. Zappa and M. Van Den Broek, “Analysing the potential of integrating wind and solar power in europe using spatial optimisation under various scenarios,” *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 1192–1216, 2018.

[18] R. Zárate-Miñano, M. Anghel, and F. Milano, “Continuous wind speed models based on stochastic differential equations,” *Applied Energy*, vol. 104, pp. 42–49, 2013.

[19] D. P. Larson, L. Nonnenmacher, and C. F. Coimbra, “Day-ahead forecasting of solar power output from photovoltaic plants in the american southwest,” *Renewable Energy*, vol. 91, pp. 11–20, 2016.

[20] Y. Zhang, N. Akkurt, J. Yuan, Z. Xiao, Q. Wang, and W. Gang, “Study on model uncertainty of water source heat pump and impact on decision making,” *Energy and Buildings*, vol. 216, p. 109950, 2020.

[21] F. Morlock, B. Rolle, M. Bauer, and O. Sawodny, “Forecasts of electric vehicle energy consumption based on characteristic speed profiles and real-time traffic data,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1404–1418, 2020.

[22] J. D. Glover, M. S. Sarma, and T. Overbye, *Power system analysis & design, SI version*. Cengage Learning, 2012.

[23] Y. Xu, M. Korkali, L. Mili, X. Chen, and L. Min, “Risk assessment of rare events in probabilistic power flow via hybrid multi-surrogate method,” *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1593–1603, 2020.

[24] Y. V. Makarov, Z. Huang, P. V. Etingov, J. Ma, R. T. Guttromson, K. Subbarao, and B. B. Chakrabarti, “Wind energy management system ems integration project: Incorporating wind generation and load forecast uncertainties into power grid operations,” tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2010.

[25] G. Pierrou, *Long-Term Voltage Stability of Power Systems Under Uncertainty: Analytical Methods and Data-Driven Control*. McGill University (Canada), 2021.

[26] G. Pierrou and X. Wang, “Analytical study of the impacts of stochastic load fluctuation on the dynamic voltage stability margin using bifurcation theory,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 4, pp. 1286–1295, 2020.

[27] J. Liu, X. Wang, and X. Wang, “A sparse polynomial chaos expansion-based method for probabilistic transient stability assessment and enhancement,” in *2022 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2022.

[28] X. Wang, H.-D. Chiang, J. Wang, H. Liu, and T. Wang, “Long-term stability analysis of power systems with wind power based on stochastic differential equations: Model development and foundations,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1534–1542, 2015.

[29] J. McCalley, S. Asgarpoor, L. Bertling, R. Billinion, H. Chao, J. Chen, J. Endrenyi, R. Fletcher, A. Ford, C. Grigg, G. Hamoud, D. Logan, A. Meliopoulos, M. Ni, N. Rau, L. Salvaderi, M. Schilling, Y. Schlumberger, A. Schneider, and C. Singh, “Probabilistic security assessment for power system operations,” in *IEEE Power Engineering Society General Meeting, 2004.*, pp. 212–220 Vol.1, 2004.

[30] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, T. Van Cutsem, and V. Vittal, “Definition and classification of power system stability ieee/cigre joint task force on stability terms and definitions,” *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1387–1401, 2004.

[31] N. Balu, T. Bertram, A. Bose, V. Brandwajn, G. Cauley, D. Curtice, A. Fouad, L. Fink, M. Lauby, B. Wollenberg, and J. Wrubel, “On-line power system security analysis,” *Proceedings of the IEEE*, vol. 80, no. 2, pp. 262–282, 1992.

[32] A. Gomez-Exposito, A. J. Conejo, and C. Cañizares, *Electric energy systems: analysis and operation*. CRC press, 2018.

[33] S. Kalyani and K. S. Swarup, "Classification and assessment of power system security using multiclass svm," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 5, pp. 753–758, 2010.

[34] K. Morison, L. Wang, and P. Kundur, "Power system security assessment," *IEEE power and energy magazine*, vol. 2, no. 5, pp. 30–39, 2004.

[35] T. Mühlpfordt, L. Roald, V. Hagenmeyer, T. Faulwasser, and S. Misra, "Chance-constrained ac optimal power flow: A polynomial chaos approach," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4806–4816, 2019.

[36] H. Sheng and X. Wang, "Probabilistic power flow calculation using non-intrusive low-rank approximation method," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3014–3025, 2019.

[37] C. Safta, R. L.-Y. Chen, H. N. Najm, A. Pinar, and J.-P. Watson, "Efficient uncertainty quantification in stochastic economic dispatch," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2535–2546, 2016.

[38] Y. Xu, L. Mili, M. Korkali, K. Karra, Z. Zheng, and X. Chen, "A data-driven nonparametric approach for probabilistic load-margin assessment considering wind power penetration," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4756–4768, 2020.

[39] X. Kou and F. Li, "Interval optimization for available transfer capability evaluation considering wind power uncertainty," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 250–259, 2020.

[40] R. Marceau, J. Endrenyi, R. Allan, F. Alvarado, G. Bloemhof, T. Carlsen, G. Couto, E. Dialynas, N. Hatziargyriou, D. Holmberg, *et al.*, "Power system security assessment: A position paper," *Electra*, vol. 175, 1997.

[41] M. Ilić, F. Galiana, L. Fink, A. Bose, P. Mallet, and H. Othman, "Transmission capacity in power networks," *International Journal of Electrical Power & Energy Systems*, vol. 20, no. 2, pp. 99–110, 1998.

