

Uncertainty Modeling and Optimization in Smart Grid with Renewable Generation

by

Zhichao Shi

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Energy Systems

Department of Electrical and Computer Engineering
University of Alberta

©Zhichao Shi, 2020

Abstract

Electricity is an essential part of our daily life which can be supplied by power systems with fossil fuels or renewable energy sources. Nowadays, traditional power systems are evolving towards new smart grid with the development of advanced information and communication technology. Compared with the general electrical grid, smart grid features a two-way flow of electricity and information as a distributed energy delivery network. With lots of benefits such as increasing system efficiency, robustness and reducing outage costs, smart grid has attracted worldwide attention in recent years both in industry and academia.

Conventional electricity generation method mainly relies on the combustion of fossil fuels. However, considering the limited supply of fossil fuels and increasing environmental pollution, it is necessary and urgent to enhance the use of alternative energy sources. Hence, the clean renewable energy sources such as wind and solar have drew more public attention recently. Although renewable energy has many advantages, the main drawback is the intermittent and random characteristic. With the increasing integration of renewable generation in smart grid, many new technical challenges have also emerged in regard to the reliable operation of power systems, especially the uncertainty related problems. Therefore, it is imperative to handle the uncertainties in smart grid to achieve reliable and stable operation.

The focus of this research is to study uncertainty modeling and related optimization problems in smart grid. There are various uncertain sources in smart grid such as renewable generation, load demand and electricity price, among which the uncertain renewable generation has attracted more attentions in recent years. In this work, we focus on the uncertainties caused by renewable generation such as wind power generation.

In order to deal with the possible uncertainties in system operation, different approaches have been studied including the direct forecast methods and some mathematical modeling methods. Although point forecast methods have been widely studied for wind energy, the forecast errors cannot be fully eliminated which may bring

significant influence power system operational decision. In this thesis, probabilistic forecast is investigated, and a recurrent neural network (RNN) based interval prediction model is first proposed to forecast uncertain wind power which can generate intervals with a predefined confidence level. As the source of wind power, wind speed forecast is also investigated in a multiobjective interval prediction framework.

Furthermore, microgrid is an important part of future smart grid, and microgrid energy management has been a popular topic for a long time. To capture the uncertainties of wind power in microgrid, mathematical modeling methods based on distributionally robust optimization (DRO) and robust optimization (RO) are investigated for energy management problems in this thesis. First, a distributionally robust chance-constrained energy management model is proposed for islanded microgrids, which considers the possible power imbalance due to uncertain wind power, and a novel moment based ambiguity set is designed. The chance constraint for power balance is processed with DRO technique and the problem is reformulated as a tractable second-order conic programming (SOCP) problem. The effectiveness of the proposed approach has been validated by case studies.

Based on the research of microgrid energy management, the uncertainty modeling in transmission system is also studied in this thesis. Particularly, two-stage chance-constrained unit commitment (UC) problem and energy and reserve dispatch problem are studied with DRO method. The statistical-distance based ambiguity set is proposed to describe the uncertain distribution of wind power in such problems.

To overcome the anticipativity of uncertainty in single-stage or two-stage models, multistage energy management problem is investigated for grid-connected microgrids which takes the non-anticipativity into account. The uncertainty modeling methods based on RO and DRO techniques are analyzed for wind power with interval uncertainty set or second-order conic representable ambiguity set, respectively. The effectiveness of the proposed multistage model is verified by case studies.

Preface

The contents presented in this thesis are based on original work by Zhichao Shi. As detailed in the following, material from some chapters of this thesis has been published as journal papers, and in conference proceedings under the supervision of Dr. Hao Liang and Dr. Venkata Dinavahi in concept formation and by providing comments and corrections to the article manuscript.

Chapter 2 includes the results published in the following papers:

- Z. Shi, H. Liang and V. Dinavahi, "Direct interval forecast of uncertain wind power based on recurrent neural networks", *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1177-1187, July 2018.
- Z. Shi, H. Liang and V. Dinavahi, "Wavelet neural network based multiobjective interval prediction for short-term wind speed", *IEEE Access*, vol. 6, pp. 63352-63365, 2018.

Chapter 3 includes the results published in the following paper:

- Z. Shi, H. Liang, S. Huang and V. Dinavahi, "Distributionally robust chance-constrained energy management for islanded microgrids", *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2234-2244, Mar. 2019.

The contents presented in Chapter 4 have been published or accepted as follows:

- Z. Shi, H. Liang and V. Dinavahi, "Data-driven distributionally robust chance-constrained unit commitment with uncertain wind power", *IEEE Access*, vol. 7, pp. 135087-135098, 2019.
- Z. Shi, H. Liang and V. Dinavahi, "Energy and reserve dispatch with renewable generation using data-driven distributionally robust optimization", *North American Power Symposium*, pp. 1-6, Oct. 2019.

Chapter 5 includes the results in the following papers that have been published or submitted:

- Z. Shi, H. Liang, S. Huang and V. Dinavahi, "Multistage robust energy management for microgrids considering uncertainty", *IET Gener. Transm. Distrib.*, vol. 13, no. 10, pp. 1906-1913, June 2019.
- Z. Shi, H. Liang and V. Dinavahi, "Multi-period energy management for CCHP microgrids with distributionally robust optimization", submitted to *IEEE Trans. Sustain. Energy*, under review.

Acknowledgements

I would like to express my deepest appreciation to my supervisors *Dr. Hao Liang* and *Dr. Venkata Dinavahi* for their supportive attitude, encouragement, and guidance through my research at the University of Alberta. Undoubtedly, without their constant help and supervision, this dissertation would not have been possible.

It is an honor for me to extend my gratitude to all my Ph.D. committee members for reviewing my thesis and providing thoughtful comments to improve it. I also thank all my colleagues in my lab and friends I met at the University of Alberta during my Ph.D. program, specifically *Dr. Shengjun Huang* and *Dr. Ning Lin*, who helped me a lot at the beginning of my research.

In the meantime, I would like to acknowledge *Dr. Tao Zhang* and *Dr. Rui Wang* from the National University of Defense Technology who led me to my research path at the very beginning. Their encouragement and support also helped me achieve my dream of studying abroad to a great extent.

In addition, I want to express my great gratitude to my parents and my girlfriend for their endless love and unconditional support. They are always my source of strength to move forward.

This thesis work was supported by the China Scholarship Council. I greatly appreciate the financial support.

Table of Contents

1	Introduction	1
1.1	Background	1
1.2	Research Definitions	4
1.2.1	Uncertainty Modeling Based on Forecast	4
1.2.2	Neural Network Model	5
1.2.3	Microgrid Energy Management	5
1.2.4	Distributionally Robust Optimization	6
1.2.5	Unit Commitment	6
1.2.6	Energy and Reserve Dispatch	7
1.3	Literature Review	7
1.3.1	Wind Power and Wind Speed Forecast	7
1.3.2	Microgrid Energy Management with Uncertainty	10
1.3.3	Unit Commitment	11
1.3.4	Energy and Reserve Dispatch	13
1.3.5	Multistage Energy Management for Microgrids	14
1.3.6	CCHP Microgrid	15
1.4	Thesis Motivation and Objectives	16
1.5	Thesis Outline	19
2	Direct Interval Prediction for Wind Power and Wind Speed	24
2.1	Introduction	25
2.2	Interval Prediction for Wind Power	25
2.2.1	Background of PI	26
2.2.1.1	PI Evaluation Indices	26
2.2.1.2	LUBE Method	28
2.2.2	RNN-based interval prediction model	28
2.2.2.1	Elman Network	29
2.2.2.2	NARX Network	30
2.2.2.3	Dragonfly Algorithm	30
2.2.2.4	Model Implementation	31
2.2.3	Realistic Wind Energy Prediction Case Study	34
2.2.3.1	Data Set	34

2.2.3.2	Parameter Settings	34
2.2.3.3	Test Results	35
2.2.3.4	Comparison with benchmark models	39
2.2.3.5	Discussion	41
2.3	Multiobjective Interval Prediction for Wind Speed	43
2.3.1	Problem Formulation	43
2.3.1.1	PI and Evaluation Indices	43
2.3.1.2	Multiobjective Problem Formulation	44
2.3.2	Solution Methodology	45
2.3.2.1	Wavelet Neural Network Based Prediction Model	45
2.3.2.2	PICEA-g Algorithm	47
2.3.2.3	Implementation Strategy	49
2.3.3	Numerical Results	51
2.3.3.1	Datasets	51
2.3.3.2	Parameter Settings	52
2.3.3.3	Prediction Results	54
2.3.3.4	Comparison with Other Models	55
2.3.3.5	Discussion	60
2.4	Summary	61
3	Chance-constrained Energy Management for Islanded Microgrids	63
3.1	Introduction	63
3.2	Problem Formulation	64
3.2.1	Distributed Generation	64
3.2.2	Energy Storage System	65
3.2.3	Load Demand	66
3.2.4	Chance Constraint for Power Balance	66
3.3	Solution Methodology	67
3.3.1	An Ambiguity Set for Wind Power Output	67
3.3.2	Problem Reformulation	69
3.3.3	Including Unimodality Information	71
3.4	Case Study	72
3.4.1	Description of Microgrid	73
3.4.2	Simulation Results	74
3.4.3	Discussion	75
3.5	Summary	79
4	Data-driven Distributionally Robust UC and ED	81
4.1	Introduction	81
4.2	Data-driven Distributionally Robust Chance-constrained UC	82

4.2.1	Problem Formulation	82
4.2.1.1	UC Mathematical Model	83
4.2.1.2	Ambiguity Set Construction	85
4.2.2	Solution Methodology	87
4.2.2.1	Problem Reformulation	87
4.2.2.2	C&CG Decomposition Algorithm	89
4.2.3	Numerical Results	91
4.2.3.1	IEEE 6-bus Test System	92
4.2.3.2	IEEE 118-bus Test System	95
4.2.3.3	Practical-scale Power System	96
4.3	Energy and Reserve Dispatch Using Data-driven DRO	98
4.3.1	Problem Formulation	99
4.3.1.1	Two-stage Dispatch Model	99
4.3.1.2	Ambiguity Set Construction	101
4.3.2	Solution Methodology	102
4.3.3	Case Studies	103
4.3.3.1	IEEE 6-bus System	103
4.3.3.2	IEEE 118-bus System	105
4.4	Summary	107
5	Multistage Energy Management for Microgrids	109
5.1	Introduction	109
5.2	Multistage Robust Energy Management for Microgrids	110
5.2.1	Problem Formulation	110
5.2.1.1	System Model	111
5.2.1.2	Multistage Robust Optimization Problem	114
5.2.2	Solution Methodology	115
5.2.2.1	Upper Bound Problem	115
5.2.2.2	Lower Bound Problem	117
5.2.2.3	RDDP Method	117
5.2.3	Case Study	119
5.2.3.1	Microgrid Description and Data	119
5.2.3.2	Simulation Results	120
5.2.3.3	Comparison with Other Methods	122
5.3	Multi-period Distributionally Robust Energy Management for CCHP Microgrids	126
5.3.1	Problem Formulation	126
5.3.1.1	CCHP Units	127
5.3.1.2	Storage Systems	128
5.3.1.3	Load Balance and Objective	129
5.3.1.4	Multi-period Formulation	130

5.3.2	Solution Methodology	131
5.3.2.1	Ambiguity Set for Wind Power	131
5.3.2.2	Linear Decision Rule	132
5.3.2.3	Problem Reformulation	133
5.3.3	Case Studies	135
5.3.3.1	Data and Parameter Settings	135
5.3.3.2	Simulation Results	136
5.3.3.3	Comparison with Other Methods	137
5.4	Summary	138
6	Conclusions and Future Work	140
6.1	Contributions of Thesis	141
6.2	Directions for Future Work	143
	Bibliography	145
	Appendix A Dual Problem and Change of Variables	163
	Appendix B Dual of SOCP Problem	164

List of Tables

2.2	Cross-validation results of Elman model	35
2.3	PI construction results of Elman model	38
2.4	PI construction results of NARX model	38
2.5	PI construction results for different seasons	39
2.6	PI construction results of MLP model	39
2.7	Comparison of proposed model with benchmark models	41
2.8	PI construction results with different algorithms	41
2.9	Descriptive statistics of the two datasets	52
2.10	Parameters of PICEA-g algorithm	52
2.11	Comparison results for Victoria data	58
2.12	Comparison results for Edmonton data	59
2.13	Multiobjective comparison results for different models	60
2.14	Training results for different population size	60
2.15	Hypervolume results for different feature selection methods	61
3.2	Parameters of conventional generators	73
3.3	Parameters of ESS	73
3.4	Comparison of system cost with different methods	79
4.2	Influence of data size on system cost	93
4.3	Influence of confidence level on system cost	94
4.4	Influence of data size for 118-bus system	96
4.5	Influence of confidence level for 118-bus system	97
4.6	Effects of data size for 6-bus system	104
4.7	Effects of data size for 118-bus system	106
5.2	Parameters for the generators	120
5.3	Detailed cost comparison for different methods	126
5.4	Parameters of MTs	136
5.5	Main parameter values	137
5.6	UC results of MTs	137
5.7	Comparison results	138

List of Figures

1.1	Smart grid conceptual model.	2
1.2	The growth of installed wind power in Canada.	3
1.3	Outline of this thesis.	20
2.1	Schematic diagram of Elman network.	29
2.2	Model flowchart for PI construction.	32
2.3	NCWC of the best individual in the training process.	36
2.4	Optimal PIs of Elman model for test data.	37
2.5	Optimal PIs of Elman model for the last week of test data.	37
2.6	Optimal PIs of NARX and MLP model.	38
2.7	Architecture of the proposed WNN model.	46
2.8	PICEA-g implementation framework.	49
2.9	ACF and PACF analysis of Victoria wind speed.	53
2.10	Average hypervolume results.	54
2.11	Pareto front of prediction results.	55
2.12	Interval prediction results for Victoria data.	56
2.13	Interval prediction results for Edmonton data.	56
3.1	Wind power and critical load.	74
3.2	Optimal power schedule of conventional generators.	75
3.3	Optimal energy schedule of ESS and deferrable load.	76
3.4	System cost with different ϵ	77
3.5	System cost with different interval size.	78
4.1	Structure of IEEE 6-bus test system.	92
4.2	Mean wind power output.	93
4.3	Comparison with other methods for 6-bus system.	95
4.4	Out-of-sample assessment result for 6-bus system.	96
4.5	Comparison with other methods for 118-bus system.	97
4.6	Out-of-sample assessment result for 118-bus system.	98
4.7	Comparison with other methods for 319-bus system.	98
4.8	Out-of-sample assessment result for 319-bus system.	99
4.9	Influence of confidence level for 6-bus system.	105

4.10	Influence of confidence level for 118-bus system.	106
4.11	Influence of data size under different distributions.	107
5.1	Microgrid system model.	111
5.2	Schematic diagram of solution process.	119
5.3	Wind power and load data.	121
5.4	Electricity exchange price.	122
5.5	Evolution of lower and upper bounds.	123
5.6	Power generation level of DG units.	124
5.7	Schedule of storage and power exchange.	125
5.8	The total cost of different methods.	125
5.9	The scheme of a grid-connected CCHP microgrid.	127
5.10	Data profiles.	136
5.11	Output of recourse variables with LDR.	138

Nomenclature

Chapter 2

a, c, e, f, s, w	Weight coefficients in DA algorithm.
a_j, b_j	Scale and shift parameters of wavelets in WNN model.
c_1, c_2	Random numbers in Levy flight equation.
d_o, d_u	The output and input delay in NARX network.
$iter$	The iteration number.
L_i	The lower bound of the i th PI.
m	The embedding dimension in delay embedding theorem.
N	The number of test samples.
N_{pop}	The population size of candidate solutions.
o	The output of NARX network.
P	Position vector in DA algorithm.
u	The input of NARX network.
U_i	The upper bound the i th PI.
V	Step vector in DA algorithm.
w_{ij}	Weight parameters in WNN model.
W_1, W_2, W_3	Weight vectors in Elman network.
x	The output of the hidden layer in Elman network.
x_c	The output of the output layer in Elman network.
y_i	The real target of test data sample i .
z_1	The output of the output layer in Elman network.
α	The parameter related to the nominal confidence level.
δ_i	A binary value corresponding to the i th PI.
η, μ, γ	Controlling parameters and boolean function in CWC index.
λ, ρ	Parameters in Levy flight equation.
τ	Time delay in delay embedding theorem.
θ_i	The width of the i th PI.

Chapter 3

a_i, b_i, c_i	Coefficients of quadratic cost function of generator i .
c^{emis}	Emission cost coefficient.
$CE_{i,t}$	Emission cost of generator i at time t .
$CG_{i,t}$	Generation cost of generator i at time t .
$CS_{j,t}$	Degradation cost of ESS j at time t .
d_i, e_i, f_i	Coefficients of quadratic emission function of generator i .

E_j^{min}, E_j^{max}	Minimum/maximum energy storage of ESS j .
$E_{j,t}$	Energy storage of ESS j at time t .
$h^0(x_t)$	Auxiliary functions of x_t .
$h(x_t)$	Auxiliary coefficient functions of x_t .
L_t	Critical load at time t .
N_{dg}	Number of conventional generators.
N_{ess}	Number of energy storage systems.
N_{def}	Number of deferrable loads.
P_i^{min}, P_i^{max}	Minimum/maximum output power of generator i .
$P_{i,t}$	The output power of generator i at time t .
$P_{j,t}^{ch}$	Charging power of ESS j at time t .
$P_j^{ch,max}$	Maximum charging power of ESS j .
$P_{j,t}^{dch}$	Discharging power of ESS j at time t .
$P_j^{dch,max}$	Maximum discharging power of ESS j .
P_k^{def}	Total demand of deferrable load k over the specified time interval.
$P_k^{def,min}, P_k^{def,max}$	Minimum/maximum serving rate for deferrable load k .
$P_{k,t}^{def}$	Demand of deferrable load k at time t .
R_i^{up}, R_i^{dn}	Ramp-up/ramp-down limits of generator i .
T_k^a, T_k^b	Start time and end time to serve deferrable load k .
w_t	Uncertain wind power at time t .
x_t	General representation of decision variables.
\mathbf{x}	Set of decision variables.
ϵ	Violation probability parameter in the chance constraint.
$\eta_j^{ch}, \eta_j^{dch}$	Charging/discharging efficiency of ESS j .
$\mathbb{E}_{\mathbb{P}_t}$	Expectation under distribution \mathbb{P}_t .
\mathbb{P}_t	Distribution of random variable w_t .
\mathcal{P}_t^0	The set of all the probability distributions of random variables.
$\mathcal{P}_t^1, \mathcal{P}_t^2$	The ambiguity set of the probability distribution of wind power.
μ_t	Mean value of random variable w_t .
$\underline{\mu}_t, \bar{\mu}_t$	Lower and upper bounds of μ_t .
σ_t^2	Variance of random variable w_t .
$\underline{\sigma}_t^2, \bar{\sigma}_t^2$	Lower and upper bounds of σ_t^2 .
β, v, τ, z, s	Auxiliary variables in second-order conic constraint formulation.
η_1, η_2, r, ξ	Auxiliary variables in second-order conic constraint formulation.

Chapter 4

a_i, e_i, c_i	Coefficients of quadratic cost function of generator i .
a_i^m, b_i^m	Piecewise linear cost coefficients for generator i .
b/\mathcal{B}	Index/set of buses.
c_i^u, c_i^d	Up and down re-dispatch cost coefficients for unit i .
c_j^w	Wind power curtailment cost coefficient.
c_k^{Lc}	Load shedding cost coefficient.
d_i^u, d_i^d	Up and down reserve cost coefficients.
d_t^b	Load demand at bus b at time t .
$d_t^{ls,b}$	Load shedding of load bus b at time t .