[42] U. for the Co-ordination of Transmission of Electricity (UCTE), "Policy 3: Operational security," 2009. Final Policy, approved by SC on 19.03.2009.

[43] G. Sombuttwilailert and B. Eua-Arporn, "A novel sensitivity analysis for total transfer capability evaluation," in *PICA 2001. Innovative Computing for Power - Electric Energy Meets the Market. 22nd IEEE Power Engineering Society. International Conference on Power Industry Computer Applications (Cat. No.01CH37195)*, pp. 342–347, 2001.

[44] C. I. S. O. C. (CAISO), "Transmission reliability margin implementation document," 2023. NERC Reliability Standard MOD-008-1, Version No. 3.1, Effective on May 04, 2023.

[45] J. Zhang, J. Yan, D. Infield, Y. Liu, and F.-s. Lien, "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and gaussian mixture model," *Applied Energy*, vol. 241, pp. 229–244, 2019.

[46] Y. Wang, Q. Hu, D. Meng, and P. Zhu, "Deterministic and probabilistic wind power forecasting using a variational bayesian-based adaptive robust multi-kernel regression model," *Applied energy*, vol. 208, pp. 1097–1112, 2017.

[47] T. Nikolakakis, *A Mixed Integer Linear Unit Commitment and Economic Dispatch Model for Thermo-Electric and Variable Renewable Energy Generators With Compressed Air Energy Storage*. PhD thesis, Columbia University, 2017.

[48] W. Wangdee and R. Billinton, "Bulk electric system well-being analysis using sequential monte carlo simulation," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 188–193, 2006.

[49] G. Carpinelli, P. Caramia, and P. Varilone, "Multi-linear monte carlo simulation method for probabilistic load flow of distribution systems with wind and photovoltaic generation systems," *Renewable Energy*, vol. 76, pp. 283–295, 2015.

[50] A. Berizzi, C. Bovo, M. Delfanti, M. Merlo, and M. S. Pasquadibisceglie, "A monte carlo approach for ttc evaluation," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 735–743, 2007.

[51] D. Xiu and J. S. Hesthaven, "High-order collocation methods for differential equations with random inputs," *SIAM Journal on Scientific Computing*, vol. 27, no. 3, pp. 1118–1139, 2005.

[52] 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 Transactions on Power Systems*, vol. 24, no. 2, pp. 661–667, 2009.

[53] Y. Chen, J. Wen, and S. Cheng, "Probabilistic load flow method based on nataf transformation and latin hypercube sampling," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 294–301, 2013.

[54] M. Hajian, W. D. Rosehart, and H. Zareipour, "Probabilistic power flow by monte carlo simulation with latin supercube sampling," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1550–1559, 2012.

[55] Z. Q. Xie, T. Y. Ji, M. S. Li, and Q. H. Wu, "Quasi-monte carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2239–2247, 2018.

[56] A. B. Krishna and A. R. Abhyankar, "Uniform experimental design-based nonparametric quasi-monte carlo for efficient probabilistic power flow," *IEEE Transactions on Power Systems*, vol. 38, no. 3, pp. 2318–2332, 2023.

[57] S. V. Dhople, Y. C. Chen, L. DeVille, and A. D. Domínguez-García, "Analysis of power system dynamics subject to stochastic power injections," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 12, pp. 3341–3353, 2013.

[58] L. Sanabria and T. Dillon, "Stochastic power flow using cumulants and von mises functions," *International Journal of Electrical Power & Energy Systems*, vol. 8, no. 1, pp. 47–60, 1986.

[59] J. Usaola, "Probabilistic load flow with wind production uncertainty using cumulants and cornish–fisher expansion," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 9, pp. 474–481, 2009.

[60] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2251–2261, 2012.

[61] P. Zhang and S. Lee, "Probabilistic load flow computation using the method of combined cumulants and gram-charlier expansion," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 676–682, 2004.

[62] E. Rosenblueth, "Point estimates for probability moments," *Proceedings of the National Academy of Sciences*, vol. 72, no. 10, pp. 3812–3814, 1975.

[63] E. Rosenblueth, "Two-point estimates in probabilities," *Applied mathematical modelling*, vol. 5, no. 5, pp. 329–335, 1981.

[64] H. Hong, "An efficient point estimate method for probabilistic analysis," *Reliability Engineering & System Safety*, vol. 59, no. 3, pp. 261–267, 1998.

[65] C.-L. Su, "Probabilistic load-flow computation using point estimate method," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1843–1851, 2005.

[66] C.-L. Su and C.-N. Lu, "Two-point estimate method for quantifying transfer capability uncertainty," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 573–579, 2005.

[67] G. Verbic and C. Canizares, "Probabilistic optimal power flow in electricity markets based on a two-point estimate method," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1883–1893, 2006.

[68] X. Ai, J. Wen, T. Wu, and W.-J. Lee, "A discrete point estimate method for probabilistic load flow based on the measured data of wind power," *IEEE Transactions on Industry Applications*, vol. 49, no. 5, pp. 2244–2252, 2013.

[69] J. M. Morales and J. Perez-Ruiz, "Point estimate schemes to solve the probabilistic power flow," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1594–1601, 2007.

[70] L. A. Gallego, J. F. Franco, and L. G. Cordero, "A fast-specialized point estimate method for the probabilistic optimal power flow in distribution systems with renewable distributed generation," *International Journal of Electrical Power & Energy Systems*, vol. 131, p. 107049, 2021.

[71] Y. Xu, M. Korkali, L. Mili, X. Chen, and L. Min, "Risk assessment of rare events in probabilistic power flow via hybrid multi-surrogate method," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1593–1603, 2020.

[72] Y. Xu, Z. Hu, L. Mili, M. Korkali, and X. Chen, "Probabilistic power flow based on a gaussian process emulator," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3278–3281, 2020.

[73] K. Ye, J. Zhao, H. Zhang, and Y. Zhang, "Data-driven probabilistic voltage risk assessment of miniwecc system with uncertain pvs and wind generations using realistic data," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 4121–4124, 2022.

[74] Y. Yang, Z. Yang, J. Yu, B. Zhang, Y. Zhang, and H. Yu, "Fast calculation of probabilistic power flow: A model-based deep learning approach," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2235–2244, 2020.

[75] S. R, R. S. Kumar, and A. T. Mathew, "Online static security assessment module using artificial neural networks," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4328–4335, 2013.

[76] M. M. Rajabi, "Review and comparison of two meta-model-based uncertainty propagation analysis methods in groundwater applications: polynomial chaos expansion and gaussian process emulation," *Stochastic environmental research and risk assessment*, vol. 33, pp. 607–631, 2019.

[77] D. Xiu and G. E. Karniadakis, "Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos," *Computer methods in applied mechanics and engineering*, vol. 191, no. 43, pp. 4927–4948, 2002.

[78] G. Blatman, *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. PhD thesis, Clermont-Ferrand 2, 2009.

[79] M. D. Spiridonakos, E. N. Chatzi, and B. Sudret, "Polynomial chaos expansion models for the monitoring of structures under operational variability," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, vol. 2, no. 3, p. B4016003, 2016.

[80] M. A. Hariri-Ardebili and B. Sudret, "Polynomial chaos expansion for uncertainty quantification of dam engineering problems," *Engineering Structures*, vol. 203, p. 109631, 2020.

[81] L. Hawchar, C.-P. El Soueidy, and F. Schoefs, "Principal component analysis and polynomial chaos expansion for time-variant reliability problems," *Reliability Engineering & System Safety*, vol. 167, pp. 406–416, 2017.

[82] J. Xu and D. Wang, "Structural reliability analysis based on polynomial chaos, voronoi cells and dimension reduction technique," *Reliability Engineering & System Safety*, vol. 185, pp. 329–340, 2019.

[83] H. Dai, Z. Zheng, and H. Ma, "An explicit method for simulating non-gaussian and non-stationary stochastic processes by karhunen-loève and polynomial chaos expansion," *Mechanical Systems and Signal Processing*, vol. 115, pp. 1–13, 2019.

[84] G. Liang, H. Li, Y. Huang, L. Liu, J. Lin, X. Chen, and W. Qiu, "Multilevel sampling based sparse polynomial expansion for efficient uncertainty analysis of mechanical system," *Mechanical Systems and Signal Processing*, vol. 186, p. 109913, 2023.

[85] B. Chouvion and E. Sarrouy, "Development of error criteria for adaptive multi-element polynomial chaos approaches," *Mechanical Systems and Signal Processing*, vol. 66, pp. 201–222, 2016.

[86] Z. Ren, W. Li, R. Billinton, and W. Yan, "Probabilistic power flow analysis based on the stochastic response surface method," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2307–2315, 2016.

[87] T. Mühlpfordt, T. Faulwasser, and V. Hagenmeyer, "Solving stochastic ac power flow via polynomial chaos expansion," in *2016 IEEE Conference on Control Applications (CCA)*, pp. 70–76, 2016.

[88] Mühlpfordt, Tillmann and Roald, Line and Hagenmeyer, Veit and Faulwasser, Timm and Misra, Sidhant, "Chance-constrained AC optimal power flow: A polynomial chaos approach," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4806–4816, 2019.

[89] F. Ni, P. H. Nguyen, and J. F. Cobben, "Basis-adaptive sparse polynomial chaos expansion for probabilistic power flow," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 694–704, 2016.

[90] Y. Xu, L. Mili, and J. Zhao, "Probabilistic power flow calculation and variance analysis based on hierarchical adaptive polynomial chaos-anova method," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3316–3325, 2019.

[91] K. Ye, J. Zhao, Y. Zhang, X. Liu, and H. Zhang, "A generalized computationally efficient copula-polynomial chaos framework for probabilistic power flow considering nonlinear correlations of pv injections," *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107727, 2022.

[92] J. Laowanitwattana and S. Uatrongjit, "Probabilistic power flow analysis based on partial least square and arbitrary polynomial chaos expansion," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1461–1470, 2022.

[93] H. Sheng and X. Wang, "Applying polynomial chaos expansion to assess probabilistic available delivery capability for distribution networks with renewables," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6726–6735, 2018.

[94] K. Panda, R. King, J. Maack, I. Satkauskas, and K. Potter, "Multi-fidelity stochastic economic dispatch for operating low carbon power grids," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2022.

[95] D. Hu and S. M. Ryan, "Stochastic vs. deterministic scheduling of a combined natural gas and power system with uncertain wind energy," *International Journal of Electrical Power & Energy Systems*, vol. 108, pp. 303–313, 2019.