$\mathcal{D}/\mathcal{D}_1/\mathcal{D}_\infty$	Distance based ambiguity set.
f_0^s/f^s	Reference/true probability for scenario s .
f_u^s/f_l^s	Upper/lower bound of the unknown probability for scenario s .
$F_i(\cdot)$	Fuel cost function of generator i .
i/\mathcal{I}	Index/set of generation units.
K_l^b	Flow distribution factor for line l at bus b .
K_0	Number of pieces for the piecewise linear approximation.
l/\mathcal{G}	Index/set of transmission lines.
L_l	Power flow capacity of transmission line l .
N_g, N_w, N_L	Number of generators, wind farms and loads.
p^n/\hat{p}^n	True probability/nominal probability for a certain scenario.
p_i^g	Output of generator i in the first stage.
p_i^{gd}	Down re-dispatch power of unit i .
p_i^{gu}	Up re-dispatch power of unit i .
\tilde{p}_i^g	Actual power output of unit i .
p_i^{min}, p_i^{max}	Minimum and maximum output of unit i .
p_k^L	Load demand k .
p_k^{Lc}	Load shedding in second stage.
r_i^u/r_i^d	Up/down reserve of unit i in first stage.
r_{it}^u/r_{it}^d	Up/down re-dispatch power at time t for unit i .
\overline{RD}_i	Shut-down ramp-down rate limit for generator i .
\overline{RU}_i	Start-up ramp-up rate limit for generator i .
RU_i/RD_i	Ramp-up/ramp-down limit for generator i .
SU_i/SD_i	Start-up/shut-down cost of generator i .
T_i^{up}/T_i^{dn}	Minimum up-time/down-time for generator i .
u_{it}/v_{it}	Binary start-up/shut-down variable for unit i at time t .
$w_t^b(\xi)$	Wind power output at bus b at time t for scenario ξ .
w_j^c	Wind power curtailment in second stage.
w_j^f	Forecasted output for wind generator j .
\tilde{w}_j	Actual power output of wind generator j .
x_{it}	Output of generator i at time t .
$\underline{x}_i/\bar{x}_i$	Minimum/maximum output of generator i .
y_{it}	Binary variable indicating the commitment status of unit i at time t .
$\alpha_t, \beta_t^n, \gamma_t^n, \lambda_t$	Dual variables for the chance constraint reformulation at time t .
δ	Power imbalance tolerance level.
ϵ_1	Risk level in the chance constraint.
Γ	Budget of uncertainty in the uncertainty set.
π_{il}^g	Power transfer distribution factor (PTDF) from generator i to line l .
π_{jl}^w	PTDF from wind generator j to line l .
π_{kl}^L	PTDF from load k to line l .
π_t^{gen}	Cost coefficient for generator up/down re-dispatch.
π_t^{ls}	Cost coefficient for load shedding.
ρ_{bt}^+, ρ_{bt}^-	Binary variables used to define the uncertainty set of wind power.
θ	Tolerance level in the ambiguity set.
$\sigma_{bt}^{1+}, \sigma_{bt}^{1-}, \sigma_{bt}^{2+}, \sigma_{bt}^{2-}$	Auxiliary variables to linearize bilinear terms in the second stage problem.
$\mu_t^1, \mu_{it}^2, \mu_{it}^3, \mu_{it}^4$	Dual variables in the second stage problem.

$\mu_{it}^5, \mu_{it}^6, \mu_{it}^7$
 $\tilde{\xi}_j$

Dual variables in the second stage problem.
The forecast error of wind generation j .

Chapter 5

a_i^{dg}, b_i^{dg}

Coefficients of linear cost function for generator i .

\mathbf{a}, \mathbf{b}

Coefficient vector of first-stage and second-stage problem.

c^{air}

Coefficient of the air specific heat capacity.

c^{gas}

Price of natural gas.

c^{loss}

Penalty cost coefficient.

c_i^{NL}

No load cost of generator i .

c_t

Coefficient vector in the objective function.

c_j^{ess}

Degradation cost coefficient of ESS j .

c_m^q

Degradation cost coefficient of TSS m .

c_t^{buy} / c_t^{sell}

Electricity purchase and sale prices at time t .

C_{jt}^{ess}

Degradation cost of ESS j at time t .

C_t^{grid}

Power exchange cost with the main grid.

D/\bar{D}

Ambiguity set and lifted ambiguity set.

E_{jt}

Energy storage of ESS j at time t .

$\underline{E}_j / \bar{E}_j$

Minimum and maximum energy storage level of unit j .

$\bar{g}_{t+1,k}$

The intercept of k th cut.

G_{it}

Natural gas consumption of MT generator i at time t .

G_{gt}^{GF}

Gas consumption of gas furnace g at time t .

h_{it}^{MT}

Heat output of MT i at time t .

h_{gt}^{GF}

The heat output of gas furnace g at time t .

$\underline{h}_g^{GF} / \bar{h}_g^{GF}$

Minimum and maximum output limit of gas furnace.

N^{MT} / N^{GF}

Number of MT and gas furnace.

p_t^{HC} / p_t^{AC}

Heat power supplied to heating coil and absorption chiller.

$\bar{p}_t^{HC} / \bar{p}_t^{AC}$

Maximum heat power supplied to heating coil and absorption chiller.

P_{it}

Output of generator i at time t .

$\underline{P}_i / \bar{P}_i$

The minimum/maximum output of generator i .

$P_{jt}^{ch} / P_{jt}^{dch}$

Charging/discharging power of ESS j at time t .

$\bar{P}_j^{ch} / \bar{P}_j^{dch}$

Maximum charging and discharging limits.

P_t^{load}

Aggregated load demand at time t .

P_t^{loss}

Auxiliary variable for power balance at time t .

q_t^{HE} / q_t^{CE}

The energy conveyed to the heating/cooling demand.

Q_t

Worst-case cost-to-go function for time t .

Q^G

Heat value of natural gas.

$r_{jt}^{e+} / r_{jt}^{e-}$

Charging and discharging power of ESS j at time t .

$\underline{r}_j^{e+} / \bar{r}_j^{e+}$

The minimum and maximum limit of charging power for unit j .

$\underline{r}_j^{e-} / \bar{r}_j^{e-}$

The minimum and maximum limit of discharging power for unit j .

$r_{mt}^{q+} / r_{mt}^{q-}$

Charging and discharging power of TSS m at time t .

$\underline{r}_m^{q+} / \bar{r}_m^{q+}$

The minimum and maximum limit of charging power for TSS unit m .

$r_m^{q-} / \bar{r}_m^{q-}$	The minimum and maximum limit of discharging power for TSS unit m .
R_i^{dn} / R_i^{up}	Ramp-down/ramp-up limit of unit i .
R^{tr}	The thermal resistance of building envelop.
T	Time horizon of the problem.
\mathbf{u}, \mathbf{v}	Auxiliary variables in the lifted ambiguity set.
$\bar{u}_{nt}, \bar{v}_{kt}$	The upper bounds of auxiliary variables in lifted support set.
w_t / w_{nt}	Random wind power at time t .
$\mathbf{w}^{[t]}$	Vector of wind power realizations from time 1 to t .
$\bar{\mathbf{w}}, \underline{\mathbf{w}}$	Maximum and minimum value vector of wind power.
W, \bar{W}, \tilde{W}	Support set and lifted support set of wind power.
x_t	Vector of decision variables.
\mathbf{x}^0	Constant vector in LDR method.
$\mathbf{X}^w, \mathbf{X}^u$	Coefficients matrices in LDR method.
$y_{it}, y_{it}^+, y_{it}^-$	Binary variables representing the on/off status, start-up and shut-down operation of MT i at time t .
η_g^{GF}	Output efficiency of gas furnace g .
$\eta_i^{MT} / \eta_{i,loss}^{MT}$	Efficiency coefficient and loss coefficient of MT unit i .
$\eta_j^{e+} / \eta_j^{e-}$	Charging/discharging rates of ESS j .
$\eta_m^{q+} / \eta_m^{q-}$	Charging/discharging rates of TSS m .
ξ_t	Random vector at time t including wind power and load.
$\tilde{\xi}_t$	Uncertainty realization at time t .
π_t	Lagrange multiplier vector in the lower bound problem.
θ_t^{am}	Ambient temperature at time t .
θ_t^{in}	Indoor temperature at time t .
$\underline{\theta}_t^{in} / \bar{\theta}_t^{in}$	Lower and upper bound of indoor temperature at time t .
$\boldsymbol{\mu}$	Mean vector of wind power.
$\sigma_{nt} / \gamma_{kt}$	Second-order moment parameters in the ambiguity set.
$\alpha_{kt}, \beta_{nt}, \lambda, \boldsymbol{\eta}$	Dual variables in worst-case expectation reformulation.
$\bar{\delta}, \tilde{\delta}, \bar{\theta}, \tilde{\theta}, \hat{\theta}, \bar{\rho}, \tilde{\rho}, \hat{\rho}$	Dual variables in conic dual problem.

1

Introduction

The main focus of this thesis is on the uncertainty modeling and related optimization problems in smart grid which combines multidisciplinary knowledge including artificial intelligence, operation research and power systems. The uncertainty mainly comes from renewable generation such as wind power and the research problems include forecast, microgrid energy management, unit commitment and economic dispatch problem.

1.1 Background

Energy is one of the most important substances in the world which is required almost all the time in our daily life. It is an essential part for all the lives on the earth including humans, animals and plants. As an important invention in human history, electrical energy is much more close to our everyday life nowadays, and we consume electricity every day no matter at home or at work with a great range of electrical devices such as computers, televisions, lights, smart phones and so on. Therefore, it is important to study power and energy systems for the stable development of our society and economy.

With the rapid development of advanced information and communication technology, traditional power systems have evolved towards the new smart grid which is also called smart electrical grid or intelligent grid. Compared with traditional power systems which usually carry power from some generation plants to a great number of consumers, smart grid is an advanced and automated power delivery network with two-way flows of electricity and information [1]. A general conceptual model of smart grid is shown in Fig. 1.1. As can be seen from this figure, based on the traditional power systems composed of generation, transmission, distribution and consumption, the information

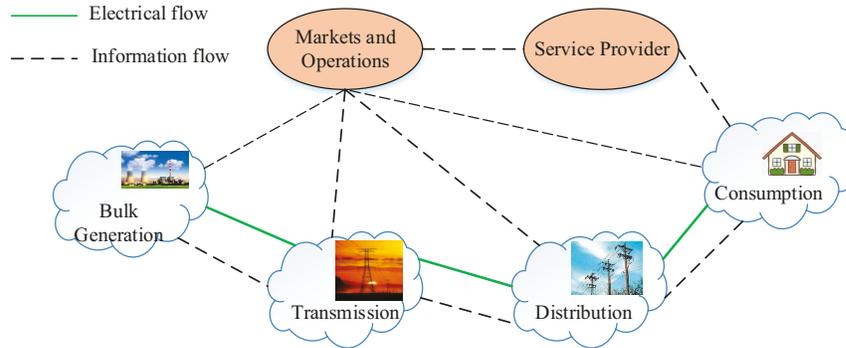


Figure 1.1: Smart grid conceptual model.

flow is included by using modern communication technologies which can help smart grid increase the power delivery efficiency and event response rate. In other words, the structure of a smart grid can be regarded as the interconnection of a traditional power system, a communication system and an information system [2]. Moreover, smart grid has many other advantages such as improving the power reliability and quality, enhancing resilience to disturbance, self-healing and so on.

In smart grid, distributed generation (DG) is an important generation form which utilizes distributed energy resources (DERs). Compared with traditional power systems, more renewable energy based DG units and storage units including battery storage systems and electrical vehicles can be integrated into smart grid. The integration of renewable generation also makes the traditional electricity users become energy producers with the two-way energy flow design, i.e., users can also produce and feed excess energy from DERs to the grid. DG also promotes the development of a new grid paradigm called microgrid which is an important component of future smart grid. A microgrid can be regarded as an interconnection of a group of DG units, energy storage devices and different loads in close proximity. It can operate in either an islanded mode or a grid-connected model to enhance the reliability.

Thermal power generation is the most common traditional electricity generating way which primarily consumes fossil fuels such as coal and crude oil. However, the amount of these fossil fuels on our earth is limited, and the great combustion of them causes greenhouse effect and severe environmental pollution problems. As a result, renewable energies such as wind and solar energy which can be integrated into smart grid have attracted significant attention recently. Renewable energies have many advantages such as cleanness and wide availability, and their development and utilization can effectively overcome the problems of burning fossil fuels.

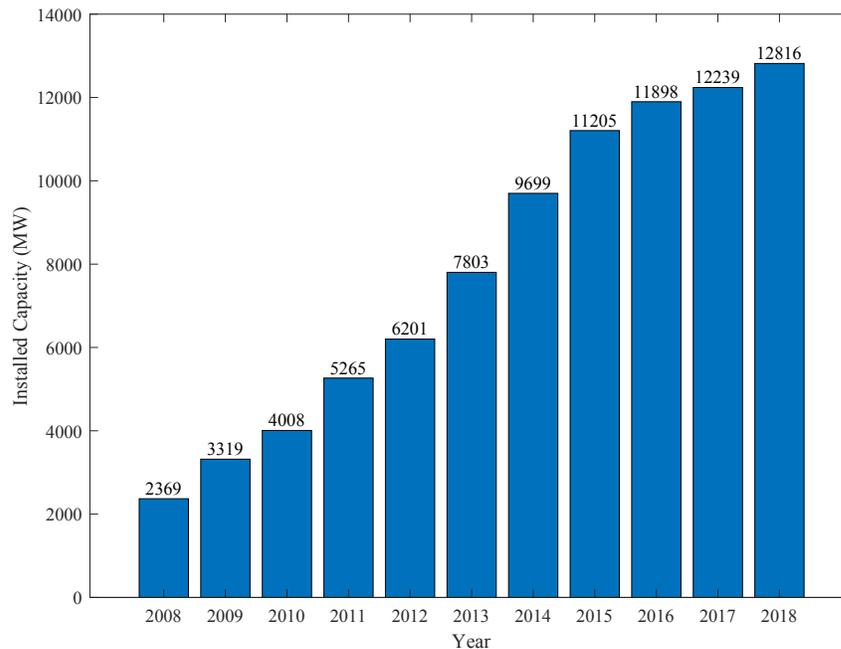


Figure 1.2: The growth of installed wind power in Canada.

There are different kinds of renewable energies that can be utilized on the earth including the common wind energy and solar energy, hydroelectric power, tidal energy and geothermal energy. With the increasing application in practice, wind energy has become one of the most popular renewable sources in the world [3]. Take Canada as an example, approximately 6% of the electricity demand is supplied by wind energy which satisfies the power demand of more than three million Canadian homes. In addition, more wind energy has been installed between 2008 and 2018 compared with other generation sources, with an yearly average growth rate of 20% over the past decade as shown in Fig. 1.2, and the installed capacity of wind energy reached 12,816 MW in 2018 [4]. Consequently, integrating renewable energies such as wind energy into smart grid is important to alleviate the increasing energy crisis and worth studying.

Although renewable energies have many distinct advantages, the main drawback that cannot be ignored is their uncertainty, i.e., they are intermittent and volatile which depend on the variable weather to a large extent. As a result, the high penetration of renewable generation in smart grid will lead to new challenges to the stable operation of power systems. In addition to the uncertainty from renewable generation, there are also many other uncertain sources in smart grid including the random load demand and variable electricity price. All these uncertainties can influence the reliability and stability of system operation. Therefore, research on how to quantify the uncertainty or how to eliminate the impact of uncertainty are of great significance to the development of robust smart grid.

In order to deal with different kinds of uncertainties, various methods have been studied in previous research including forecast method and mathematical modeling analysis. Among these methods, the most direct method is the forecast approach which intends to estimate the unknown values of random variables. During the past decades, numerous forecast methods have been proposed and applied in diverse fields involving renewable energy [5], load demand [6] and electricity price [7]. Compared with the initial point forecast, probabilistic forecast was studied later which shows more advantages in quantifying the uncertainty. For mathematical modeling analysis approach, stochastic programming (SP) and robust optimization (RO) are most studied to handle the random variables, where the former usually uses a certain probability distribution or scenario representations and the latter adopts uncertainty set to describe the uncertainty. However, both SP and RO methods have some shortcomings with respect to the distribution assumption and over-conservativeness of the solution, respectively, which may be further improved.

1.2 Research Definitions

In this section, the critical terms and problems investigated in this thesis are clarified to identify the scope of this research work.

1.2.1 Uncertainty Modeling Based on Forecast

High penetration of renewable generation has posed many new challenges to the stable operation of power systems. To deal with the uncertainty problems, the most direct method is forecast which aims to determine the value of a random variable based on historical data. Statistical modeling technique and learning method are usually adopted in forecast approach. Different forecast methods have been proposed for wind power or wind speed prediction over the past decades. According to the forecast output, forecast method can be classified into point forecast and probabilistic forecast. Compared with point forecast method which only generate a deterministic forecast value, probabilistic forecast could provide additional quantitative information about the uncertainty [8]. Probabilistic forecast actually includes interval forecast, quantiles forecast and probability density function forecast and so on, among which interval forecast is the most visualized one. Interval prediction method can generate estimated intervals of random variables with a confidence level which provides more uncertainty information than point forecast.

Interval prediction is actually equivalent to an optimization problem with several new evaluation indices such as the reliability index and interval width index [9]. By combining these two indices into a single objective, the interval prediction can be handled as a

single-objective optimization problem. In addition, by directly optimizing the two indices simultaneously, the prediction interval (PI) construction problem is transformed into a multi-objective optimization problem.

1.2.2 Neural Network Model

Artificial neural network (ANN) which is also referred to as neural network (NN) has been motivated from the different computational way of human brain and natural intelligence. ANN is a complex system network composed of a great number of processing units called neurons and it has strong nonlinear learning ability. There are many properties or capabilities in a ANN model such as nonlinearity, input-output mapping, parallel and distributed processing and adaptivity [10]. The connection weights in a NN are gradually adjusted in the learning process and used to store the obtained knowledge.

NN models are data-driven techniques which can represent a complex nonlinear mapping relationship between input and output by learning from abundant historical data. Due to the strong learning and generalization ability, NN models have been widely studied in many problems including pattern recognition, classification and forecast. Generally, NN models have two structures: feedforward and feedback. The feedforward model conducts the computation from the input to the output in a forward direction and it may contain one or more hidden layers. For NN models with feedback structure, or recurrent NNs (RNNs), they are very useful in modeling a dynamic system.

1.2.3 Microgrid Energy Management

With the development of DG in smart grid, a new grid paradigm called microgrid has been proposed which is regarded as one of the cornerstones of future smart grid [1]. A microgrid a low-voltage distribution system with the integration of various DERs and controllable loads [11], and DERs contain different kinds of DG sources such as conventional generators, wind turbines, photovoltaics (PV), hydro, biomass and distributed storage systems [12]. In a microgrid, all DERs work coordinately to provide desired reliable and stable power, and feed the power directly to the consumers or the main utility grid [13]. Microgrids have different functionality by operating in either a grid-connected mode or an islanded mode, and they should have the transition ability between two operation modes.

Microgrids have attracted considerable attention in recent years due to their advantages in dealing with the energy supply problems with DERs. In order to achieve reliable and economic operation of a microgrid, an energy management system (EMS) is usually required to schedule and dispatch the generation units considering various constraints. With the increasing penetration of renewable generation in microgrids which

are intermittent and uncertain, microgrid energy management problems also become more complex and challenging.

1.2.4 Distributionally Robust Optimization

SP can effectively describe many uncertain decision-making problems where uncertainty can be represented by a probability distribution. Although SP problem is a convex optimization problem, it is usually solved via Monte Carlo approximations with many scenarios which results in heavy computational burden. Another challenging difficulty of SP is how to obtain the exact distribution of random variable in practice. In order to address these issues, a robust version of SP was proposed by Scarf [14] which is also called distributionally robust optimization (DRO).

DRO method can be seen as an intermediate approach between SP and RO. Instead of making an assumption of certain distribution, the DRO method optimizes the expected objective over a set of unknown distributions sharing certain statistical characteristics such as moment information including mean and variance, and this set is usually called ambiguity set. By incorporating such properties of distribution information, DRO method can produce less conservative solutions compared with distribution free optimization method thus addressing the limitations of SP and RO method. The research about this method has made much progress [15] [16] and it has been effectively applied to solve optimization problems in power systems recently.