[96] C.-L. Chen, T.-Y. Lee, and R.-M. Jan, "Optimal wind-thermal coordination dispatch in isolated power systems with large integration of wind capacity," *Energy conversion and management*, vol. 47, no. 18-19, pp. 3456–3472, 2006.

[97] R.-H. Liang and J.-H. Liao, "A fuzzy-optimization approach for generation scheduling with wind and solar energy systems," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1665–1674, 2007.

[98] Lorca and X. A. Sun, "Adaptive robust optimization with dynamic uncertainty sets for multiperiod economic dispatch under significant wind," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1702–1713, 2015.

[99] Y. Sasaki, N. Yorino, Y. Zoka, and F. I. Wahyudi, "Robust stochastic dynamic load dispatch against uncertainties," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5535–5542, 2018.

[100] B. Wang, P. Zhang, Y. He, X. Wang, and X. Zhang, "Scenario-oriented hybrid particle swarm optimization algorithm for robust economic dispatch of power system with wind power," *Journal of Systems Engineering and Electronics*, vol. 33, no. 5, pp. 1143–1150, 2022.

[101] Y. Sasaki, N. Yorino, Y. Zoka, and F. I. Wahyudi, "Robust stochastic dynamic load dispatch against uncertainties," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5535–5542, 2018.

[102] P. Li, M. Yang, and Q. Wu, "Confidence interval based distributionally robust real-time economic dispatch approach considering wind power accommodation risk," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 58–69, 2021.

[103] T. Ding, X. Zhang, R. Lu, M. Qu, M. Shahidehpour, Y. He, and T. Chen, "Multi-stage distributionally robust stochastic dual dynamic programming to multi-period economic dispatch with virtual energy storage," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 1, pp. 146–158, 2022.

[104] Y. Rychener, D. Kuhn, and T. Sutter, "End-to-end learning for stochastic optimization: A bayesian perspective," in *International Conference on Machine Learning*, pp. 29455–29472, PMLR, 2023.

[105] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," *IEEE TRANSACTIONS ON POWER SYSTEMS*, vol. 23, no. 2, 2008.

[106] P. A. Ruiz, C. R. Philbrick, E. Zak, K. W. Cheung, and P. W. Sauer, "Uncertainty management in the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 642–651, 2009.

[107] D. Ke, C. Y. Chung, and Y. Sun, "A novel probabilistic optimal power flow model with uncertain wind power generation described by customized gaussian mixture model," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 200–212, 2016.

[108] Y. Gu and L. Xie, "Stochastic look-ahead economic dispatch with variable generation resources," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 17–29, 2017.

[109] L. Yang, M. He, V. Vittal, and J. Zhang, "Stochastic optimization-based economic dispatch and interruptible load management with increased wind penetration," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 730–739, 2016.

[110] S. Feng, T. Ding, Z. Zhang, C. Mu, W. Jia, H. Deng, and Z. Zhou, "Parallel dual-dqam for multi-scenario stochastic economic dispatch model by temporal and scenario decompositions," *IEEE Transactions on Automation Science and Engineering*, 2023.

[111] X. Ye, Z. Lu, Y. Qiao, Y. Min, and M. O'Malley, "Identification and correction of outliers in wind farm time series power data," *IEEE Transactions on power systems*, vol. 31, no. 6, pp. 4197–4205, 2016.

[112] P. Algikar, Y. Xu, S. Yarahmadi, and L. Mili, "A robust data-driven process modeling applied to time-series stochastic power flow," *IEEE Transactions on Power Systems*, 2023.

[113] T. Senjyu, R. Sakamoto, N. Urasaki, T. Funabashi, H. Fujita, and H. Sekine, "Output power leveling of wind turbine generator for all operating regions by pitch angle control," *IEEE Transactions on Energy Conversion*, vol. 21, no. 2, pp. 467–475, 2006.

[114] J. Hu, L. Sun, X. Yuan, S. Wang, and Y. Chi, "Modeling of type 3 wind turbines with df/dt inertia control for system frequency response study," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2799–2809, 2017.

[115] C.-L. Nguyen and H.-H. Lee, "Optimization of wind power dispatch to minimize energy storage system capacity," *Journal of electrical engineering & technology*, vol. 9, no. 3, pp. 1080–1088, 2014.

[116] I. Zenelis, *Inter-area oscillations in power systems with uncertainties: data-driven wide-area damping control design and monitoring enhancement using energy storage*. McGill University (Canada), 2021.

[117] D. L. Yao, S. S. Choi, K. J. Tseng, and T. T. Lie, "Determination of short-term power dispatch schedule for a wind farm incorporated with dual-battery energy storage scheme," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 74–84, 2012.

[118] W. Wang, B. Yuan, Q. Sun, and R. Wennersten, "Application of energy storage in integrated energy systems—a solution to fluctuation and uncertainty of renewable energy," *Journal of Energy Storage*, vol. 52, p. 104812, 2022.

[119] H. Alsharif, M. Jalili, and K. N. Hasan, "Power system frequency stability using optimal sizing and placement of battery energy storage system under uncertainty," *Journal of Energy Storage*, vol. 50, p. 104610, 2022.

[120] N. Navid and G. Rosenwald, "Market solutions for managing ramp flexibility with high penetration of renewable resource," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 784–790, 2012.

[121] S. Nojavan, M. Majidi, and N. N. Esfetanaj, "An efficient cost-reliability optimization model for optimal siting and sizing of energy storage system in a microgrid in the presence of responsible load management," *Energy*, vol. 139, pp. 89–97, 2017.

[122] V. Jani and H. Abdi, "Optimal allocation of energy storage systems considering wind power uncertainty," *Journal of Energy Storage*, vol. 20, pp. 244–253, 2018.

[123] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Reliability engineering & system safety*, vol. 93, no. 7, pp. 964–979, 2008.

[124] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[125] K. N. Hasan, R. Preece, and J. V. Milanović, "Priority ranking of critical uncertainties affecting small-disturbance stability using sensitivity analysis techniques," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2629–2639, 2017.

[126] P. Ju, C. Qin, F. Wu, H. Xie, and Y. Ning, "Load modeling for wide area power system," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 4, pp. 909–917, 2011.

[127] Y. Jin, C. Lu, P. Ju, C. Rehtanz, F. Wu, and X. Pan, "Probabilistic preassessment method of parameter identification accuracy with an application to identify the drive train parameters of dfig," *IEEE transactions on power systems*, vol. 35, no. 3, pp. 1769–1782, 2019.

[128] G. Ökten and Y. Liu, "randomized quasi-monte carlo methods in global sensitivity analysis," *Reliability Engineering & System Safety*, vol. 210, p. 107520, 2021.

[129] K. Ye, J. Zhao, C. Huang, N. Duan, Y. Zhang, and T. E. Field, "A data-driven global sensitivity analysis framework for three-phase distribution system with PVs," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4809–4819, 2021.

[130] X. Xu, Z. Yan, M. Shahidehpour, H. Wang, and S. Chen, "Power system voltage stability evaluation considering renewable energy with correlated variabilities," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3236–3245, 2018.

[131] F. Ni, M. Nijhuis, P. H. Nguyen, and J. F. G. Cobben, "Variance-based global sensitivity analysis for power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1670–1682, 2018.

[132] X. Xu, Z. Yan, M. Shahidehpour, S. Chen, H. Wang, Z. Li, and Q. Zhou, "Maximum loadability of islanded microgrids with renewable energy generation," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 4696–4705, 2019.

[133] S. E. Tuller and A. C. Brett, "The goodness of fit of the weibull and rayleigh distributions to the distributions of observed wind speeds in a topographically diverse area," *Journal of climatology*, vol. 5, no. 1, pp. 79–94, 1985.

[134] G. Wang, H. Xin, D. Wu, P. Ju, and X. Jiang, "Data-driven arbitrary polynomial chaos-based probabilistic load flow considering correlated uncertainties," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3274–3276, 2019.

[135] Powertech Labs Inc., *Dynamic Security Assessment Software (DSATools): Voltage Security Assessment Tool (VSAT) User Manual*. Surrey, British Columbia, Canada, July 2022.

[136] M. Moustapha, C. Lataniotis, P. Wiederkehr, , P.-R. Wagner, D. Wicaksono, S. Marelli, and B. Sudret, "UQLib user manual," tech. rep., Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2022. Report UQLab-V2.0-201.

[137] S. Marelli and B. Sudret, "UQLab user manual – Polynomial chaos expansions," tech. rep., Chair of Risk, Safety Uncertainty Quantification, ETH Zurich, 2018. Report UQLab-V1.1-104.

[138] S. Marelli, C. Lamas, K. Konakli, C. Mylonas, P. Wiederkehr, and B. Sudret, "UQLab user manual – Sensitivity analysis," tech. rep., Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019. Report # UQLab-V1.3-106.

[139] N. Wiener, "The homogeneous chaos," *American Journal of Mathematics*, vol. 60, no. 4, pp. 897–936, 1938.

[140] S. Oladyshkin and W. Nowak, "Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion," *Reliability Engineering & System Safety*, vol. 106, pp. 179–190, 2012.

[141] J. Li, N. Ou, G. Lin, and W. Wei, "Compressive sensing based stochastic economic dispatch with high penetration renewables," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1438–1449, 2019.

[142] Y. Xu, L. Mili, A. Sandu, M. R. v. Spakovsky, and J. Zhao, "Propagating uncertainty in power system dynamic simulations using polynomial chaos," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 338–348, 2019.

[143] J. Liu, X. Wang, and X. Wang, "A sparse polynomial chaos expansion-based method for probabilistic transient stability assessment and enhancement," in *2022 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2022.

[144] B. Sudret, "Uncertainty propagation and sensitivity analysis in mechanical models–contributions to structural reliability and stochastic spectral methods," *Habilitationa diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, France*, vol. 147, p. 53, 2007.

[145] C. Soize and R. Ghanem, "Physical systems with random uncertainties: chaos representations with arbitrary probability measure," *SIAM Journal on Scientific Computing*, vol. 26, no. 2, pp. 395–410, 2004.

[146] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann, "On the convergence of generalized polynomial chaos expansions," *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 46, no. 2, pp. 317–339, 2012.

[147] R. H. Cameron and W. T. Martin, "The orthogonal development of non-linear functionals in series of fourier-hermite functionals," *Annals of Mathematics*, pp. 385–392, 1947.

[148] J. M. Morales, A. J. Conejo, and J. Perez-Ruiz, "Simulating the impact of wind production on locational marginal prices," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 820–828, 2010.