1.2.5 Unit Commitment

Unit commitment (UC) and economic dispatch (ED) are two important problems for energy scheduling and grid management. In a deregulated electricity market, UC is usually used for day-ahead market clearing, reliability evaluation and intra-day operations by independent system operators. The output of UC generally includes the commitment status and dispatch decisions of generation units which satisfy the related system constraints like power balance and power flow limits and specific unit constraints such as ramping up/down and capacity limits [17], and the objective is usually to minimize the total system cost.

In UC problem, binary variables are introduced to represent the status of generation units (on/off) which makes the problem difficult to solve together with many other constraints when problem size becomes large. A lot of different formulations as well as solution methods for UC problem have been proposed in previous literature [18]. Although UC problems have been widely studied, many new challenges emerge as the use of renewable generation increases, particularly, a large amount of uncertainty is added to the UC problem and generation scheduling in grid management [19]. As a result, new UC

models and solution approaches should also be developed to incorporate the uncertainty modeling.

1.2.6 Energy and Reserve Dispatch

ED is also an important problem in power system operation which intends to control the total generation to match the total load and minimize the system operating cost. ED is often formulated as an optimization problem with the economic objective and related constraints [20]. However, ED is also influenced now by the uncertainty from increasing integration of renewable generation. To resolve the uncertainty of renewable generation in power systems, one method is to schedule more reserves and use more flexible resources to ensure sufficient energy supply. Although adequate reserves can protect the reliability of electricity supply, the economic cost with reserve provision will also increase. Therefore, it is significant to co-optimize the energy and reserve dispatch to improve the reliable and economic operation of power systems with uncertain renewable energies [21] [22].

1.3 Literature Review

In this section, previous studies which are closely related to this research are reviewed as follows.

1.3.1 Wind Power and Wind Speed Forecast

Wind power is the most common renewable energy studied in smart grid and wind forecast has been widely investigated. Generally, wind forecast consists of direct wind power forecast and indirect wind speed forecast.

- **Wind power forecast:** Based on the forecast time scale, wind power prediction can be classified into short-term, medium-term and long-term prediction [23]. While according to the forecast models, it can be divided into two main categories: physical and statistical [24]. Physical methods are developed by modeling the relationship between physical variables and the specifications of wind turbines with some physical-based equations. For the statistical methods, time series models (e.g., autoregressive integrated moving average ARIMA) [25], data mining approaches [26], artificial neural networks (ANNs) [27], [28], and support vector machines (SVM) [29] are widely studied in the literature. Compared with physical models, statistical models, as a data-driven technique, are usually simpler and more adaptive. In addition, hybrid methods which combine different models have also been researched [30] to improve the forecast performance. This kind of approach aims to retain the advantages of the individual model and it seems to be better than other methods.

Conventional point forecast methods as mentioned above suffer from the problem that they cannot eliminate the forecast error and the forecast accuracy may be highly variable. They only generate a deterministic forecast value for a certain time step without any indication of the associated uncertainty [31]. From the decision making point of view, the use of point forecast will have significant impact on the stable and reliable operation of power system. Therefore, recent research of wind power prediction have focused more on probabilistic methods which can include the forecast uncertainty [32].

Compared with the widely used point prediction methods, probabilistic forecasts could provide additional quantitative information about wind power generation uncertainty [8]. As a result, the probabilistic forecast of wind energy has attracted much attention recently [33], [34]. In probabilistic forecast, the uncertainty can be expressed by probabilistic measures such as probability density functions (PDFs), quantiles and intervals, moments of distribution (mean and variance) and so on [8]. As the most visualized representation, interval forecast of wind energy gained more popularity [35], [36].

Among different kinds of interval forecast approaches, the recently proposed lower upper bound estimation (LUBE) method is a nonparametric procedure that can construct appropriate prediction intervals (PIs) directly in an unsupervised learning mode based on a feedforward neural network (NN) [9]. It makes no assumption about the forecast error distribution and its computational burden is almost negligible in comparison with other traditional NN-based PI construction methods [37]. To obtain high quality PIs with narrow width and large coverage probability, both single-objective procedure [38] and multi-objective framework [39] using LUBE method are proposed in the literature. In [40], a constrained single-objective framework is designed to minimize the PI width while constraining the coverage probability. Similarly, a hybrid model based on back propagation NN is proposed for wind speed interval forecast [41]. In this study, the wavelet de-noising technique is employed to preprocess the data, and cuckoo search optimization algorithm is used to train the NN model. In the multi-objective framework, Pareto optimal solutions [39] and fuzzy inference method [42] are investigated to construct the optimal PIs. In these studies, the prediction models are all based on the feedforward NN model. In addition, rough neural networks which combine rough and conventional neurons to deal with uncertainties of the data are also studied for wind power or speed prediction [43]. In [44], the rough neurons are integrated in a deep neural architecture to improve the accuracy of short-term wind speed forecast. In [45], the most informative input parameters of an NN are

determined through attribute reduction of rough set theory and the training time is also reduced. Although good results are reported in these works, they concentrate on wind speed point forecast instead of interval forecast. The combination of deterministic and probabilistic forecast for wind power is also studied [46]. It is an indirect probabilistic forecast method which firstly implements point forecast with support vector regression (SVR) models and then the confidence intervals for forecast values are obtained with quantile regression (QR) method.

Although NN-based LUBE method has been widely studied as aforementioned, the existing research, to the author's best knowledge, focus on the feedforward NN prediction model, especially the multilayer perceptron (MLP) NN model. Generally, NN models have two basic structures: feedforward and feedback. Compared with the feedforward NN, recurrent neural network (RNN) with a feedback structure has been shown to excel at time series forecast [47]. In a feedback structure, the network behaves like a dynamic system which can better capture the characteristics of variable wind speed [48]. However, interval prediction for wind power with RNN model is rarely reported in the literature.

- **Wind speed forecast:** Similarly, wind speed forecast has also been researched for several decades. In particular, the research on short-term wind speed prediction with ANN model or other statistical models has attracted much attention. In [49], different weather data including temperature, relative humidity and air pressure are considered to predict the wind speed by using the common backpropagation neural network. In [50], SVM model is studied to conduct short-term wind speed forecast and in this forecast method, the dataset is preprocessed by empirical model decomposition (EMD) and the parameters of SVM model are tuned by an improved cuckoo search algorithm. In addition, hybrid methods have also been studied to enhance the prediction accuracy by combining several models [51], [52]. The RNN model has been extensively studied for wind power and speed forecast. For instance, the local RNNs are studied for the long-term wind speed and power prediction in [53]. These models were trained by two on-line learning methods based on recursive forecast error algorithm. Similarly, the long-term wind speed prediction was studied using a composite method of statistical and NN approaches in [23]. Based on the general trend and pattern extracted from the historical data, the nonlinear autoregressive with exogenous inputs (NARX) network was trained and used to forecast the next year hourly data. Moreover, the Elman neural network was also explored for wind speed forecast [54].

Interval prediction method has also been applied to wind speed forecast as

mentioned above. Generally, the PI construction problem is defined as a single-objective optimization problem. Considering the complicated nonlinear and non-differentiable objective function, evolutionary algorithms instead of traditional gradient based methods such as simulated annealing (SA) [9] and particle swarm optimization (PSO) [40] are adopted to solve the problem. However, PI construction is actually a multiobjective optimization problem as high quality PIs need both sufficient reliability and narrow width.

Compared with abundant single-objective interval prediction research, there are only a few studies about multiobjective interval prediction for wind speed. Based on a simple multilayer perceptron NN model, the short-term wind speed interval prediction is performed in a multiobjective framework in [39]. The NN model was trained by a multiobjective evolutionary nondominated sorting genetic algorithm II (NSGA-II) [55]. Similarly, radial basis function (RBF) NN model is also investigated for wind speed multiobjective interval prediction [56]. In addition, SVM model is also applied to predict wind speed which is trained by the multiobjective differential evolution algorithm [57]. However, unlike the NN model, two SVMs are used to create the PIs' lower and upper bounds in this study which may increase the computational burden.

1.3.2 Microgrid Energy Management with Uncertainty

Microgrid energy management has become challenging with the increasing penetration of renewable generation. To deal with the uncertainty caused by renewable energy generation in microgrid energy management, various methods have been investigated including robust optimization, stochastic programming and chance constrained programming. Among these methods, the application of robust optimization [58] in microgrid energy management has been extensively explored recently. For example, the energy management for a grid-connected microgrid with high penetration of wind power and demand-side management is studied in [59]. The formulated robust optimization problem is solved by a distributed algorithm to minimize the system net cost. Similarly, a scenario-based robust energy management method for a grid-connected microgrid is proposed to deal with the uncertain renewable generation and load in [60]. By exploring the worst-case scenario, a robust energy management solution for the proposed model could be obtained. In [61], a two-stage adaptive robust optimization method is studied for microgrid energy management which considers the uncertainties of renewable generation and grid-connection condition. The uncertainties are controlled by the "budget of uncertainty" parameters in this work.

Stochastic programming is another common approach to handle the uncertainty

in microgrid operation and energy scheduling. With this method, the uncertainties associated with renewable generation and load are usually represented by certain distributions such as normal distribution [62], [63]. Although not so many as the research on robust and stochastic optimization model, chance constrained programming (CCP) method is also investigated in microgrid energy management [64]. For instance, in [65], two CCP problems are formulated for grid-connected microgrids energy management and the problems are solved by a linear programming transformation. Also, a grid-connected microgrid based on combined heat and power (CHP) system is studied in [66]. The CCP method is employed to describe the uncertainty of renewable generation and load and the optimal schedule problem is solved by a particle swarm optimization (PSO) based algorithm. In [67], chance constrained optimization is employed to solve the demand response problem for a home energy management system and the chance constraint is used to describe the variable power interaction between the household and the utility grid due to uncertain load demand. In addition, the CCP method has also been studied for other power system problems such as transmission expansion planning problem with load and wind uncertainties [68].

Although the microgrid energy management has been widely studied, most of the research focuses on the grid-connected microgrids with uncertainties compared with few research works on islanded microgrids [69], [70]. In addition, the constraints with uncertainties have not been sufficiently studied which mostly concentrate on the balance of generation and load demand [65].

1.3.3 Unit Commitment

As a critical application in power system operation, UC problem has been studied for a long time which aims to reduce system cost and improve reliability by optimal scheduling of generation units. Due to the high-level penetration of intermittent and unpredictable renewable energy, new challenges about secure and reliable system operation also arise such as large reserve capacity demand, divergence of area control error and possible power imbalance [71]. Consequently, it is imperative to incorporate associated uncertainties into UC problems with renewable generation so that more reliable solutions can be attained.

Although UC problem is usually nonconvex and very difficult to solve, many efforts have been developed for UC problems with uncertainties in the past decades. In particular, the stochastic optimization methods [17] attract the most attention, which can be typically classified into two categories: stochastic programming and robust optimization methods. Stochastic programming is a traditional method to deal with data uncertainty and was first investigated to solve uncertain UC problems [72]. For example, in [73], a security-constrained UC (SCUC) problem with uncertain wind power is studied

and the wind power is assumed to follow normal distribution. In [74], a stochastic UC model considering the uncertain load and outages is proposed, and the random variables are represented by scenarios trees. Generally, it is assumed that the probability distributions of random variables are known in stochastic programming methods, and the objective is to minimize the expected total system cost. However, the exact probability distribution is usually hard to be known in practice. In addition, stochastic programming methods suffer from heavy computational burden as substantial scenarios are required to comprehensively represent the probability distribution.

Robust optimization is another popular method to deal with uncertainties in UC problems. Compared with stochastic programming, the true probability distribution is not required in robust optimization, and the random variables are represented by some uncertainty sets. A vast number of literature about robust UC problems has been reported, such as the typical two-stage robust UC models [75] [76] [77] which consist of first-stage commitment decision and second-stage recourse action. In addition, multistage robust UC models have also been studied recently by taking into account the non-anticipativity of dispatch decisions [71] [78]. Compared with stochastic programming, robust optimization ignores the probabilistic information and tries to find the minimal cost under the worst-case scenario within the uncertainty set. Although the solution is robust against all uncertainty realizations, it may be over-conservative since the worst-case scenario rarely occurs in practice.

To address the shortcomings of stochastic programming and robust optimization methods, an alternative method, distributionally robust optimization (DRO), has attracted much attention recently [15]. The DRO method aims at optimizing an uncertain problem under the worst-case distribution from a so-called ambiguity set. The ambiguity set is a family of probability distributions which share some certain statistical information. Since partial distribution information is utilized in DRO method, the conservativeness of the solution from this method is between robust optimization and stochastic programming. In recent years, the DRO method has been widely applied to solve power system optimization problems, especially the moment-based DRO method [16]. For instance, in [79], a moment-based DRO model is proposed for UC problem, and linear decision rule is used to reformulate and solve the intractable problem. In [80], DRO approach is used to solve the contingency-constraint UC problem and the ambiguity set of contingency probability distributions is constructed based on available moment information. Similarly, the moment-based DRO method is used for co-optimization of energy and reserve dispatch in [81], and the problem is reformulated to a tractable semidefinite programming (SDP) problem.

In addition, DRO method has also been applied to optimal power flow [82], power system planning problem [83], distributed generation capacity assessment of active distribution networks [84], optimal bidding problem of electricity markets [85], transmission line congestion management problem [86] and so on. A general framework of these works is that an ambiguity set is proposed first based on the assumption of the research problem and a tractable reformulation is then derived based on dual and approximation techniques.

Moment-based DRO method only considers moment information such as expectation and variance. However, the actual distribution contains more information than moments. In practice, a number of historical data of random variable are usually available from which we can obtain more valuable distribution information, e.g., an estimated distribution by data fitting. Therefore, distance-based or data-driven DRO methods have been investigated in some very recent studies [87]. In [88], a distance-based DRO model is studied for UC problem, and the ambiguity set is constructed based on Kullback-Leibler (KL) divergence. Data-driven DRO method is also reported for UC problem [89] where the confidence band of cumulative distribution function (CDF) is used to construct the ambiguity set. Similarly, L_1 norm and L_∞ norm are used to construct confidence sets in a data-driven manner for stochastic UC problem [90] [91]. With the same confidence sets, a new duality-free decomposition method is proposed to solve the distributionally robust UC problem in [92]. Wasserstein metric based DRO method has also been studied for UC problem [93] [94] which constructs the ambiguity set based on Wasserstein ball. In addition, data-driven DRO method is also studied for reserve and energy scheduling problem [95] which considers the Wasserstein distance. According to abundant related works, DRO based optimization problems are usually very complicated and even intractable, and different ambiguity set construction methods lead to various problem reformulation methods.

1.3.4 Energy and Reserve Dispatch

Energy and reserve dispatch is another important problem in power system operation. To ensure the reliability of power system operation, new types of reserves are considered to compensate the uncertain renewable generation, which makes it important to co-optimize the energy and reserve dispatch from the economic perspective [81]. Similar with UC problem, SP and RO method have been investigated in energy and reserve dispatch problem [96] [97] [98].

Considering the deficiencies of stochastic and robust optimization methods, the new technique DRO has also been applied to deal with uncertainty according to partial distribution information for energy and reserve co-dispatch [99]. In [100], a

distributionally robust reserve model is proposed to minimize the cost of generation and reserves, and the uncertainty from wind power forecast error is captured by an ambiguity set based on first and second-order moment information. In [101], a two-stage hydro-thermal-wind economic dispatch model considering distributional robustness is studied, and the uncertain probability distribution of wind power is also described by moment-based ambiguity set. In addition, distance-based DRO methods have also been studied for energy and reserve dispatch problems. In [95], a two-stage data-driven distributionally robust energy and reserve scheduling model with wind power is studied, and the Wasserstein ball based ambiguity set is used to contain all possible probability distributions. Also, a two-stage risk-averse stochastic model is proposed for energy and reserve dispatch problem in [102], and Kernel density estimation is used to estimate the probability distributions of wind power captured by the L_1 -norm based ambiguity set.

1.3.5 Multistage Energy Management for Microgrids

As analyzed before, microgrid energy management with uncertainty has been widely studied with various modeling methods including SP, RO and DRO method. However, most of these research works focus on single-stage or two-stage models, for example, the single-stage robust energy management model [60], two-stage stochastic energy management or scheduling model [62] [103] [104], two-stage robust energy management model [61] [105] and so on. However, one problem in single-stage or two-stage models is that the non-anticipativity of uncertainty is not considered.

As discussed above, single-stage or two-stage energy management models have been widely studied for microgrids. In such models, the first stage is to determine the day-ahead schedule under uncertainty, and the second stage aims to re-adjust the optimal schedule with fixed uncertainty realization. Compared with two-stage models, multistage models with uncertainty (e.g., multistage stochastic programming models) are more complicated and computationally difficult. To overcome the computational tractability problems of multistage models, various methods have been proposed, including the sample average approximation and the popular decomposition method, stochastic dual dynamic programming (SDDP). SDDP method was first proposed in [106] for the hydrothermal generation scheduling problem. Recently, this method has been applied to deal with power system multistage optimization problems, such as the real-time economic dispatch [107], energy storage management [108], and DER aggregators operation under multiple sources of uncertainty [109]. Although SDDP method has been demonstrated to solve the computational challenge of multistage stochastic optimization problems, a limitation is that the assumed distribution of random variables is hard to be known in practice, and it is usually approximated by fitting the historical data. In comparison with the study of multistage stochastic programming models, multistage robust or distributionally robust

models have been seldom studied for microgrid energy management problems.

1.3.6 CCHP Microgrid

Among various microgrids, one interesting kind is the combined cooling, heating and power (CCHP) based microgrid which is also known as tri-generation system and can provide electric and thermal power simultaneously [110] [111]. A CCHP based microgrid usually consists of renewable generation, CCHP units such as micro turbines (MTs), heating and refrigeration system and different kinds of loads. The energy utilization efficiency can be significantly improved (e.g., to be 80%) by implementing CCHP in a microgrid compared with traditional independent energy system [112]. Therefore, CCHP microgrid is considered as a leading power generation method in electricity market with the efficiency and environmental concern.

There are generally two operational modes for CCHP units in practice, i.e., following the electric load and following the thermal load depending on the priority of load satisfaction [113]. To decouple the electric and thermal output of CCHP units, storage systems are usually utilized in the microgrid operation. Energy management or dispatch for CCHP microgrids has been widely studied with various strategies and methods. For example, in [114], a coordinated operation strategy was proposed for a distribution system integrated with gas-electricity and CCHP systems, and the accurate forecast of renewable generation is used. Similarly, an optimal dispatch strategy for a CCHP system was proposed to minimize the total operation cost with forecasted wind power in [115]. However, forecast errors for renewable generation cannot be eliminated fully, and exact forecast values can hardly be obtained in practice.

Similar with the above discussion, SP and RO method have also been applied to manage the energy scheduling for CCHP microgrids. For example, the optimal short-term scheduling of combined heat and power (CHP) based microgrids is studied through a stochastic programming formulation in [116] and the wind speed was assumed to follow a Weibull distribution. The coordinated day-ahead scheduling and real-time dispatch models are developed for the coupled co-optimization of cooling and electricity energy in [117], and the uncertainty of wind power is represented by multiple scenarios generated from normal distribution. The limitation of stochastic programming method is that it usually suffers from high computational burden with many scenarios. In addition, the true probability distribution cannot be known exactly in practice.

Compared with stochastic optimization method, robust optimization does not need the true distribution assumption, and it has also been a popular method to handle the uncertainties. In [118], a two-stage new robust coordinated operation method is proposed

for a grid-connected CCHP microgrid with multiple uncertainties which are described by uncertainty sets. Similarly, a two-stage adaptive robust optimization approach is developed in [61] for energy management of a microgrid with CHP units and uncertain wind power. In [119], a robust model based on information gap decision theory was formulated to derive the optimal operation strategy for CHP units, and the envelop bound model is used for uncertainty modeling in this work. Based on the above analysis, we can find that distributionally robust energy management has seldom been studied for CCHP microgrids.