[149] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, pp. 559–572, 1901.

[150] K. Ye, J. Zhao, N. Duan, and D. A. Maldonado, "Stochastic power system dynamic simulation and stability assessment considering dynamics from correlated loads and pvs," *IEEE Transactions on Industry Applications*, vol. 58, no. 6, pp. 7764–7775, 2022.

[151] Torre, Emiliano and Marelli, Stefano and Embrechts, Paul and Sudret, Bruno, "Data-driven polynomial chaos expansion for machine learning regression," *J. Comput. Phys.*, vol. 388, pp. 601–623, 2019.

[152] F. E. R. Commission *et al.*, "Order no. 889 final rule (issued 24 april 1996). 75 ferc 61,078. open access same-time information system (formerly real-time information networks) and standards of conduct, docket no," tech. rep., RM95-9-000, 1996.

[153] W. Li, E. Vaahedi, and Z. Lin, "Bc hydro's transmission reliability margin assessment in total transfer capability calculations," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4796–4802, 2013.

[154] O. Onohaebi, "Power outage in the nigeria transmission grid," *Research Journal of Applied Sciences*, vol. 4, no. 1, pp. 1–9, 2009.

[155] H. Sheng and X. Wang, "A non-intrusive low-rank approximation method for assessing the probabilistic available transfer capability," *arXiv preprint arXiv:1810.08156*, 2018.

[156] K. Audomvongseree and A. Yokoyama, "Consideration of an appropriate ttc by probabilistic approach," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 375–383, 2004.

[157] G. Hamoud, "Assessment of available transfer capability of transmission systems," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 27–32, 2000.

[158] R. D. Christie, B. F. Wollenberg, and I. Wangensteen, "Transmission management in the deregulated environment," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 170–195, 2000.

[159] G. C. Ejebe, J. G. Waight, M. Sanots-Nieto, and W. F. Tinney, "Fast calculation of linear available transfer capability," *IEEE Transactions on Power Systems*, vol. 15, no. 3, pp. 1112–1116, 2000.

[160] M. Khosravifard and M. Shaaban, "Risk-based available transfer capability assessment including nondispatchable wind generation," *International Transactions on Electrical Energy Systems*, vol. 25, no. 11, pp. 3169–3183, 2015.

[161] F. W. Mohn and A. Z. De Souza, "Tracing pv and qv curves with the help of a cric continuation method," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1115–1122, 2006.

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

[163] L. Min and A. Abur, "Total transfer capability computation for multi-area power systems," *IEEE Transactions on power systems*, vol. 21, no. 3, pp. 1141–1147, 2006.

[164] Y. Ou and C. Singh, "Improvement of total transfer capability using tcsc and svc," in *2001 Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No. 01CH37262)*, vol. 2, pp. 944–948, IEEE, 2001.

[165] Y. Ou and C. Singh, "Assessment of available transfer capability and margins," *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 463–468, 2002.

[166] A. B. Rodrigues and M. G. Da Silva, "Probabilistic assessment of available transfer capability based on monte carlo method with sequential simulation," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 484–492, 2007.

[167] X. Zhang, Z. Bie, and G. Li, "Reliability assessment of distribution networks with distributed generations using monte carlo method," *Energy Procedia*, vol. 12, pp. 278–286, 2011.

[168] M. Hajian, W. D. Rosehart, and H. Zareipour, "Probabilistic power flow by monte carlo simulation with latin supercube sampling," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1550–1559, 2012.

[169] M. Ramezani, C. Singh, and M. Haghifam, "Role of clustering in the probabilistic evaluation of ttc in power systems including wind power generation," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 849–858, 2009.

[170] R. Chang, C. Tsai, C. Su, and C. Lu, "Method for computing probability distributions of available transfer capability," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 149, no. 4, pp. 427–431, 2002.

[171] K. Wang, G. Li, and X. Jiang, "Applying probabilistic collocation method to power flow analysis in networks with wind farms," in *2013 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2013.

[172] E. Haesen, C. Bastiaensen, J. Driesen, and R. Belmans, "A probabilistic formulation of load margins in power systems with stochastic generation," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 951–958, 2009.

[173] H. Sheng and X. Wang, "A non-intrusive low-rank approximation method for assessing the probabilistic available transfer capability," *arXiv preprint arXiv:1810.08156*, 2018.

[174] Y. Liu, J. Zhao, L. Xu, T. Liu, G. Qiu, and J. Liu, "Online ttc estimation using nonparametric analytics considering wind power integration," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 494–505, 2018.

[175] Y. Atwa, E. El-Saadany, M. Salama, and R. Seethapathy, "Optimal renewable resources mix for distribution system energy loss minimization," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 360–370, 2009.

[176] "Nrel western wind data set." `https://www.nrel.gov/grid/western-wind-data.html`.

[177] X. Kou and F. Li, "Interval optimization for available transfer capability evaluation considering wind power uncertainty," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 250–259, 2018.

[178] X. Zhang, J. J. Thomas, and S. Grijalva, "Fast evaluation of probabilistic total transfer capability considering multiple wind farms," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pp. 1–5, IEEE, 2016.

[179] J. Yan, C. Li, and Y. Liu, "Deep learning based total transfer capability calculation model," in *2018 International Conference on Power System Technology (POWERCON)*, pp. 952–957, IEEE, 2018.