1.4 Thesis Motivation and Objectives

As discussed above, the increasing penetration of renewable generation has brought many new challenges to smart grid. The common uncertainty modeling methods for renewable generation include (1) forecast technique and (2) mathematical modeling analysis method. Although a lot of related research works have been done to deal with uncertainty modeling as introduced in the literature review, there are still many aspects that can be further explored or improved. In particular, compared with the common point forecast for renewable energy, probabilistic method can be further investigated. In addition, DRO may be studied to model the uncertainty instead of the general SP and RO method. Therefore, exploring novel forecast technique and new modeling method based on DRO to capture the uncertainty of renewable generation constitute the main motivation of this research. The specific research issues consist of wind power and wind speed forecast, energy management for islanded microgrids and grid-connected microgrids, UC problem, and energy and reserve dispatch. Based on the challenges in these research problems illustrated in the literature review, the main motivation and objectives of this work can be described as follows.

- **Wind Power Interval Prediction Based on RNN**

It is known that forecast errors cannot be fully eliminated and the forecast accuracy of point forecast may be highly variable. Compared with the common point forecast, probabilistic forecast, specifically the interval forecast could provide more quantitative information about uncertainty. Although NN-based interval prediction has been widely studied as mentioned before, those works mainly focus on the feedforward NN prediction model. However, RNN model has not been studied for interval prediction which shows better performance for time series forecast. As a result, in this study, the short-term wind power interval prediction based on RNN model and LUBE method is investigated for the first time. Particularly, the single-objective framework is employed and a novel aggregated cost function is designed as the objective of model training. Considering the high complexity and

nonlinearity of the cost function, the dragonfly algorithm (DA)-a new intelligent and powerful optimization algorithm-is introduced to solve the problem effectively. In addition, a new modification is proposed for DA to reinforce its search ability. To cope with the chaotic historical wind power data, delay embedding theorem instead of the general correlation analysis is applied to reconstruct the data in this study.

- **Multiobjective Interval Prediction for Wind Speed**

Wind speed is the direct source of wind energy, and the interval forecast for wind speed is also investigated in this research. For interval prediction models, most of them focus on the single-objective framework which combines different evaluation indices into a comprehensive objective function, while there are only a few studies about multiobjective interval prediction for wind speed. Actually, the construction of PIs is essentially a multiobjective optimization problem which needs to optimize various evaluation indices simultaneously. Therefore, the research of multiobjective wind speed interval forecast is still not sufficient, and this thesis intends to further improve the multiobjective interval prediction method from both the prediction model and the optimization algorithm. In particular, the interval prediction for short-term wind speed is performed based on NN models in a multiobjective framework in this study.

- **Chance-constrained Energy Management for Islanded Microgrids**

Microgrids are important components of future smart grid which have attracted much attention over the last decade. As discussed above, microgrid energy management has been widely studied, however, most of the works investigated the grid-connected microgrids with uncertain renewable generation. Moreover, the effect of uncertainty has not been sufficiently studied in these works which mostly employ the power balance constraint. Therefore, in this research, the islanded microgrid will be first studied with the emphasis of uncertainty influence. Note that for real islanded microgrids integrated with renewable energy, the power balance may not always be satisfied due to the uncertain output of renewable energy. Motivated by this point, a chance-constrained (CC) problem for islanded microgrids energy management involving renewable generation is proposed to ensure the system reliability. The uncertainty of renewable generation is represented by a new box-type ambiguity set based on which the CC problem can be solved by DRO method.

- **Data-driven Distributionally Robust Chance-constrained UC**

UC problem is one of the most fundamental tasks in power system operation which can also be integrated into microgrid energy management. Traditional UC problems have been extensively studied, however, with the development of new uncertainty modeling method, distributionally robust UC has been proposed recently. Compared with moment-based DRO method which only considers moment information such as mean and variance, distance-based DRO method can usually extract more distribution information of random variable from a number of available historical data. Therefore, a distributionally robust chance-constrained UC problem based on distance-based DRO technique is studied in this work which further enriches the research of UC problem. The studied UC problem here is formulated as a two-stage model. In the first stage, with a chance constraint restricting the probability of power imbalance, the commitment decision and a base-case dispatch plan are determined, and in the second stage, the operational risk or expected re-dispatch cost caused by load curtailment or wind power spillage under the worst-case wind power distribution is considered.

- **Energy and Reserve Dispatch Using DRO Method**

Economic dispatch is another basic and significant problem in power system operation and optimization. After the UC decision is determined, the optimal dispatch problem will become the focus. To tackle the uncertainty of renewable generation, new types of reserves emerge, and co-optimization of energy and reserve dispatch has become an important problem for reliable system operation. Different from previous literature, a data-driven two-stage energy and reserve dispatch problem using distance-based DRO method is studied in this thesis. Particularly, a two-stage model is formulated which minimizes the generation and reserve cost with forecasted wind power in the first stage and minimizes the expected re-dispatch cost considering the worst-case probability distribution in the second stage.

- **Multistage Robust Energy Management for Grid-connected Microgrids**

Among various studies of microgrid energy management with uncertainty, single-stage or two-stage models have been broadly employed as mentioned before. In such models, the day-ahead schedule is usually determined in the first stage, and the second stage aims to re-adjust the optimal dispatch with fixed uncertainty realization. These models have limitations in capturing the uncertainty despite their prevalence. With a simple uncertainty description method, it is typically assumed that the information is perfect and the uncertainty is anticipative in the second stage. More specifically, after the determination of day-ahead schedule, the

uncertainties are all revealed at the same time for the scheduling horizon (e.g., 24 hours) with perfect information assumption in the second stage [109]. However, in practice, the uncertainties can only be revealed gradually and the future information is unknown at current time. Thus, a multistage model would be more practical and proper to capture the inter-temporal uncertainties than a two-stage model. Therefore, multistage energy management problem for grid-connected microgrids is studied in this work which considers non-anticipative uncertainty, and RO method is adopted to describe the uncertainty. To deal with the computational difficulty of multistage robust model, a novel decomposition method similar to SDDP method, i.e., the robust dual dynamic programming method (RDDP), is proposed to solve the problem.

- **Multi-period Energy Management for CCHP Microgrids with DRO**

With the introduction of CCHP units in microgrid, the energy utilization efficiency can be significantly improved. Hence, CCHP microgrid is considered as an important generation method in current electricity market, and the general energy management method can also be applied to CCHP microgrids. As discussed in the literature review, the research of distributionally robust energy management for CCHP microgrids is rarely reported. In addition, the existing works about CCHP microgrids also mainly focus on the single-stage or two-stage models which ignore the non-anticipativity of dispatch decisions. Consequently, to enforce the non-anticipativity of dispatch decisions, a multi-period energy management model for CCHP microgrids with DRO technique is proposed in this thesis. The proposed multi-period model is included in a two-stage framework, and here we use multi-period instead of multi-stage to make a difference. To capture the uncertain distribution of renewable generation such as wind power, DRO method is investigated and a new second-order conic representable ambiguity set is designed.

1.5 Thesis Outline

In this thesis, uncertainty modeling and related optimization problems in smart grid with renewable generation are studied. The studied uncertainty modeling techniques encompass direct interval forecast, DRO method and RO method. The relevant problems involve wind power and wind speed interval prediction, microgrid energy management, UC and energy and reserve dispatch. The relationship of these items is illustrated in Fig. 1.3. As shown in this figure, interval prediction of wind energy is first conducted which can estimate the intervals of uncertain renewable generation. Based on these interval results, microgrid energy management in distribution system is studied which considers the interval based ambiguity set and uncertainty set. Moreover, both the chance-constrained

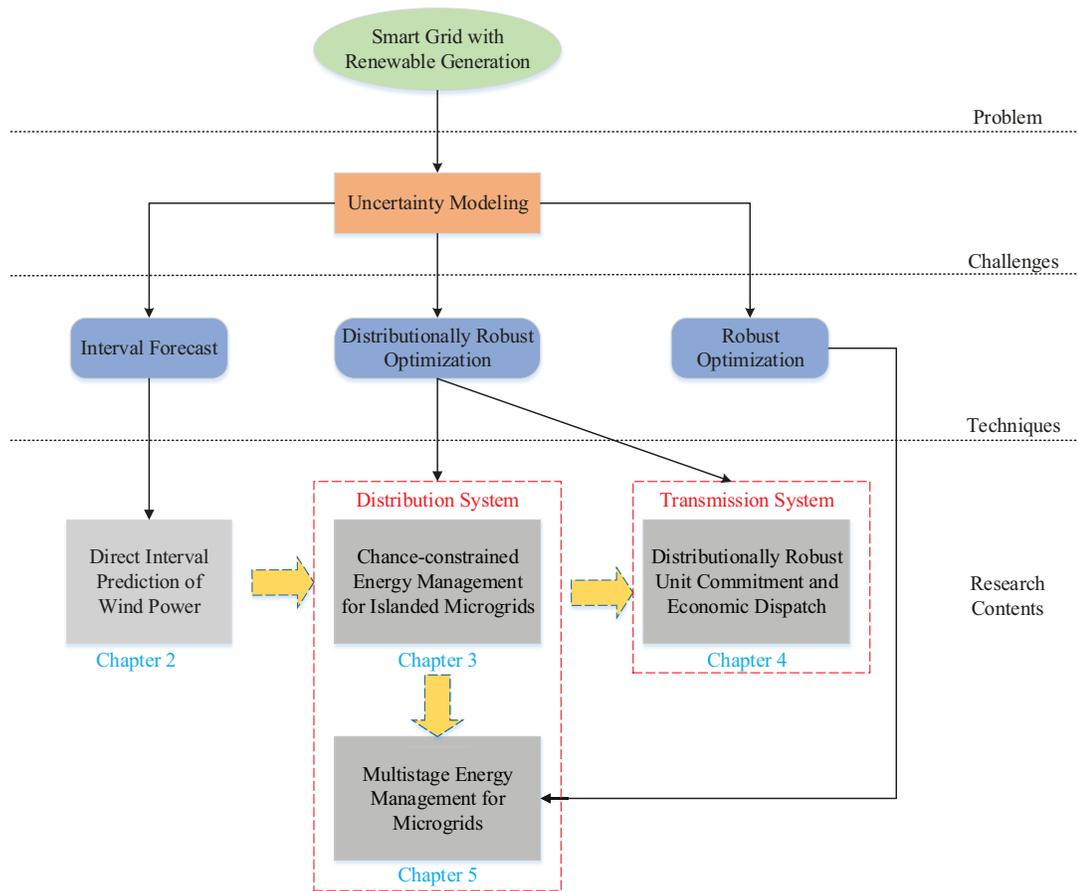


Figure 1.3: Outline of this thesis.

energy management for islanded microgrids and multistage model for grid-connected microgrids are studied. In addition, based on the research of chance-constrained energy management model, UC and ED problems in transmission system are also investigated with DRO method. More specifically, this thesis consists of six chapters and the detailed descriptions are given as follows.

- **Chapter 1: Introduction** - The research background is first briefly introduced in this chapter, followed by the definition of research problems and scope. Then relevant literature is reviewed to highlight the research challenges for each problem. Finally, the motivation, objectives and outline of this thesis are presented.
- **Chapter 2: Direct Interval Prediction for Wind Power and Wind Speed** - This chapter presents interval forecast method for wind power and wind speed, respectively. First, interval prediction based on RNN model is proposed to construct

PIs with LUBE method for wind power. In addition, a novel comprehensive cost function with a new PI evaluation index is designed with the purpose of enhancing the model training. To tune the parameters of RNN prediction model, the dragonfly algorithm with a linearly random weight update method is introduced as the optimization tool. The performance of the proposed prediction model is validated by a case study using a real world wind power dataset. Second, a multiobjective interval prediction method based on wavelet neural network (WNN) for short-term wind speed forecast is proposed. This method can generate a set of Pareto optimal solutions which represents a set of prediction models that can directly construct the prediction intervals. An advanced multiobjective evolutionary algorithm, preference inspired co-evolutionary algorithm using goal vectors, is investigated to train the WNN model. Two case studies are carried out with real wind speed data of Victoria and Edmonton in Canada to justify the effectiveness of the proposed method. The numerical results also show the superiority of the proposed forecast approach compared with some benchmark methods.

- **Chapter 3: Chance-constrained Energy Management for Islanded Microgrids** - In this chapter, a chance-constrained energy management model is developed for an islanded microgrid, which includes distributed generators, energy storage system and renewable generation like wind power. The objective function of this model consists of generation cost, emission cost and degradation cost of storage system. To capture the uncertainty of renewable generation, a novel ambiguity set is introduced without knowing its probability distribution or exact moment information. With this ambiguity set, the chance constraint is processed with DRO method and the energy management problem is reformulated as a tractable second-order conic programming (SOCP) problem. The proposed approach is tested with a case study and simulation results indicate that it is effective and reliable. Moreover, the comparison with the method based on known moment information and some other methods is also conducted to show the performance of the proposed approach.
- **Chapter 4: Data-driven Distributionally Robust UC and ED** - Based on the research in chapter 3, distributionally robust UC and ED problem are studied in this chapter, respectively. First, a data-driven distributionally robust chance-constrained (DDRC) UC model is developed. The proposed two-stage UC model focuses on the commitment decision and dispatch plan in the first stage, and considers the worst-case expected cost for possible power imbalance or re-dispatch in the second stage. For uncertainty modeling of wind power, a distance-based ambiguity set is designed which can be constructed in a data-driven manner. Based on the ambiguity set, the original complicated UC problem is reformulated to a tractable optimization

problem which is then solved by the column-and-constraint generation (C&CG) algorithm. The performance of the the proposed approach is validated by case studies with different test systems including the IEEE 6-bus test system, modified IEEE 118-bus system and a practical-scale system, especially the value of data in controlling the conservativeness of the problem. Similarly, a two-stage data-driven distributionally robust energy and reserve dispatch problem with uncertain wind power is studied. A different distance-based ambiguity set is proposed, based on which the second-stage worst-case expectation of the problem is reformulated to a combination of conditional value-at-risk (CVaR) and an expected cost with respect to a reference distribution. Thus, the proposed two-stage DR model becomes a two-stage stochastic optimization problem which can be readily solved. Case studies are also carried out to verify the effectiveness of the proposed approach.

- **Chapter 5: Multistage Energy Management for Microgrids** - Microgrids can be operated in islanded or grid-connected mode. In this chapter, multistage energy management for grid-connected microgrids are investigated based on RO and DRO technique, respectively. First, a novel multistage robust energy management model for grid-connected microgrids is developed which considers the uncertainty of renewable generation and load demand. To solve this complex and computationally difficult problem, a robust version of dual dynamic programming method is proposed which includes a forward pass and a backward pass procedure and has a similar framework with the common stochastic dual dynamic programming (SDDP) method. Based on real datasets, a case study is carried out to validate the effectiveness of the proposed model and solution methodology. Numerical results show that the proposed approach can effectively achieve the robust optimal solution, and the comparison with other methods also testifies the advantage of the proposed multistage robust model. Second, multi-period energy management for CCHP microgrids is studied with DRO technique. Specifically, a two-stage multi-period distributionally robust energy management model is proposed which considers the non-anticipativity of uncertainty in dispatch process. A new second-order conic representable ambiguity set is designed for uncertain wind power. To address this complicated problem, linear decision rule (LDR) approximation is adopted and the proposed problem is transformed into a tractable mixed-integer second-order conic programming (MISOCP) problem. Case studies are conducted with real-world data to validate the performance of the proposed approach, particularly, the comparison with robust optimization and deterministic method are carried out to show the effectiveness of the proposed method.

- **Chapter 6: Conclusions and Future Work** - The contributions of this research and

the future work are summarized in this chapter.

2

Direct Interval Prediction for Wind Power and Wind Speed

Acronyms

ACE	Average Coverage Error
ACF	Autocorrelation Function
ARIMA	Autoregressive Integrated Moving Average
AWD	Accumulated Width Deviation
CWC	Coverage Width-based Criterion
DA	Dragonfly Algorithm
GA	Genetic Algorithm
GPR	Gaussian Process Regression
LUBE	Lower Upper Bound Estimation
MLP	Multilayer Perceptron
NARX	Nonlinear Autoregressive with Exogenous Inputs
NCWC	New Coverage Width-based Criterion
NN	Neural Network
NSGA-II	Nondominated Sorting Genetic Algorithm II
PACF	Partial Autocorrelation Function
PICP	Prediction Interval Coverage Probability
PICEA-g	Preference-inspired Coevolutionary Algorithm Using Goal Vectors
PIMSE	Prediction Interval Mean Squared Error
PINAW	Prediction Interval Normalized Average Width
PINC	Prediction Interval Nominal Confidence
PSO	Particle Swarm Optimization
PM	Polynomial Mutation
QR	Quantile Regression
RBF	Radial Basis Function
RNN	Recurrent Neural Network

SA	Simulated Annealing
WNN	Wavelet Neural Network

2.1 Introduction

With the development of advanced generation technologies, there has been an enormous increase in the amount of renewable generation such as wind power in recent years. Different wind power forecast methods have been proposed aimed at reducing its uncertain impact on power system operation, and wind energy forecast consists of direct wind power forecast and indirect wind speed forecast. For conventional point forecast methods, only a deterministic forecast value can be generated for a certain time step without any indication of the associated uncertainty [31] which may impact the following system operation with this value. Therefore, probabilistic forecast methods including interval prediction have attracted more attention recently which can include the forecast uncertainty [32].

Interval forecast is one of the most popular probabilistic forecast methods, and a nonparametric procedure LUBE method is proposed recently based on NN model. NN-based LUBE interval prediction method has been widely studied due to the light computational burden and no distribution assumption. However, such models are all developed based on feedforward NN model. Considering the advantage of RNN model, the RNN-based LUBE method is first investigated for short-term wind power interval prediction, and a new intelligent algorithm called DA algorithm is introduced to optimize the RNN model.

As a direct source of wind energy, wind speed interval forecast is also studied in this chapter. Currently, most of the research works about wind speed interval prediction adopt a single-objective optimization framework. However, interval prediction is actually a multiobjective optimization problem, and only a few studies focus on the multiobjective interval prediction with NN model. Consequently, improvement for multiobjective interval prediction may be achieved from both the forecast model and optimization algorithm. In this work, short-term wind speed interval prediction is performed based on WNN model in a multiobjective framework, and a novel multiobjective evolutionary algorithm is investigated to train the NN model with two objectives.

2.2 Interval Prediction for Wind Power

The direct interval prediction for wind power is studied in this section. The background of PI is first introduced in subsection 2.2.1, followed by the presentation of the proposed RNN interval prediction model and the DA algorithm in subsection 2.2.2. The performance of

the proposed forecast approach is validated by case studies in subsection 2.2.3.

2.2.1 Background of PI

The construction of a PI is to estimate the upper and lower bound of an interval with a confidence level which shows the accuracy and the quality of PIs need to be evaluated by some measures. In this subsection, the PI evaluation indices are introduced first. Then, the NN-based LUBE interval prediction method is explained.

2.2.1.1 PI Evaluation Indices

A high quality PI is expected to have larger reliability and narrower width. To assess these two aspects of PIs, two indices, PI coverage probability (PICP) and PI normalized average width (PINAW) [37], are mostly employed to quantitatively measure the forecast intervals. PICP which is also called PI confidence level is used to show the probability that target values will be covered by the forecasted intervals. Obviously, a larger PICP value indicates that more targets will lie in the constructed PIs. This index is usually considered as the critical indicator of PIs and it can be mathematically defined as follows:

$$\text{PICP} = \frac{1}{N} \sum_{i=1}^N \delta_i \quad (2.1)$$

where N is the number of test samples and δ_i is a binary value which is described as follows:

$$\delta_i = \begin{cases} 1, & y_i \in [L_i, U_i] \\ 0, & y_i \notin [L_i, U_i] \end{cases} \quad (2.2)$$

where y_i is the target, L_i and U_i are lower and upper bound of the PI, respectively. Generally, the PICP value should be greater than the preassigned confidence level in the training process, otherwise the PIs are invalid and should be discarded.

Although PICP index is the key feature of PIs, we can not just focus on this objective and ignore the interval width. With sufficiently wide intervals, the PICP objective can be easily achieved. However, very wide intervals hardly yield any valuable information and may be useless in practice. Therefore, a quantitative measure of interval width, PINAW, is defined to limit the interval width, as follows:

$$\text{PINAW} = \frac{1}{N \cdot Rg} \sum_{i=1}^N (U_i - L_i) \quad (2.3)$$

where Rg means the range of the targets (difference between the maximum and the minimum).

PICP and PINAW can only evaluate one of the aspects of PIs, respectively. In order to assess the overall quality of PIs, the index coverage width-based criterion (CWC) which is a comprehensive cost function is designed [38]. Furthermore, a modified CWC cost function is proposed to overcome the multiplication drawback [120] as follows:

$$\text{CWC} = \text{PINAW} + \gamma(\text{PICP})e^{-\eta(\text{PICP}-\mu)} \quad (2.4)$$

where $\gamma(\text{PICP})$ is a boolean function given by

$$\gamma = \begin{cases} 1, & \text{if PICP} < \mu \\ 0, & \text{if PICP} \geq \mu \end{cases} \quad (2.5)$$

where η and μ are two controlling parameters. The former is usually a large constant to penalize the invalid PIs, while the latter can be determined according to the nominal confidence level.

With the CWC function, the primary multi-objective problem can be transformed into a single-objective minimization problem. Although the CWC cost function is used as a comprehensive evaluation index, we can find that it is only determined by the estimated upper and lower bounds and the known information is not fully used in the training process. Similar to the frequently used mean squared error (MSE) index in point forecast, a new PI width evaluation criterion, PIMSE, is designed to make better use of the known target values in this study:

$$\text{PIMSE} = \frac{1}{N} \sum_{i=1}^N [(U_i - y_i)^2 + (L_i - y_i)^2]. \quad (2.6)$$

By introducing this index, a new CWC (NCWC) function can be developed for the training:

$$\text{NCWC} = \text{PINAW} + \gamma(\text{PICP})e^{-\eta(\text{PICP}-\mu)} + \text{PIMSE}. \quad (2.7)$$

With this new cost function, we can combine the unsupervised learning and supervised learning by using the known information in the training process. Besides, it is expected that a more symmetric interval, which is closer to the true confidence interval, will be obtained by minimizing the PIMSE index. Therefore, the NCWC is used to enhance the model training in this work.

In the model training process with the NCWC objective, the reliability index PICP will be the influential factor at the beginning stage due to the high penalty cost. If PICP is less than the predefined confidence level, the NCWC will be large regardless of the interval width. As the training continues, the PICP will become greater than the nominal confidence level and the sum of PINAW and PIMSE should be the dominant factor. Note that it is necessary to consider both criteria PINAW and PIMSE here as PINAW mainly

focuses on narrower intervals and PIMSE will make full use of the known information and make the intervals more close to the actual symmetric confidence intervals. If only PIMSE index is considered, we may get symmetric but wide intervals which will lead to a large PINAW value.

2.2.1.2 LUBE Method

The LUBE method is a nonparametric method that can directly construct PIs [9]. It is simple and fast to generate PIs without any assumption about the forecast errors. In previous work, LUBE method is implemented with a feedforward NN. The NN model has two output nodes including the upper and lower bounds of PIs. Actually, this method belongs to unsupervised learning since the upper and lower bounds are not known during the training process. In practice, it is better to have narrower PIs with a larger coverage probability. Therefore, the primary problem based on LUBE method is a multi-objective problem with two conflicting objectives. This multi-objective problem can be transformed into a single-objective problem by introducing the nonlinear and nondifferentiable CWC cost function and some basic constraints [9], [40]. To optimize the cost function in the NN training process, different gradient-free optimization methods such as simulated annealing (SA), particle swarm optimization (PSO) algorithm have been applied in previous study.

In each iteration of the training process, two outputs representing the lower and upper bounds of a PI are generated based on the model inputs. Then the two measures PICP and PINAW as well as the corresponding CWC cost function can be calculated for all training samples. As the training procedure continues with a certain optimization algorithm, the NN parameters are tuned gradually so that the PICP meets the predefined confidence level and the interval width PINAW decreases. When the maximum number of iterations is reached or there is no further improvement on the objective for a certain number of consecutive iterations, the model training terminates and the resulting optimal model can be used for construction of new PIs [121]. The key features of LUBE method includes simplicity, low computation cost and distribution-free assumption compared with traditional interval forecast methods. More details about the LUBE method can be found in [9].

2.2.2 RNN-based interval prediction model

As mentioned above, the LUBE method is a simple and efficient method to construct high quality PIs. Due to the easy implementation and low computational cost, LUBE method has become popular in quantifying forecast uncertainty in a very short time. It has attracted much attention and abundant research works are carried out based on feedforward NN model to do interval prediction. Compared with the feedforward

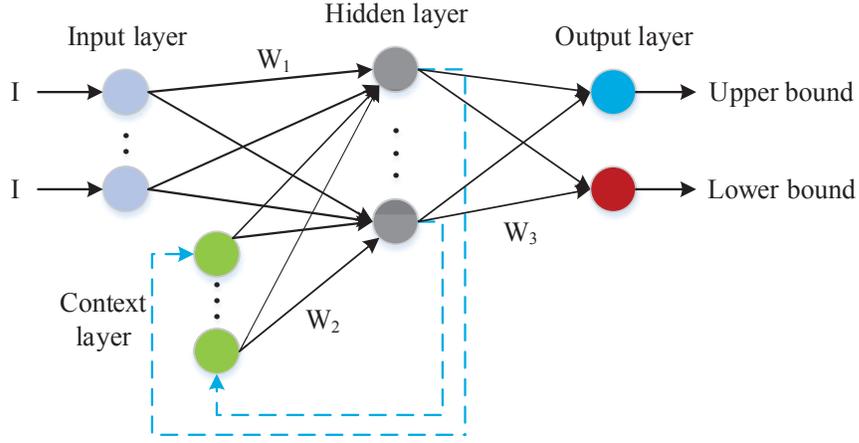


Figure 2.1: Schematic diagram of Elman network.

NN which can only learn a static input-output mapping relationship, an RNN with feedback structure behaves like a dynamic system which is more suitable to model temporal sequences. Therefore, the RNN-based interval prediction model for wind power generation is developed in this study. Two typical RNN models, Elman network and NARX model, are investigated.

2.2.2.1 Elman Network

As a first-order RNN, the Elman network is a simple recurrent network [122] which employs a context layer to feedback the outputs of the hidden layer. The context layer is a copy of the hidden layer outputs at the previous time step. Although simple in structure, it has the ability to perform complex tasks. Based on the Elman network, the interval prediction model can be developed and the schematic diagram of the model is shown in Fig. 2.1.

In this three layer Elman network model, there are two output nodes representing the upper and lower bounds of a PI. The only feedback is from the hidden layer to the context layer and the connection weights are constants, while the other feedforward weights are adjustable which could be optimized by intelligent optimization method. According to the Elman structure, the dynamic change of this model can be mathematically expressed as follows:

$$x(k) = \phi[W_1 I(k-1) + W_2 x_c(k) + b_1] \quad (2.8)$$

$$x_c(k) = x(k-1) \quad (2.9)$$

$$z_1(k) = f(W_3 x(k) + b_2) \quad (2.10)$$

where x and z_1 represent the output of the hidden layer and output layer, respectively, x_c

is the output of the context layer, $\phi(\cdot)$ is the transfer function which is usually nonlinear hyperbolic-tangent-sigmoid function, and $f(\cdot)$ is pure linear activation function.

2.2.2.2 NARX Network

NARX is another class of RNN that is suitable to model time series data. Unlike Elman network, the feedback in this input-output recurrent model is from the output to the input. The dynamic behaviour of NARX model can be mathematically described by [123]:

$$o(k+1) = g[o(k), \dots, o(k-d_o); u(k), \dots, u(k-d_u)] \quad (2.11)$$

where o is the output, u is the input, d_o and d_u are the output and input delay, respectively, $g(\cdot)$ is also a nonlinear mapping function which can be approximated by a standard MLP model. Without loss of generality, the typical three layer structure is used in this study. Thus, the NARX-based interval forecast model is similar to the previously described Elman forecast model and the schematic diagram is omitted.

2.2.2.3 Dragonfly Algorithm

Inspired by the behaviours of dragonflies, DA is a population-based optimization algorithm that was proposed in 2016 [124]. In DA, each dragonfly represents a promising solution for the optimization problem. In this study, the dragonfly or solution is a vector that consists of the NN connection weight values. These weights are adjusted to find the optimal values by minimizing the cost function. To solve the problems with DA, each dragonfly should have two vectors: position (P) and step (V). The step vector here is similar with the velocity vector in PSO, while the position updating of individuals is determined by five main behaviours including separation, alignment, cohesion, attraction towards food and distraction from an enemy. These behaviours can be described as follows [124]:

$$Se_i = - \sum_{j=1}^{Num} P_j - P_i \quad (2.12)$$

$$Al_i = (\sum_{j=1}^{Num} V_j) / Num \quad (2.13)$$

$$Co_i = (\sum_{j=1}^{Num} P_j) / Num - P_i \quad (2.14)$$

$$Fo_i = P^+ - P_i, En_i = P^- + P_i \quad (2.15)$$

where P_i is the position of current individual, P_j and V_j represent the j -th neighbour individual's position and corresponding velocity, respectively, Num is the number of neighbour individuals, P^+ and P^- denotes the position of dragonflies' food and enemy, respectively.

Based on the above behaviours of the dragonflies, the position vector of each individual can be updated as follows:

$$P^{iter+1} = P^{iter} + V^{iter+1} \quad (2.16)$$

$$V^{iter+1} = (sSe_i + aAl_i + cCo_i + fFo_i + eEn_i) + wV^{iter} \quad (2.17)$$

where $iter$ is the iteration number, $s, a, c, f, e,$ and w are corresponding weight coefficients which control the exploration and exploitation search during the optimization process. The above updating rules are applicable for individuals with neighbours. When the individuals have no neighbours, the Levy flight which is a random walk [125] is used to improve their exploration and stochastic behaviours, and the position update is as follows:

$$P^{iter+1} = P^{iter} + Levy(dim) \times P^{iter} \quad (2.18)$$

$$Levy(dim) = 0.01 \times \frac{c_1 \times \rho}{|c_2|^{\frac{1}{\lambda}}} \quad (2.19)$$

where dim is the dimension of the vector P , c_1 and c_2 are two random numbers in $[0,1]$, respectively, λ is a constant value and ρ can be calculated by the following equation:

$$\rho = \left(\frac{\Gamma(1 + \lambda) \times \sin(\frac{\pi\lambda}{2})}{\Gamma(\frac{1+\lambda}{2}) \times \lambda \times 2^{\frac{\lambda-1}{2}}} \right)^{1/\lambda} \quad (2.20)$$

where $\Gamma(n) = (n - 1)!$.

The DA has been demonstrated to perform better than other well-known optimization algorithm such as PSO and genetic algorithm (GA) on the test functions [124]. Therefore, it is introduced to tune the RNN parameters by optimizing the comprehensive cost function in this study. In DA, the inertia weight w is adjusted adaptively by the typical linearly decreasing manner. In order to enhance the total search ability of this algorithm, inspired by the random inertia weight update method [126], we can further improve the weight update by using Levy flight as follows:

$$w^{iter} = w_{max} - (iter/Maxt) \times (w_{max} - w_{min}) \quad (2.21)$$

$$w^{iter+1} = w^{iter} + Levy(dim) \quad (2.22)$$

where w_{max} and w_{min} are the maximum and minimum weights, respectively, $Maxt$ is the maximum iteration number, the Levy function is the same as that defined in (2.20).

2.2.2.4 Model Implementation

Based on the Elman and NARX network model, the LUBE method was implemented to construct PIs with DA optimization algorithm. The single-objective problem is formulated with the NCWC cost function. The model implementation flowchart is shown in Fig. 2.2

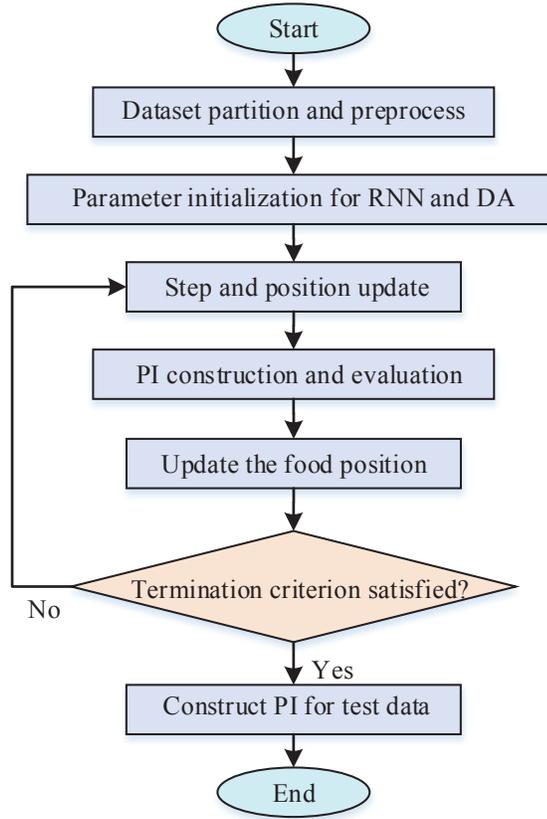


Figure 2.2: Model flowchart for PI construction.

and details of the main steps are discussed below.

1) Dataset partition and preprocess. For the forecast model, the main input is the historical wind power data. The original dataset should be split into training data, validation data and test data. The training and validation data are combined to train the model, while the test data are used to verify the model's generalization ability. In order to accelerate the model training process, the original data are usually normalized to $[-1,1]$ after partition.

To construct PIs with the RNN model, the original time series data should be transformed into a suitable form for training the model. A dynamic system in discrete time can be depicted as:

$$X(t+1) = \mathbf{F}(X(t)) \quad (2.23)$$

where $X(t)$ is the system state at time step t and \mathbf{F} is nonlinear vector valued function. Since the wind power and speed data are volatile and chaotic [127] from a dynamic system, the state space reconstruction technique with the delay embedding theorem [128] was employed to process the original data. By this theorem, the one-dimensional chaotic data

is supposed to compress the information of higher dimension. Hence, the time series data $X(t)$ can be reconstructed as follows:

$$X_0(t) = [X(t), X(t - \tau), \dots, X(t - (m - 1)\tau)] \quad (2.24)$$

where τ is time delay and m is the embedding dimension. Therefore, creating the delay embedding comes down to finding the optimal values of parameters τ and m . For a given dataset, these two parameters can be determined by the mutual information function and false nearest neighbour method [129].

Once the delay embedding was completed, the reconstructed time series was generated and it was used to train the RNN model for one-step ahead prediction task. In this case, the number of model input was also determined which equals the embedding dimension.

2) Parameter initialization. There are two sets of parameters corresponding to the DA algorithm and RNN model. In DA algorithm, the step and position vectors are initialized with small random numbers. The connection weights of RNN model are represented by the position vectors of dragonflies, thus their initialization is finished. To find the optimal number of hidden nodes in RNN model, five-fold cross-validation method can be applied with the training dataset.

3) Update step and position vectors. The position and step of each dragonfly are updated according to the equations (2.16) and (2.17). The individual with better fitness (smaller cost function value in this study) will be retained.

4) Model evaluation. Each individual corresponds to one prediction model. With the connection weights, PIs can be constructed and the corresponding measures PICP, PINAW and PIMSE can be calculated. The index NCWC is considered as the fitness in RNN training process. The individual with the best fitness is recorded as dragonflies' food source and its position vector represents the best model weights.

5) Termination criterion. The training is terminated when the maximum iteration is reached in this work. If the termination condition is not met, then it will return to update the step and position vectors.

6) PI construction for test dataset. When the training is completed, we can get the optimal connection weights for RNN model. With this optimal prediction model, PI construction can be easily accomplished for the test data. The relevant indices are also calculated to evaluate the PI quality.

2.2.3 Realistic Wind Energy Prediction Case Study

To validate the forecast performance of the proposed RNN-based LUBE method, a realistic case study of wind energy interval prediction is carried out in this subsection. First of all, the data source and relevant parameter settings are explained. Then, the numerical results and discussion are presented.

2.2.3.1 Data Set

The historical wind power data is taken from the Adelaide wind farm located in Ontario, Canada and it can be obtained from the IESO website [130]. Ontario takes the lead in the clean wind power utilization with 4781 MW of installed capacity, which supplies about 5% of the total electricity demand in the province [131].

The chosen dataset consists of hourly wind power data in MW from 1 January 2016 to 7 April 2017. During this time period, the wind farm performance is assumed to be normal and there is no missing or false data. The whole dataset is further partitioned into three subsets for training, validation and test. The training set and validation set account for about 80% of the whole dataset in this study, i.e., the whole year data in 2016. The remaining data are used to test the model's prediction performance.

2.2.3.2 Parameter Settings

As mentioned above, two sets of parameters about RNN model and DA algorithm should be determined in the proposed prediction model. For the RNN model, we should design the best structure by finding the optimal number of input nodes and hidden nodes. The number of input nodes is related to the dimension of delay vectors which can be determined by state space reconstruction technique. According to the delay embedding approach, a time series is a series of observations of a dynamic system and the forecast is about forecasting the system's state. To forecast the system, we should construct a state space that is equivalent to the original one by using a small set of the most recent previous observations [128]. By delay embedding theorem, we need to find two parameters: the embedding dimension m which represents the size of the set of most recent observations and the time delay τ which means the optimal autocorrelation level in each delay vector. In this study, τ is determined to be 16 by the mutual information method where the first local minimum occurs and m is 7 by false nearest neighbour method, which can be accomplished by the utility functions *mutual* and *false_nearest* in TISEAN toolbox [132]. In this case, the dimension of reconstructed delay vectors is 7 and the number of input nodes for the RNN model is also 7. After determining the value of τ and m , delay vectors can also be obtained for the prediction model.

Table 2.2: Cross-validation results of Elman model

Nodes	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)
3	94.88	74.29	1.8427	2.5856	74.29
5	94.94	65.66	1.5107	2.1672	65.66
8	93.90	69.04	1.6974	2.3878	69.04
10	94.13	73.69	2.0186	2.7555	73.69

As for the number of hidden layer nodes, five-fold cross-validation method is employed to explore it with the training dataset in this work. This method is frequently used in the literature [9] which can help establish a stable model. The cross-validation results of some typical hidden nodes numbers for Elman network are shown in Table 2.2. From this table, we can see that the model has the best performance according to the NCWC index when the number of hidden nodes is 5. Therefore, the optimal structure of Elman network is 7-5-2 and the number of nodes in the context layer is also 5. The optimal structure of NARX model can be determined similarly and the number of hidden nodes is also 5 in this study. In addition, for the typical three layer structure, the common hyperbolic tangent and pure linear activation functions are used in the hidden and output layer, respectively.

Another parameter set is about the optimization algorithm. It is suggested that the weight coefficients in (2.17) can update in an adaptively tuning method to balance the exploration and exploitation during the optimization [124]. Generally, the inertia weight w varies from 0.9 to 0.4 [133]. In this study, its range is set to be [0.7, 1] by trial and error method. In the NCWC or CWC cost function, the controlling parameter μ is set to be the nominal confidence level 0.9 and η equals 50 [31]. In addition, the population size is 30 and the maximum number of iteration is set to be 1000 during the optimization process.

2.2.3.3 Test Results

The proposed Elman network and NARX interval prediction model have been applied to construct wind power PIs. After the training process, the dragonflies' food position vector obtained by DA corresponds to the optimal prediction model. With the best model, PIs can be constructed for the wind power test dataset. The performance measures are also calculated to evaluate the obtained PIs' quality including PICP, PINAW, PIMSE, NCWC and the CWC cost function which is frequently used in the literature.