[180] G. Wang, H. Xin, D. Wu, P. Ju, and X. Jiang, "Data-driven arbitrary polynomial chaos-based probabilistic load flow considering correlated uncertainties," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3274–3276, 2019.

[181] A. Atputharajah and T. K. Saha, "Power system blackouts-literature review," in *2009 International Conference on Industrial and Information Systems (ICIIS)*, pp. 460–465, IEEE, 2009.

[182] H. Farahmand, M. Rashidi-Nejad, and M. Fotuhi-Firoozabad, "Implementation of facts devices for atc enhancement using rpf technique," in *2004 Large Engineering Systems Conference on Power Engineering (IEEE Cat. No.04EX819)*, pp. 30–35, 2004.

[183] K. Iba, H. Suzuki, M. Egawa, and T. Watanabe, "Calculation of critical loading condition with nose curve using homotopy continuation method," *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 584–593, 1991.

[184] North American Electric Reliability Council (NERC), "Available Transfer Capability: Definitions, Determination, and Utilization - Final Report," tech. rep., North American Electric Reliability Council, June 1996.

[185] California Independent System Operator (CAISO), "Transmission Reliability Margin Implementation Document," 2016. [Online; accessed 9-May-2023].

[186] National Renewable Energy Laboratory, "Nrel western wind data set." `https://www.nrel.gov/grid/western-wind-data.html`, 2023. Accessed on 10 May 2023.

[187] M. Aien, M. Fotuhi-Firuzabad, and M. Rashidinejad, "Probabilistic optimal power flow in correlated hybrid wind–photovoltaic power systems," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 130–138, 2014.

[188] E. Camm, M. R. Behnke, O. Bolado, M. Bollen, M. Bradt, C. Brooks, W. Dilling, M. Edds, W. J. Hejdak, D. Houseman, S. Klein, F. Li, J. Li, P. Maibach, T. Nicolai, J. Patino, S. V. Pasupulati, N. Samaan, S. Saylors, T. Siebert, T. Smith, M. Starke, and R. Walling, "Characteristics of wind turbine generators for wind power plants," in *2009 IEEE Power Energy Society General Meeting*, pp. 1–5, 2009.

[189] National Renewable Energy Laboratory, "Solar geospatial data & tools." `https://www.nrel.gov/gis/solar-geospatial-data-tools.html`, Accessed 2023.

[190] Y. M. Atwa, E. F. El-Saadany, M. M. A. Salama, and R. Seethapathy, "Optimal renewable resources mix for distribution system energy loss minimization," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 360–370, 2010.

[191] W. E. C. C. (WECC), "Wecc pv plant power flow modeling guidelines," tech. rep., August 2010.

[192] U. E. I. Administration, "Short-term energy outlook." EIA Website, 9 2021. [Accessed: May 11, 2023].

[193] A. Douglas, A. Breipohl, F. Lee, and R. Adapa, "Risk due to load forecast uncertainty in short term power system planning," *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1493–1499, 1998.

[194] R. Billinton and D. Huang, "Effects of load forecast uncertainty on bulk electric system reliability evaluation," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 418–425, 2008.

[195] K. Bae and J. S. Thorp, "A stochastic study of hidden failures in power system protection," *Decision Support Systems*, vol. 24, no. 3-4, pp. 259–268, 1999.

[196] A. Ilin and T. Raiko, "Practical approaches to principal component analysis in the presence of missing values," *The Journal of Machine Learning Research*, vol. 11, pp. 1957–2000, 2010.

[197] R. Ahlfeld, F. Montomoli, E. Scalas, and S. Shahpar, "Uncertainty quantification for fat-tailed probability distributions in aircraft engine simulations," *Journal of Propulsion and Power*, vol. 33, no. 4, pp. 881–890, 2017.

[198] S. Marelli and B. Sudret, "Uqlab user manual–polynomial chaos expansions," *Chair of risk, safety & uncertainty quantification, ETH Zürich, UQLab-V1.1-104*.

[199] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," 2004.

[200] S. Marelli, N. Lüthen, and B. Sudret, "UQLab user manual – Polynomial chaos expansions," tech. rep., Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2022. Report UQLab-V2.0-104.

[201] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[202] C. Josz, S. Fliscounakis, J. Maeght, and P. Panciatici, "Ac power flow data in matpower and qcqp format: itesla, rte snapshots, and pegase," *arXiv preprint arXiv:1603.01533*, 2016.

[203] S. Fliscounakis, P. Panciatici, F. Capitanescu, and L. Wehenkel, "Contingency ranking with respect to overloads in very large power systems taking into account uncertainty, preventive, and corrective actions," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4909–4917, 2013.

[204] J. Cheng, R. L.-Y. Chen, H. N. Najm, A. Pinar, C. Safta, and J.-P. Watson, "Chance-constrained economic dispatch with renewable energy and storage," *Computational Optimization and Applications*, vol. 70, no. 2, pp. 479–502, 2018.

[205] T. Summers, J. Warrington, M. Morari, and J. Lygeros, "Stochastic optimal power flow based on convex approximations of chance constraints," in *2014 Power Systems Computation Conference*, pp. 1–7, IEEE, 2014.

[206] R. A. Jabr, S. Karaki, and J. A. Korbane, "Robust multi-period opf with storage and renewables," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2790–2799, 2014.