In the training process, the variation of the best individual's objective function is shown in Fig. 2.3. As shown in this figure, the NCWC function of both Elman and NARX model decreases dramatically in the first few generations to get the satisfied PICP. When PICP satisfies the nominal coverage probability, more attentions are paid to the interval

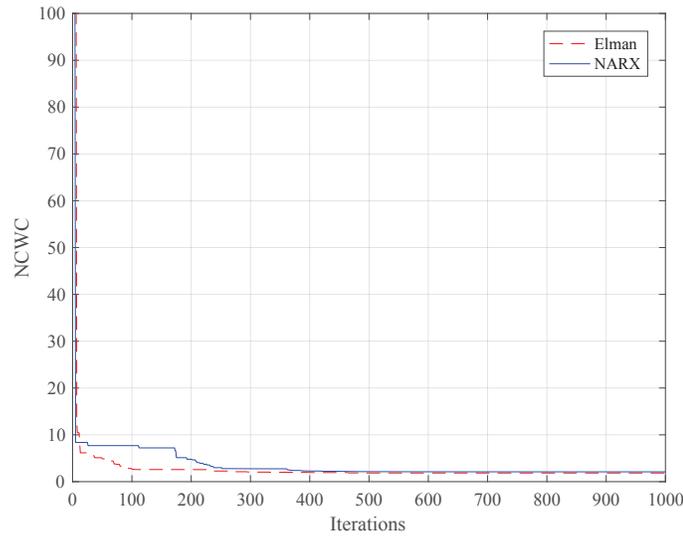


Figure 2.3: NCWC of the best individual in the training process.

width, i.e., to minimize the sum of PINAW and PIMSE. As the optimization proceeds, the NCWC objective continues to decrease and it converges to the optimal value eventually for both models. The convergence validates the strong search ability of DA algorithm with random linear weight update method. It can also be seen that the optimal objective value of Elman network is a little better than that of NARX model.

For Elman network model, the PI construction results for test data are shown in Fig. 2.4. For better visualization, the PIs for the last week of test data are also given in Fig. 2.5. From the results, we can see that most of the target values (the green dash line with star) lie in the constructed upper and lower bounds. As shown in Fig. 2.5, both the predicted upper and lower bounds have a similar trend with the real data, which implies that the prediction model can capture the dynamic feature of wind power data well. Note that the lower bound that is below zero is set to be zero in this study. In addition to the similar trend of those three lines, we can also find that most of the constructed intervals are approximately symmetric about the targets resulting from the involvement of PIMSE in the training process. A symmetric interval is more closer to the true confidence interval which can be obtained based on the known distribution information.

In order to verify the repeatability of the prediction model and get convincing forecast results, the case study with Elman network is repeated for five times. Results of each run including PICP, PINAW, PIMSE, NCWC and CWC are shown in Table 2.3. As can be seen from this table, the PICP values of all five runs satisfy the nominal coverage probability (90%), which indicates that the prediction model is reliable since the PICP index is usually considered as the key feature of the PIs. The results are also consistent as the variances of

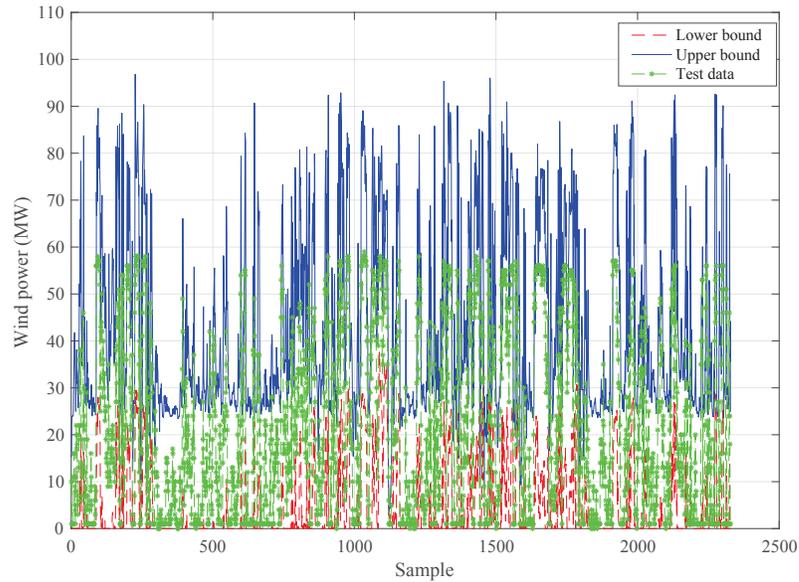


Figure 2.4: Optimal PIs of Elman model for test data.

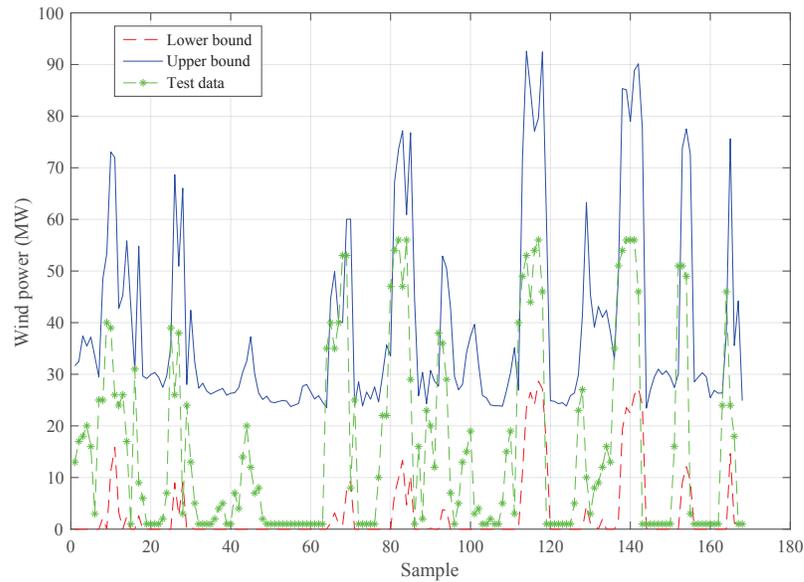


Figure 2.5: Optimal PIs of Elman model for the last week of test data.

these measures are quite small, e.g., the standard deviation of CWC for five runs in Table 2.3 is 1.7912. Note that the PIMSE value here is calculated from the normalized data. As the CWC cost function is a comprehensive index which is frequently used in the literature, we will take this index for convenient comparison later. The median CWC value (63.71) instead of the best one is used to represent the average performance of Elman prediction model. Moreover, the median value is less influenced by outliers compared with the mean value.

Table 2.3: PI construction results of Elman model

No.	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)
1	93.86	61.95	1.2362	1.8558	61.95
2	93.90	65.20	1.3421	1.9940	65.20
3	93.81	63.71	1.4188	2.0559	63.71
4	93.94	66.43	1.3564	2.0207	66.43
5	94.24	62.90	1.3364	1.9654	62.90

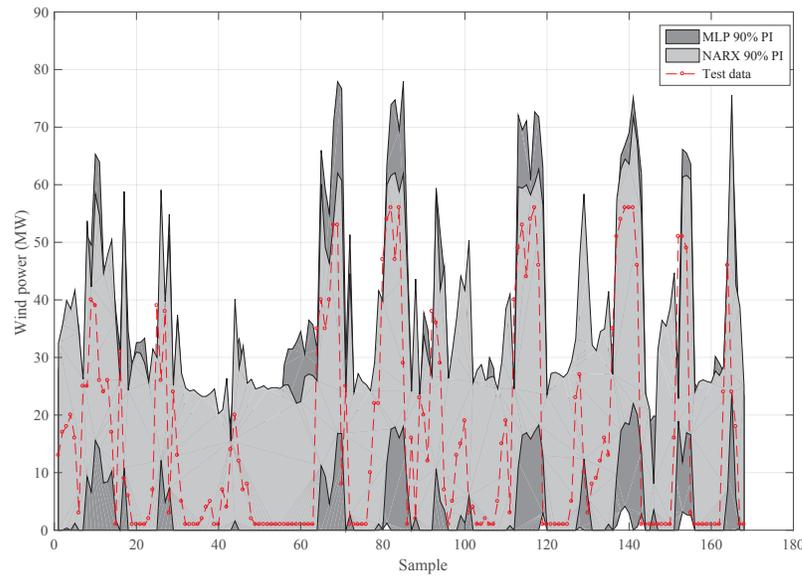


Figure 2.6: Optimal PIs of NARX and MLP model.

Table 2.4: PI construction results of NARX model

No.	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)
1	94.03	65.23	1.5537	2.2060	65.23
2	94.12	68.78	1.6633	2.3511	68.78
3	94.20	69.72	1.5718	2.2690	69.72
4	93.47	63.23	1.3599	1.9922	63.23
5	93.04	68.63	1.6428	2.3290	68.63

As for NARX model, the PI construction is similar with that of Elman model. Hence, the results of only the last week are given in Fig. 2.6 for simplicity, from which we can obtain similar conclusions. The NARX model is also run for five times and the results are shown in Table 2.4. As can be seen, the NARX model is also reliable and its average performance is represented by the median CWC value as well. The standard deviation of CWC for NARX model in Table 2.4 is 2.7613.

Table 2.5: PI construction results for different seasons

Season	Elman model					NARX model				
	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)
Spring	92.31	61.95	1.3643	1.9838	61.95	95.10	60.99	1.2494	1.8593	60.99
Summer	96.01	57.93	1.2364	1.8157	57.93	95.30	55.80	1.1230	1.6810	55.80
Autumn	91.73	66.30	1.4424	2.1054	66.30	91.25	65.68	1.8004	2.4572	65.68
Winter	94.93	64.50	1.3238	1.9687	64.50	95.17	67.02	1.3988	2.0690	67.02

Table 2.6: PI construction results of MLP model

No.	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)
1	95.92	70.65	2.2352	2.9416	70.65
2	96.13	74.10	1.6891	2.4302	74.10
3	96.99	76.97	1.7817	2.5514	76.97
4	96.56	72.36	1.6312	2.3547	72.36
5	95.53	80.57	2.4104	3.2161	80.57

To further validate the effectiveness of the proposed method, the dataset is expanded according to different seasons which are tested with the proposed method, respectively. The average test results of four seasons with Elman model and NARX model are given in Table 2.5. From this table, we can see that both Elman model and NARX model can get good prediction results for different seasons. Moreover, the forecast results of spring and summer are better than those of autumn and winter by a comparison of CWC and NCWC values. This may be due to abrupt change of wind speed in autumn and winter seasons.

2.2.3.4 Comparison with benchmark models

For comparison purpose, some benchmark models are employed to construct PIs with the same dataset including naive method, ARIMA model, Gaussian process regression (GPR), QR and feedforward MLP model [40]. As a fundamental model, a typical three layer MLP was designed and the implementation procedure for PI construction is similar with that of RNN model. DA is also utilized to optimize the model with the same parameter settings. The optimal number of hidden nodes is determined to be 5 by five-fold cross-validation method. In addition, the activation functions are hyperbolic tangent and pure linear function for the hidden layer and output layer [40], respectively. The PI results of MLP model for the last week are shown in Fig. 2.6, where we can intuitively see that its interval width is larger than that of NARX model. To get quantitative results, MLP model is also repeated for five times and the median result is used to illustrate its average performance as shown in Table 2.6. From Table 2.6, it can be observed that the performance of MLP model seems not to be very stable due to the large standard deviation of CWC which is 3.9238.

ARIMA is a classical model used in time series forecast. Generally, ARIMA model

performs better for one-step ahead forecast. In this study, the seasonal ARIMA model [134] is considered for direct one-step prediction [40]. The naive model is another well-known benchmark and it is similar to the persistence model in point forecast. According to this method, the forecast for the next step is generated based on the previous values. For example, the maximum and minimum values of the previous 20 samples are considered as the upper bound and lower bound for next step, respectively [120]. GPR model is an effective nonlinear prediction method which can be applied in many areas including regression and classification [135]. It adopts the Gaussian white noise assumption in the model and it is suitable to handle small sample problem. In addition, QR is another common statistical method that can be used for probabilistic forecast [136].

The comparison of PI construction results is summarized in Table 2.7. From Table 2.7, it can be observed that our proposed RNN prediction model outperforms the other benchmark models except for the GPR model. Note that in the GPR forecast model, the noise of the data is assumed to follow Gaussian distribution and the joint distribution of any finite number of variables is also Gaussian. However, the Gaussian distribution assumption is usually not applicable in practice. On the contrary, our method makes no assumptions on the data noise. In addition, the calculation time of GPR model for large dataset is very long and it is 3152.69s in our case which is almost close to the prediction time scale of one hour. Therefore, the application of GPR forecast model may not be feasible in practice.

In addition to the GPR model, Elman model achieves the lowest CWC value as well as the NCWC value, followed by the NARX model. All of the PICP values can satisfy the preassigned nominal coverage probability 90% except for the QR and naive model. Although the PINAW values are quite low for QR and naive model, their CWC values are very large due to the penalty on the unsatisfied PICP or low reliability. When PICP is satisfied, the PI quality depends on the interval width, which can be revealed by the comprehensive index CWC. By comparing the CWC values in Table 2.7, we can see that the PI quality has been significantly improved with the proposed RNN interval forecast model. The percentage improvements of Elman model are 14.02%, 21.74%, 62.71% and 84.90% in comparison with MLP, ARIMA, QR and naive model, respectively. They are 7.38%, 15.70%, 59.83% and 83.73%, respectively, for NARX model.

To verify the performance of DA optimization algorithm, it is compared with other population based approaches including GA and PSO. For the GA algorithm, the real-coded technique is employed and the probability of crossover and mutation are set to be 0.9 and 0.1, respectively [137]. The parameters of PSO algorithm are taken from [40]. With the same implementation procedure, we can obtain the average prediction results of Elman

Table 2.7: Comparison of proposed model with benchmark models

Method	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)
Elman	93.81	63.71	1.4188	2.0559	63.71
NARX	93.04	68.63	1.6428	2.3290	68.63
MLP	96.13	74.10	1.6891	2.4302	74.10
ARIMA	92.01	81.41	2.0675	2.8816	81.41
GPR	90.51	43.12	0.6897	1.1209	43.12
QR	89.86	63.75	1.3373	3.0459	170.87
Naive	87.46	65.28	1.6303	5.8493	421.90

Table 2.8: PI construction results with different algorithms

Algorithm	Elman model						NARX model					
	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)	time(s)	PICP(%)	PINAW(%)	PIMSE	NCWC	CWC(%)	time(s)
DA	93.81	63.71	1.4188	2.0559	63.71	785.75	93.04	68.63	1.6428	2.3290	68.63	770.81
GA	94.85	74.78	1.8622	2.6100	74.78	771.49	94.07	71.48	1.7260	2.4408	71.48	770.80
PSO	93.21	86.30	2.5299	3.3930	86.30	1051.09	95.88	76.40	1.9980	2.7620	76.40	1022.86

model and NARX model with GA and PSO algorithm which are summarized in Table 2.8. The training terminates when the maximum number of iterations is reached for all three algorithms, and the corresponding training times are also given in Table 2.8. As can be seen, the prediction results with the DA algorithm are the best according to the CWC and NCWC values. GA and PSO are easy to be trapped in local optima when the training terminates. In addition, the training time of DA and GA are almost the same, while the time of PSO is longer. The training time of the proposed method is about 10 minutes which is much less than the forecast time scale of one hour. In addition, when the training process terminates, we can use the well-trained model offline, and the forecast time will be shorter. In other words, the proposed method can be used for real-time forecast. Note that the execution time is related to the optimal structure of our model. If the number of hidden nodes in the recurrent model increases, then the average running time will also be longer.

2.2.3.5 Discussion

As the forecast time range prolongs, the forecast accuracy significantly decreases as a result of more uncertainties. Therefore, only one-step ahead prediction is considered in this work which is more accurate in practice. However, multi-step forecast may also be possible for our model with a proper data preprocess method. For wind power point forecast, we know that some other inputs such as NWP data and nearby wind speed are also considered except for the historical data to reduce the forecast error. These data may also be useful for interval prediction which is worth studying in the future. The delay embedding approach can be considered as a feature selection operation in our method.

Also, other feature selection techniques can be used to process the input data which may help determine the most informative features and achieve higher quality PIs. Although it is possible to improve the forecast method as mentioned above, we can still find the superiority of our model from the comparison with other benchmark models.

Generally, the quality of PIs is evaluated by the coverage probability and interval width which are two conflicting objectives during PI construction process. To shorten the interval width may deteriorate the coverage probability and vice versa. Actually, the quality of PIs may be influenced by different factors such as data characteristics and forecast model. For different datasets, the PI results generated by the same forecast model are different as can be seen from Table 2.5. For forecast model, we may further study some modified RNN models or design a better comprehensive cost function in the existing single-objective framework to improve the PIs' quality in the future. In addition, considering the multiobjective characteristic of PI construction, a multiobjective problem formulation for interval prediction may also be a good choice for future research. In a single-objective framework for PI construction, the minimization of the comprehensive cost function NCWC or CWC is essentially to find an optimal compromise between these two aspects but it may not always balance them well. However, with a multiobjective framework, a set of Pareto optimal solutions can be generated and we can select a satisfactory one from them according to posterior preference information.

Due to the intrinsic randomness of DA algorithm in the training process, we may get different prediction results for each run. Therefore, to avoid a suboptimal solution, it is better to repeat the training several times and take the average as the optimal solution in a practical application. In addition, the PI results can be utilized in different ways in practice. For example, they can either be directly used in robust optimization problem of power systems which only needs the lower and upper bound, or be transformed into deterministic point forecast values with convex combination of the lower and upper bound [138].

PI construction time is a critical factor in practice. Although the training time of the proposed RNN model is a little longer than MLP model, the test time, which is more useful for online applications, is almost the same with that of MLP model. The average PI construction time of Elman and NARX model in this study are 0.0770s and 0.0737s, respectively, which are very fast. All the experiments in this study are implemented with MATLAB software on a PC with hardware configuration of Intel Core TM i7-6700 CPU 3.40 GHz and 8 GB of RAM.

2.3 Multiobjective Interval Prediction for Wind Speed

In this section, interval prediction for short-term wind speed is studied in a multiobjective framework based on WNN model. The multiobjective problem formulation for PI construction is first described in subsection 2.3.1. Then the interval prediction methodology including the WNN model and optimization algorithm is illustrated in subsection 2.3.2, and subsection 2.3.3 provides the numerical results and comparison based on real datasets.

2.3.1 Problem Formulation

Some new PI evaluation indices which are the basis to assess the PI quality are introduced and the problem formulation of multiobjective interval prediction is presented in this subsection.

2.3.1.1 PI and Evaluation Indices

The basic concept of PI and some common evaluation indices such as PICP and PINAW have been introduced in section 2.2. In addition to the general indices introduced above, there are some other indices used in the literature, such as the average coverage error (ACE), interval score [35] and the accumulated width deviation (AWD) [139]. In PI construction process, the PI nominal confidence (PINC) is usually predefined, and the PICP index aims to approach PINC as closely as possible. In this case, ACE is defined as the difference between PICP and PINC as follows:

$$ACE = PICP - PINC. \quad (2.25)$$

ACE can be utilized to assess the quality of PIs with respect to the reliability. The smaller the absolute value of ACE is, the better the quality of derived PIs is. Another index AWD can also be used for reliability evaluation of PIs. By comparing the position of the real targets and PIs, relative width deviation can be calculated, and AWD is the sum of relative width deviation as shown below:

$$AWD_i = \begin{cases} \frac{L_i - y_i}{U_i - L_i}, & y_i < L_i \\ 0, & y_i \in [L_i, U_i] \\ \frac{y_i - U_i}{U_i - L_i}, & y_i > U_i \end{cases} \quad (2.26)$$

$$AWD = \frac{1}{N} \sum_{i=1}^N AWD_i \quad (2.27)$$

where y_i represents the real target. AWD index penalizes the PIs if the real targets are not enclosed, and a smaller AWD indicates higher PI quality. Note that the two basic indices PICP and PINAW are used as the objectives of the formulated multiobjective problem in this study.

To assess the overall performance of the PIs including the calibration and sharpness, another comprehensive index interval score is introduced. Denote the width of a PI as θ_i which is calculated by $\theta_i = U_i - L_i$, then the interval score S_i of a specific interval is defined as follows:

$$S_i = \begin{cases} -2\alpha\theta_i - 4(L_i - y_i), & y_i < L_i \\ -2\alpha\theta_i, & y_i \in [L_i, U_i] \\ -2\alpha\theta_i - 4(y_i - U_i), & y_i > U_i \end{cases} \quad (2.28)$$

where α is related to the nominal confidence level ($100(1 - \alpha)\%$). Based on the interval score of each forecast point, the overall interval score can be calculated as follows:

$$\text{Score} = \frac{1}{N} \sum_{i=1}^N S_i. \quad (2.29)$$

From the definition, we can find that a lower absolute value of the interval score indicates higher quality of PIs. The Score index can be used to assess the overall skill of PIs since it considers all aspects of PI evaluation [35]. Note that a lot of evaluation indices for PIs have been studied in previous literature and several common indices are employed in this work. Some other indices such as the continuous ranking probability score [140] may also be investigated for future research.

2.3.1.2 Multiobjective Problem Formulation

According to the performance indices introduced above, the PI construction is actually an optimization problem which aims at high quality PIs. As CWC is a comprehensive evaluation index, unconstrained single-objective optimization problem based on it was first proposed as follows [9]:

$$\text{Minimize: } \text{CWC}(w_{ij}) \quad (2.30)$$

where w_{ij} is the prediction model parameters to be tuned. Furthermore, taking the coverage probability as the fundamental requirement for valid PIs, constrained single-objective problem formulation was also proposed [40]. In this problem, PICP is constrained to be larger than the supposed confidence level, and the minimization of the PINAW value is the optimization objective.

Although the single-objective problem framework has been widely studied for interval prediction, the PI construction is essentially a multiobjective problem. The problem has two objectives: maximizing the reliability index and minimizing the width index, which are two conflicting objectives, i.e., boosting one objective will deteriorate the other one. Therefore, a multiobjective problem formulation is more appropriate to describe the PI construction problem and the interval prediction for wind speed is conducted in a multiobjective framework in this work. The primary multiobjective interval forecast

problem can be expressed as follows:

$$\begin{aligned}
 &\text{Objectives: Maximize: PICP}(w_{ij}) \\
 &\quad \text{Minimize: PINAW}(w_{ij}) \\
 &\text{Constraints: } 0 \leq \text{PICP}(w_{ij}) \leq 100\% \\
 &\quad \text{PINAW}(w_{ij}) > 0.
 \end{aligned} \tag{2.31}$$

Note that during the training process, the maximization objective can be easily transformed to the minimization of $1-\text{PICP}(w_{ij})$ according to the adopted training algorithm.

As a multiobjective optimization problem can be converted into a single-objective problem with some techniques such as weighted average method, the single-objective problem formulation mentioned above can be considered as such technique. However, the difference is obvious between single-objective problem and multiobjective problem. The former only optimizes one single objective and gets one optimal solution, while the latter optimizes several objectives simultaneously and obtains a set of trade-off solutions which are called Pareto optimal solutions. These solutions form the Pareto front from which the decision maker can select a most satisfactory one. Moreover, with the development of multiobjective evolutionary algorithm, multiobjective optimization problem can be solved efficiently and effectively without being transformed into a single-objective problem.

2.3.2 Solution Methodology

For multiobjective interval prediction problem, a good prediction model and optimization algorithm should be designed to achieve high quality PIs. In this subsection, the prediction model based on WNN is first proposed, followed by the introduction of preference-inspired coevolutionary algorithm using goal vectors (PICEA-g) optimization algorithm. Then the implementation strategy of interval prediction is presented.

2.3.2.1 Wavelet Neural Network Based Prediction Model

In direct interval prediction methods based on NN models, the multilayer perceptron (MLP) model has been widely studied. In addition, RBF NN model is also reported for multiobjective interval prediction [56]. However, another feedforward NN model, WNN, has not been studied for interval prediction problem. The first WNN model was proposed to approximate arbitrary nonlinear function as an alternative of classic feedforward NN [141]. In point forecast, it was demonstrated that the WNN model outperforms the other feedforward NN models such as MLP and RBF NN models [142], [143]. Inspired by the good performance of WNN model in point forecast, it is reasonable to explore its performance in interval prediction for wind speed. In addition, considering the essence of interval prediction problem, a multiobjective optimization framework is better suited.

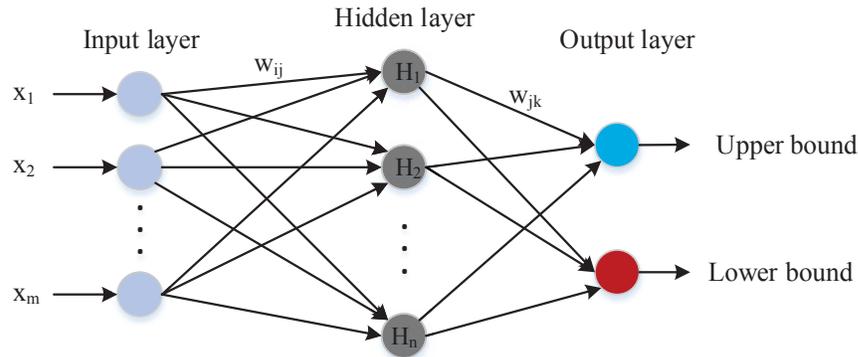


Figure 2.7: Architecture of the proposed WNN model.

Therefore, it is worthy to design a multiobjective prediction model based on the WNN model which is expected to have good prediction performance.

WNNs are developed by combining the wavelet theory and NN models. They belong to feedforward NNs and have been successfully used in some classification and time-series forecast problems [144]. Generally, there are two ways to combine the wavelet theory and NN models in forecasting tasks. One is using wavelet transformation to decompose the time-series data into some sub-series which are then combined with NN models to forecast future values [145]. Another method is to employ the wavelet basis function as the activation function of the hidden neurons to construct WNN model which is also studied in this work. WNNs can be classified into adaptive models where wavelet coefficients are variable and fixed grid WNNs where wavelet coefficients are fixed [146]. Adaptive WNNs have better generalization capability because of the wavelets' local properties and the adaption of wavelet shape corresponding to the training data. Consequently, an adaptive WNN prediction model is proposed in this study which is shown in Fig. 2.7. The proposed adaptive WNN model is a more efficient structure for forecast tasks. Note that the wavelet transformation technique may also be investigated in future research.

As can be seen in Fig. 2.7, a three layer WNN model is designed where the wavelet transformation is embedded in the hidden neurons of the WNN model [147]. The output layer has two nodes which represent the upper and lower bound of a PI, respectively. According to the universal approximation theorem that a single hidden layer feedforward NN with sigmoid activation function is able to approximate any function, we can get the superposition of sigmoid wavelet [147]. Then the key problem in designing a good WNN model is to find the optimal number of hidden nodes. Among different wavelets, the Mexican hat wavelets are symmetrical and have explicit expression which can provide exact time frequency analysis. In addition, they are based on continuous wavelet transform

and can be shifted and scaled smoothly over the entire domain [142]. Hence, in this work, the wavelet activation function used in the hidden units is the Mexican hat wavelet as follows:

$$\psi(x) = (1 - x^2)e^{-0.5x^2}. \quad (2.32)$$

Thus, the WNN model shown in Fig. 2.7 can be expressed as below:

$$H_j = \psi_{a_j, b_j} \left(\sum_{i=1}^{m_1} w_{ij} x_i \right), \quad j = 1 \cdots n_1 \quad (2.33)$$

$$\psi_{a_j, b_j}(z_0) = \psi \left(\frac{z_0 - b_j}{a_j} \right) \quad (2.34)$$

$$y_k = \sum_{j=1}^n w_{jk} H_j + g_k, \quad k = 1, 2 \quad (2.35)$$

where m_1 and n_1 are the number of input nodes and hidden nodes, respectively, w_{ij} and w_{jk} denote connection weights, a_j and b_j are scale (dilation) and shift (translation) parameters of wavelets, respectively, k is the number of output nodes and g represents the bias. Note that some other wavelet functions may also be used as the activation functions. But their performance needs to be further investigated in future. In this adaptive WNN model, the connection weights and wavelets parameters are all variable that need to be tuned to attain the best forecast performance.

The proposed interval prediction model is derived from the lower and upper bound estimation (LUBE) [9] method which is a direct unsupervised learning process to generate PIs. It can construct PIs simply and fast without making data distribution assumption. Compared with the supervised learning process, the proposed method only use the original data, and the lower and upper bounds are not required in the training process. Particularly, the proposed model directly generates unknown PIs which are gradually improved based on the evaluation indices. For the training set including the input and targets, the input is determined by correlation analysis method which is introduced in detail in Section 2.3.3.2. The real data points are used as the targets and the real lower upper bounds are unknown. During the training process, a set of preliminary lower and upper bounds is generated with the NN model as shown in (2.35) and they are compared with the real targets to calculate the evaluation indices, i.e., the optimization objectives. The PIs are adjusted iteratively based on the quality of objectives. In addition, since wavelets have shown excellent performance in nonlinear function modeling, it is expected that the proposed adaptive WNN model performs well in forecast tasks.

2.3.2.2 PICEA-g Algorithm

Various multiobjective evolutionary algorithms have been proposed such as the Pareto dominance based NSGA-II which is one of the most efficient methods by employing an

elitist and diversity preservation mechanism. Recently, a new multiobjective evolutionary algorithm, PICEA-g, has been proposed and shown to perform better than other advanced methods including NSGA-II [148]. Therefore, the PICEA-g algorithm is investigated to train the the proposed WNN prediction model in this study.

It is known that preference-based methods are helpful to generate tradeoff surfaces of interest to the decision maker in objective subspaces. With the decision maker's preferences, the incomparable solutions may become comparable. As a result, the concept of co-evolving candidate populations and a set of preferences have been proposed [149] and PICEA-g algorithm is a realization of this approach. In this approach, various preference sets help generate various regions of Pareto front. It is expected to get a good representative of the whole front with many sets of preferences as the co-evolution proceeds.

The general idea of PICEA-g is summarized as follows [148]. In PICEA-g, a set of preferences, also called goal vectors, is co-evolved with the common population of potential solutions during the search process. As for fitness assignment, the potential solutions obtain fitness by satisfying some certain goal vectors in objective space, but the fitness contribution should be shared between all the solutions that meet the goals. The goal vectors' fitness is generated by satisfactory candidate solutions and higher satisfaction implies lower fitness [150]. The aim of goal vectors is to adaptively lead the potential solutions toward the Pareto front, i.e., they co-evolve with the solution population in the process.

The implementation of PICEA-g can be illustrated in an elitist framework as shown in Fig. 2.8. A population of potential solutions, P , and a set of goal vectors, G , are co-evolved for some certain generations. For every iteration, the genetic variation operation is conducted with the parent solution population P to produce the offspring P_v . While the new goal vectors G_v are randomly regenerated according to the predefined bounds. Then the solution population and the goal vectors are pooled respectively and sorted in terms of the fitness. Lastly, truncation selection is implemented on the sorted population to produce a fixed number of potential solutions and goal vectors as the offspring population. Note that the minimization of 1-PICP and PINAW are considered as two objectives in this work which are used to calculate the fitness during the optimization process. More details about PICEA-g algorithm including the detailed fitness function can be found in the references [148] and [151].

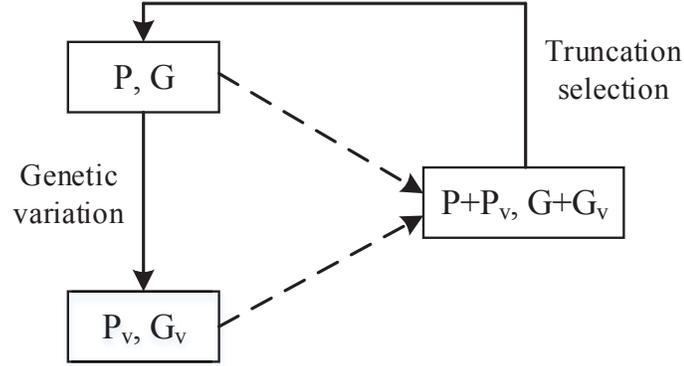


Figure 2.8: PICEA-g implementation framework.

2.3.2.3 Implementation Strategy

Based on the proposed WNN prediction model and PICEA-g training algorithm, multiobjective interval prediction for wind speed can be implemented with real datasets. The main steps of the model implementation are summarized as follows.

Step 1: Data preprocess. Although the wind speed forecast may be influenced by many factors like the weather condition and temperature, the historical wind speed is the most relevant factor which is considered as the input in this study. The original wind speed dataset should be partitioned into training set and test set. In addition, the original data are usually normalized to speed up the model training.

Step 2: Initialize the parameters of the training algorithm. For PICEA-g algorithm, the population size, maximum number of generations and parameters of genetic operators should be specified. The population are coded with real values, i.e., real-coded chromosomes are adopted. Each individual represents one WNN model and consists of all the free parameters as follows:

$$p = [w_{ij}, w_{jk}, a_j, b_j, g_k], \quad i = 1 \cdots m_1, j = 1 \cdots n_1, k = 1, 2. \quad (2.36)$$

The dilation parameter a_j and translation parameter b_j of wavelet functions are randomly initialized with uniform distribution in the intervals $[0.5, 2]$ and $[-3, 3]$, respectively [143]. The weights and bias of NN model are initialized randomly in $[-1, 1]$ with uniform distribution. The real-coded genetic operators used in this study are simulated binary crossover (SBX) and polynomial mutation (PM) [55]. For SBX operator, it can be defined with the following formulas:

$$p_{idx}^{1,t+1} = 0.5 * [(1 + \beta_{idx})p_{idx}^{1,t} + (1 - \beta_{idx})p_{idx}^{2,t}] \quad (2.37)$$

$$p_{idx}^{2,t+1} = 0.5 * [(1 - \beta_{idx})p_{idx}^{1,t} + (1 + \beta_{idx})p_{idx}^{2,t}] \quad (2.38)$$

where $p_{idx}^{1,t}$ and $p_{idx}^{2,t}$ are two parent variables in generation t , $p_{idx}^{1,t+1}$ and $p_{idx}^{2,t+1}$ are two offspring variables in generation $t+1$, and the parameter β_{idx} is calculated as follows:

$$\beta_{idx} = \begin{cases} (2r)^{\frac{1}{\eta_c+1}}, & r \leq 0.5 \\ (\frac{1}{2(1-r)})^{\frac{1}{\eta_c+1}}, & r > 0.5 \end{cases} \quad (2.39)$$

where r is a random number in the interval $[0,1]$ and η_c is the distribution index defined by the decision maker. The SBX operator intends to generate offspring near the parents which is helpful to inherit the valuable information. For PM operator, it can be expressed as follows:

$$p'_{idx} = \begin{cases} p_{idx} + \delta_{idx}(p_{idx} - p_{idx}^{low}), & r \leq 0.5 \\ p_{idx} + \delta_{idx}(p_{idx}^{up} - p_{idx}), & r > 0.5 \end{cases} \quad (2.40)$$

where p_{idx}^{low} and p_{idx}^{up} are the lower bound and upper bound of the decision variable, respectively, r is still the random number and δ_{idx} is a parameter as follows:

$$\delta_{idx} = \begin{cases} (2r)^{\frac{1}{\eta_m+1}} - 1, & r \leq 0.5 \\ 1 - (2(1-r))^{\frac{1}{\eta_m+1}}, & r > 0.5 \end{cases} \quad (2.41)$$

where η_m is the user-defined index parameter.

Step 3: Determine the optimal WNN structure. The parameters of WNN prediction model mainly includes the number of input nodes and hidden nodes. The input features can be determined by correlation analysis method. Specifically, the correlation analysis is implemented with the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) in this work, which will be introduced explicitly in next subsection. Considering the sequence of time series data, the number of hidden nodes is determined by trial and error method which has a similar idea with cross validation method [40]. The hypervolume indicator [152] is employed to assess the model performance and determine the optimal model structure. A higher hypervolume value stands for a better model.

Step 4: Model training and evaluation. After determining the parameters and optimal structure of the prediction model, the model was retrained with the training data. The termination condition is to reach the predefined maximum iteration in this study. When the training terminates, the Pareto front is attained for the test dataset which consists of the PICP and PINAW values of each individual. The hypervolume can also be calculated to evaluate the model.

Step 5: PI construction. From the multiobjective optimization method, a set of Pareto optimal solutions can be obtained. To construct high quality and satisfactory

PIs, the decision makers may select the best solution according to their preferences such as the reliability requirement. This flexible selection is obviously an advantage of multiobjective interval prediction over the single-objective interval prediction method.

In wind speed forecast, the associated uncertainty can be represented by different probabilistic approaches including probability density function, moments of distribution, quantiles and intervals [121]. The most commonly used probabilistic forecast method is based on quantiles. However, we can only get one quantile in one simulation with such method and construct the PIs indirectly with a pair of these quantiles. In contrast, the proposed interval forecast method can produce a set of optimal solutions simultaneously, and the PIs are constructed directly without the estimation of quantiles. Therefore, from the decision maker's viewpoint, the proposed multiobjective interval prediction method is more efficient and concise.

2.3.3 Numerical Results

To verify the effectiveness of the proposed multiobjective interval prediction model, case studies with real-world wind speed data are executed in this subsection. First, the datasets used as well as the parameter settings of the prediction model are depicted. Then the prediction results and comparison with other models are demonstrated.

2.3.3.1 Datasets

The wind speed data used in this study are hourly mean wind speed taken from two locations: Victoria and Edmonton in Canada [153]. The time periods of two datasets are both from 1 August 2016 to 31 July 2017 with the hour unit. However, in this time period, the Victoria dataset has 5 missing values. The missing values cannot be deleted directly to keep the wind speed distribution. As the overall data trend will not dramatically change in a very short time, the mean value of the data before and after the missing data point is used to replace the missing one in this study.

Victoria is located on Vancouver Island while Edmonton is an inland city, thus the wind speed data from these two locations are expected to have different characteristics. The descriptive statistics of the two chosen datasets are summarized in Table 2.9. In this study, 80% of the one year data (from August 2016 to May 2017) are used to train the prediction model, the remaining are utilized to test the model. In addition, the training set and the testing set are normalized to $[-1,1]$, respectively. As the forecast accuracy decreases with the increase of forecast time scale, one step ahead interval forecast is conducted in this work.

Table 2.9: Descriptive statistics of the two datasets

Location	Mean	Std.	Min.	Max.
Victoria	11.03	7.08	0	55
Edmonton	10.92	6.24	0	48

Table 2.10: Parameters of PICEA-g algorithm

Parameter	Value
Npop	40
MaxGen	1000
Crossover operator	SBX ($p_c=1, \eta_c=15$)
Mutation operator	PM ($p_m = 1/nvar, \eta_m=20$)

2.3.3.2 Parameter Settings

Two sets of parameters need to be determined in the proposed prediction model. One is about the PICEA-g algorithm, the other one is about the WNN model. The parameters of PICEA-g algorithm used in this study are collected from the reference [148] as shown in Table 2.10. N_{pop} is the population size of candidate solutions. The number of goal vectors is equal to the population size. $MaxGen$ represents the maximum number of iterations which controls the termination of the model training. For SBX crossover operator, the recombination probability p_c is 1 and the distribution index η_c is set to 15. For PM mutation, the mutation probability p_m is related to the number of decision variables $nvar$ and the distribution index η_m is equal to 20 in this study.

In time series forecast, correlation analysis is usually employed to identify the order of the model. In particular, the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are often utilized to conduct correlation analysis between the forecast value and past historical data [40]. Therefore, the ACF and PACF analysis is adopted to determine the input of the WNN model, i.e., determine the input values that have maximum correlation to the forecast values. The ACF and PACF analysis method is widely used in forecast tasks such as [40], [56]. Since the intermittent and volatile wind speed fluctuates every now and then, it shows no apparent daily and weekly trend and we can assume that it is stationary. Then the ACF and PACF analysis can be used directly without difference operation. For the Victoria dataset, the ACF and PACF are shown in Fig. 2.9. As can be seen from this figure, the ACF has an exponential decaying trend and the PACF is cut off at lag 3. Thus, the proper order of this time series should be 3. Considering x_t as the time series variable, the vector (x_{t-2}, x_{t-1}, x_t) is then used as the input to forecast the value x_{t+1} at next step. Likewise, the correlation analysis with ACF and PACF can also be implemented for Edmonton wind speed data. The proper time series order is also 3 and the similar analysis graph is omitted here.