[207] Z. Hu, Y. Xu, M. Korkali, X. Chen, L. Mili, and J. Valinejad, "A bayesian approach for estimating uncertainty in stochastic economic dispatch considering wind power penetration," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 671–681, 2021.

[208] M. M. Rajabi, "Review and comparison of two meta-model-based uncertainty propagation analysis methods in groundwater applications: polynomial chaos expansion and gaussian process emulation," *Stochastic environmental research and risk assessment*, vol. 33, pp. 607–631, 2019.

[209] R.-P. Liu, Y. Hou, Y. Li, S. Lei, W. Wei, and X. Wang, "Sample robust scheduling of electricity-gas systems under wind power uncertainty," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5889–5900, 2021.

[210] S. Wang, S. Shahidehpour, D. Kirschen, S. Mokhtari, and G. Irisarri, "Short-term generation scheduling with transmission and environmental constraints using an augmented lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1294–1301, 1995.

[211] E. Torre, S. Marelli, P. Embrechts, and B. Sudret, "A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas," *Probabilistic Engineering Mechanics*, vol. 55, pp. 1–16, 2019.

[212] "National Renewable Energy Laboratory. western wind data set." `https://www.nrel.gov/grid/western-wind-data.html`. Accessed: 2021-01-18.

[213] X. Wang, X. Wang, H. Sheng, and X. Lin, "A data-driven sparse polynomial chaos expansion method to assess probabilistic total transfer capability for power systems with renewables," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2573–2583, 2020.

[214] M. D. Morris, "Factorial sampling plans for preliminary computational experiments," *Technometrics*, vol. 33, no. 2, pp. 161–174, 1991.

[215] I. M. Sobol, "Sensitivity estimates for nonlinear mathematical models," *Mathematical Modelling and Computational Experiments*, vol. 1, no. 4, pp. 407–414, 1993.

[216] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf, "Global sensitivity analysis for systems with independent and/or correlated inputs," *The journal of physical chemistry A*, vol. 114, no. 19, pp. 6022–6032, 2010.

[217] S. Kucherenko, S. Tarantola, and P. Annoni, "Estimation of global sensitivity indices for models with dependent variables," *Computer physics communications*, vol. 183, no. 4, pp. 937–946, 2012.

[218] S. Liu, D. Jin, Z. Ma, X. Wei, and L. Zhang, "survey of morris and e-fast algorithms based on power-generation operation and assistant decision model," in *2015 8th International Conference on Ubi-Media Computing (UMEDIA)*, pp. 344–349, IEEE, 2015.

[219] T. A. Mara, "extension of the rbd-fast method to the computation of global sensitivity indices," *Reliability Engineering & System Safety*, vol. 94, no. 8, pp. 1274–1281, 2009.

[220] Y. Caniou, *Global sensitivity analysis for nested and multiscale modelling*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II, 2012.

[221] Z. Liu and Y. Choe, "Data-driven sensitivity indices for models with dependent inputs using polynomial chaos expansions," *Structural Safety*, vol. 88, p. 101984, 2021.

[222] T. A. Mara and W. E. Becker, "polynomial chaos expansion for sensitivity analysis of model output with dependent inputs," *Reliability Engineering and System Safety*, vol. 214, p. 107795, 2021.

[223] Y. Caniou, *Global sensitivity analysis for nested and multiscale modelling*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II, 2012.

[224] C. Xu and G. Z. Gertner, "Uncertainty and sensitivity analysis for models with correlated parameters," *Reliability Engineering & System Safety*, vol. 93, no. 10, pp. 1563–1573, 2008.

[225] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and computers in simulation*, vol. 55, no. 1-3, pp. 271–280, 2001.

[226] D. A. Bechrakis and P. D. Sparis, "Correlation of wind speed between neighboring measuring stations," *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, pp. 400–406, 2004.

[227] J. M. Morales, R. Minguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843–855, 2010.

[228] C. Ma and M. Iqbal, "Statistical comparison of solar radiation correlations monthly average global and diffuse radiation on horizontal surfaces," *Solar Energy*, vol. 33, no. 2, pp. 143–148, 1984.

[229] P. Wiederkehr, "Global sensitivity analysis with dependent inputs," tech. rep., Chair of Risk, Safety Uncertainty Quantification, ETH Zurich, 2018. R-MTWiederkehr-001.

[230] Z. Wang, Y. Zhou, Q. Guo, and H. Sun, "The total transfer capability assessment of transmission interfaces combining causal inference and multi-task learning," *IEEE Transactions on Power Systems*, pp. 1–13, 2023.

[231] I. Zenelis, G. Pierrou, and X. Wang, "A data-driven energy storage system-based algorithm for monitoring the small-signal stability of power grids with volatile wind power," in *2021 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2021.

[232] X. Kong, C. Hu, and Z. Duan, *Principal component analysis networks and algorithms*. Springer, 2017.

[233] N. H. Timm, *Applied multivariate analysis*. Springer, 2002.

[234] Y. Noh, K. Choi, and L. Du, "Reliability-based design optimization of problems with correlated input variables using a gaussian copula," *Structural and multidisciplinary optimization*, vol. 38, pp. 1–16, 2009.

[235] S. Wang, S. Shahidehpour, D. S. Kirschen, S. Mokhtari, and G. Irisarri, "Short-term generation scheduling with transmission and environmental constraints using an augmented lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1294–1301, 1995.