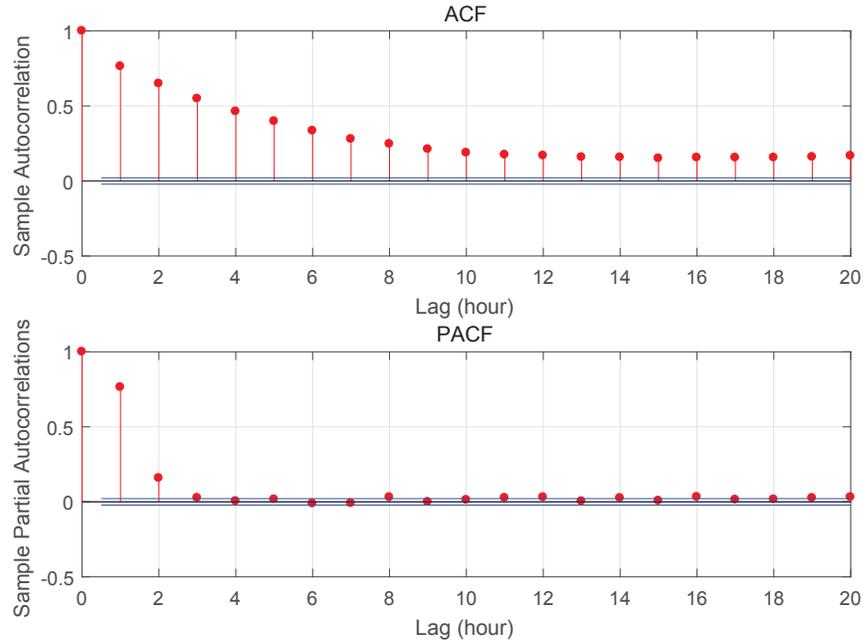


Figure 2.9: ACF and PACF analysis of Victoria wind speed.

The number of hidden nodes usually has an important effect on the NN model performance. To determine the optimal number of hidden nodes, two methods including cross-validation and trial and error method are often studied in the previous literature. In this study, the method proposed in [40] is adopted and combined with the hypervolume indicator to investigate the optimal number of hidden neurons which has a similar idea with cross-validation. For each selection of NN models, it is trained and validated for five times with the training and testing datasets. The hypervolume indicator is calculated for every simulation run and the average hypervolume value for each model was obtained. The hypervolume represents the proportion of the objective space calculated based on the obtained approximating Pareto front and a certain reference point [150]. The reference point is set to (1.2, 1.2) for the minimization problem with two objectives (1-PICP, PINAW) in this study. The hypervolume is calculated by the method developed in [154]. For the minimization problem, the model with maximum average hypervolume value is chosen to be the best model. Considering the balance of computation complexity and generalization capability, the number of hidden nodes is limited to change from 3 to 10 in this work. The average hypervolume results for Victoria dataset are given in Fig. 2.10. As can be seen from this figure, the model with 8 hidden neurons has the best performance. Therefore, the optimal structure of WNN prediction model for Victoria data is 3-8-2. Similarly, the optimal number of hidden nodes of the prediction model for Edmonton dataset was determined to be 7.

In single-objective interval prediction, CWC is used as the comprehensive index to

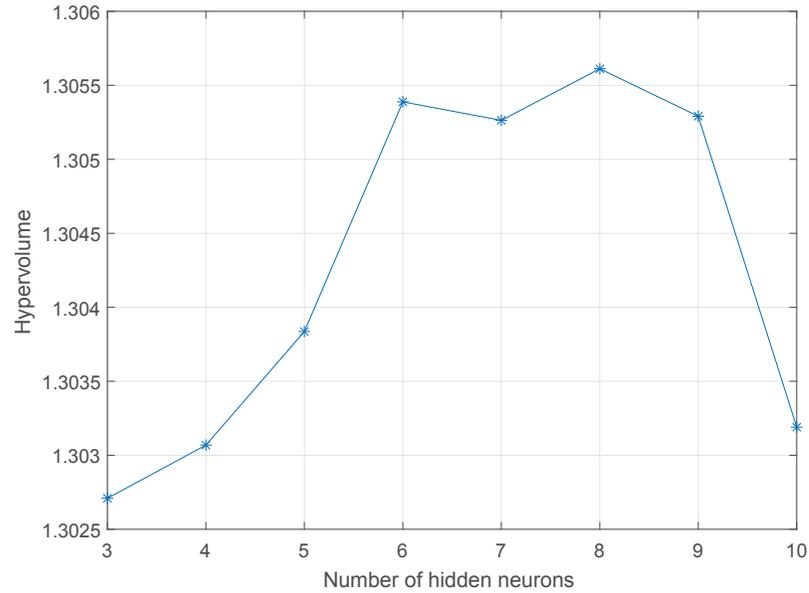


Figure 2.10: Average hypervolume results.

evaluate the PIs' quality. In order to compare multiobjective interval prediction with single-objective interval prediction methods conveniently, CWC index is also investigated in this work. For CWC parameters, the parameter μ is specified as the nominal confidence level $1 - \alpha = 0.9$, the large constant η is set to 50 [9]. These parameters may also be variable according to the decision maker.

2.3.3.3 Prediction Results

After specifying the parameters and determining the optimal model structure, multiobjective interval prediction for wind speed can be implemented. The model is first trained with the training data. After the training termination is reached, a set of Pareto optimal solutions, i.e., a set of non-dominated optimal prediction models can be obtained. Applying these models to do interval prediction with test data leads to the required Pareto front of test set.

The test results or the Pareto front with the proposed multiobjective interval prediction model for Victoria and Edmonton data are presented in Fig. 2.11. From this figure, we can see that good prediction results can be obtained for both datasets. Both the Pareto fronts show good convergence and diversity and have reasonable and valid objective values. Each point in the Pareto front indicates the result of a prediction model. Actually, the decision maker can choose a satisfactory prediction model among these Pareto solutions of training sets to construct PIs according to his posterior preference such as the interval prediction reliability requirement.

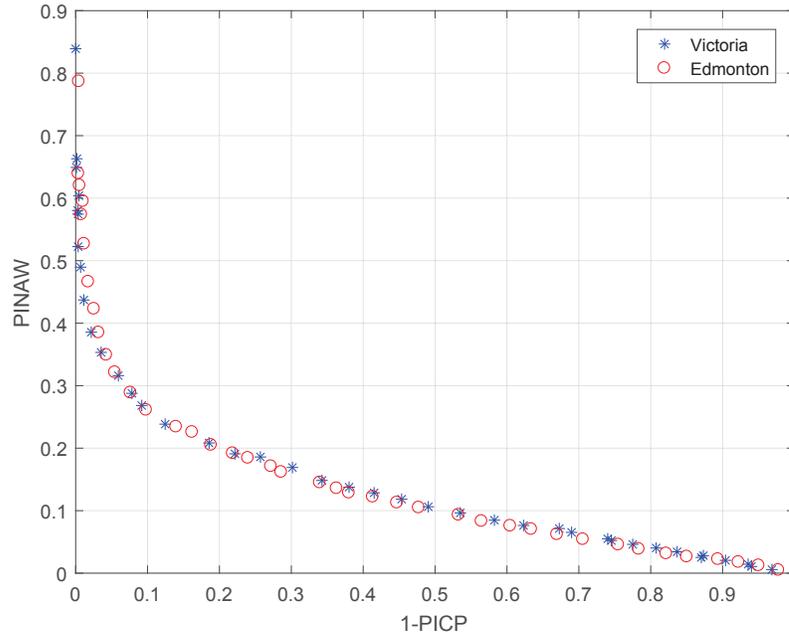


Figure 2.11: Pareto front of prediction results.

To generate high quality PIs, the comprehensive index CWC is considered and the model with the smallest CWC value is taken as the most appropriate model. Since the nominal confidence level is set to 0.9 in this work, choosing the model with smallest CWC value is equivalent to choose the one with smallest PINAW value among those with PICP not less than 0.9. The interval prediction results from the model with the smallest CWC value for Victoria dataset are shown in Fig. 2.12. Note that only the prediction results of the last week of the test dataset are shown in this figure for better visualization. In addition, as the real wind speed is impossible to be below zero, the lower bound limitation is set to zero in this study [40]. As can be seen from this figure, the PIs generated from the model are valid and narrow with PICP=90.81% and PINAW=26.83%. Both the upper bound and the lower bound vary similarly with the actual data.

Similarly, the PI construction results for Edmonton dataset can be attained as shown in Fig. 2.13. From this figure, it can be seen that the wind speed of Edmonton has a different fluctuation trend, but high quality PIs can still be generated by the proposed WNN prediction model. The constructed PIs are able to enclose the real targets well. For this case, PICP is 90.18% and PINAW equals 26.11%, which indicate narrow PIs on the condition that the reliability is guaranteed.

2.3.3.4 Comparison with Other Models

In order to substantiate the effectiveness of the proposed interval forecast approach, several benchmark models are employed to conduct interval prediction with the same

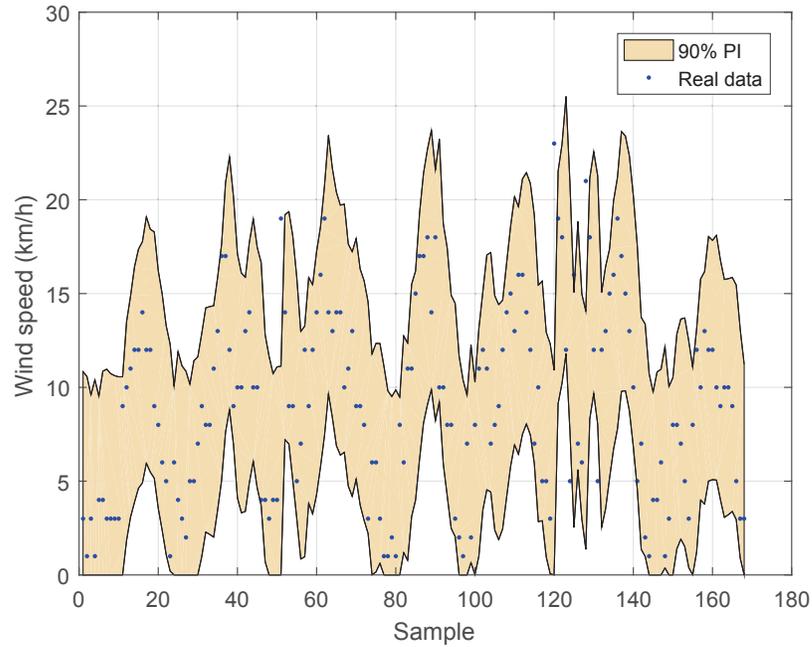


Figure 2.12: Interval prediction results for Victoria data.

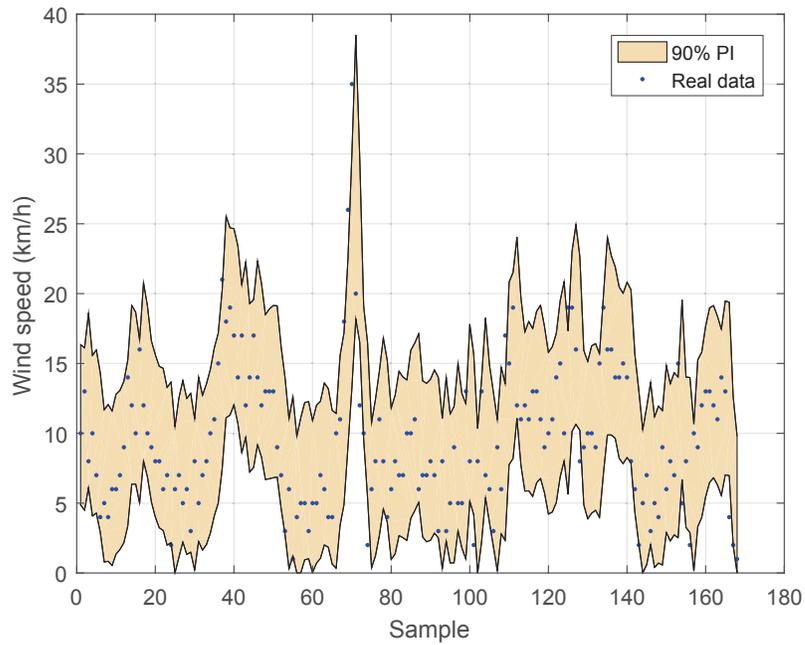


Figure 2.13: Interval prediction results for Edmonton data.

datasets for comparison purpose. The proposed multiobjective interval prediction method is first compared with other common single-objective interval forecast methods, then different NN models in a multiobjective framework are also compared.

The single-objective interval prediction methods considered in this study include NN based LUBE method [9], quantile regression (QR) method and Naive method. NN based LUBE method is widely studied for wind power interval prediction and the NN model is usually the MLP model. The NN based LUBE method proposed in [40] is adopted to conduct single-objective interval prediction for wind speed and the NN connection weights are tuned by particle swarm optimization (PSO) algorithm. The corresponding parameters can also be found in the reference. Note that some other feedforward NN models and optimization algorithms (e.g., genetic algorithm) can also be studied for single-objective interval forecast model which share the same principle. QR is another typical probabilistic forecast approach [136] which can be applied to interval prediction. Naive method is also a general benchmark forecast model and it works similarly with the persistence model in point forecast. Naive method forecasts future intervals based on the past historical data and it performs well for short-term forecast task. In this study, the forecast error is assumed to follow normal distribution, the last wind power value is used as the mean, and the variance is calculated based on the latest observations [35]. With the mean and variance, PIs can be constructed for the forecast horizon.

The interval prediction results from the proposed approach and the benchmark techniques for Victoria and Edmonton data are given in Table 2.11 and Table 2.12, respectively. In addition to PICP and PINAW, ACE, AWD and Score indices are also presented. The CWC value which is a comprehensive index is also listed in the results. Note that the results of the proposed WNN-PICEA-g method are generated from the most satisfactory solution of the Pareto solution set as mentioned above. For better persuasiveness, the experiments with different PINC values are conducted, i.e., PINC=85%, 80%, 75% and 70% are also studied as shown in the tables.

From Table 2.11, we can see that the proposed method and NN-LUBE-PSO method can generate valid PIs ($\text{PICP} \geq \text{PINC}$) for Victoria data for all experiments. However, QR and Naive methods are not so good. Obviously, the result of the proposed method has the minimum interval width and ACE value. Although the NN-LUBE-PSO method construct PIs with a high reliability, the PIs are less informative as they are too wide, and the very high probability also leads to a slightly lower AWD value. Since CWC and Score index can measure both the coverage probability and interval width of PIs, they can be used to compare the overall performance of various forecast approaches. Therefore, the proposed multiobjective interval prediction method has the best performance in Table 2.11. In addition, the multiobjective interval prediction method produces a set of Pareto solutions with a simulation run which can offer more choices to the decision maker than the single-objective prediction methods. More specifically, we can select the proper solutions from the Pareto solutions according to different PINC requirements, while

Table 2.11: Comparison results for Victoria data

PINC	Method	PICP (%)	PINAW (%)	ACE (%)	AWD	Score	CWC
90%	WNN-PICEA-g	90.81	26.83	0.81	0.0196	-3.3890	0.2683
	NN-LUBE-PSO	96.17	36.12	6.17	0.0058	-3.6507	0.3612
	QR	85.72	29.98	-4.28	0.0535	-3.9617	8.7889
	Naive	85.80	26.39	-4.20	0.0376	-3.6267	8.4143
85%	WNN-PICEA-g	85.95	23.31	0.95	0.0310	-4.4562	0.2331
	NN-LUBE-PSO	91.26	30.54	6.26	0.0144	-5.0725	0.3054
	QR	80.24	24.86	-4.76	0.0834	-5.1405	11.0542
	Naive	81.81	23.33	-3.19	0.0548	-4.8260	5.1543
80%	WNN-PICEA-g	81.27	20.77	1.27	0.0450	-5.5404	0.2077
	NN-LUBE-PSO	87.09	26.96	7.09	0.0314	-6.3132	0.2696
	QR	74.01	21.05	-5.99	0.1202	-6.1379	20.1475
	Naive	77.71	20.94	-2.29	0.0744	-5.8641	3.3548
75%	WNN-PICEA-g	75.79	18.11	0.79	0.0679	-6.4028	0.1811
	NN-LUBE-PSO	82.92	21.65	7.92	0.0475	-6.6901	0.2165
	QR	69.50	18.46	-5.50	0.1575	-7.0067	15.8027
	Naive	73.32	18.92	-1.68	0.0970	-6.7774	2.5077
70%	WNN-PICEA-g	71.56	16.39	1.56	0.1020	-7.3752	0.1639
	NN-LUBE-PSO	77.84	20.10	7.84	0.0930	-8.0008	0.2010
	QR	64.31	16.12	-5.69	0.2053	-7.7543	17.3963
	Naive	69.21	17.14	-0.79	0.1235	-7.5989	1.6534

several simulation experiments need to be conducted with a single-objective forecast model.

Similar forecast results are also obtained for Edmonton data as shown in Table 2.12. The proposed WNN-PICEA-g method can still construct valid PIs with narrow width which demonstrates the stability and consistency of the method. It is still the best forecast method according to the CWC value followed by the NN-LUBE-PSO method. For QR and Naive method, the PICP value cannot reach the nominal confidence level most of the time resulting a high CWC value. Furthermore, we can find that the ACE value of the proposed method is much closer to 0. In summary, the proposed multiobjective interval prediction method can construct PIs effectively and performs better than the benchmark approaches.

Since several other multiobjective interval prediction methods based on NN models have been reported in the previous literature [39], [56], multiobjective comparison is also implemented between the WNN model and other NN models including MLP NN and RBF NN. The implementation strategy for MLP NN and RBF NN model is the same with that of the proposed model. The PICEA-g algorithm with the same parameters is still used as the training algorithm. To compare the performance of different models quantitatively, the hypervolume indicator is adopted to measure the obtained Pareto front.

Table 2.12: Comparison results for Edmonton data

PINC	Method	PICP (%)	PINAW (%)	ACE (%)	AWD	Score	CWC
90%	WNN-PICEA-g	90.18	26.11	0.18	0.0233	-3.5946	0.2611
	NN-LUBE-PSO	94.57	33.12	4.57	0.0133	-3.9449	0.3312
	QR	81.78	27.54	-8.22	0.0664	-4.2310	61.1661
	Naive	86.43	27.85	-3.57	0.0389	-3.9903	6.2351
85%	WNN-PICEA-g	87.49	23.60	2.49	0.0389	-4.9093	0.2360
	NN-LUBE-PSO	90.92	27.73	5.92	0.0237	-5.1288	0.2773
	QR	75.44	22.61	-9.56	0.1039	-5.4662	119.1755
	Naive	81.70	24.64	-3.30	0.0557	-5.2556	5.4560
80%	WNN-PICEA-g	80.18	20.19	0.18	0.0728	-5.9536	0.2019
	NN-LUBE-PSO	83.27	23.04	3.27	0.0410	-6.0864	0.2304
	QR	70.19	19.20	-9.81	0.1466	-6.4896	135.2586
	Naive	78.34	22.12	-1.66	0.0749	-6.3506	2.5200
75%	WNN-PICEA-g	76.98	19.11	1.98	0.0755	-6.9284	0.1911
	NN-LUBE-PSO	77.27	21.75	2.27	0.0715	-7.5641	0.2175
	QR	64.76	16.61	-10.24	0.1947	-7.3444	167.2513
	Naive	74.80	19.98	-0.20	0.0969	-7.2948	1.3047
70%	WNN-PICEA-g	70.42	16.14	0.42	0.1033	-7.4815	0.1614
	NN-LUBE-PSO	74.87	17.89	4.87	0.0861	-7.6088	0.1789
	QR	59.91	14.52	-10.09	0.2495	-8.0989	155.4964
	Naive	70.58	18.10	0.58	0.1226	-8.1209	0.1810

The average hypervolume results of different NN forecast models for Victoria and Edmonton data are given in Table 2.13. The hypervolume values in this table are average results for the test dataset from five independent simulation runs and a larger hypervolume value means a better result for multiobjective minimization problem. As can be seen from this table, the WNN model is slightly better than the MLP NN model, but it performs much better than the RBF NN model, especially for the Edmonton dataset. Therefore, the proposed WNN forecast model performs best which has the maximum hypervolume.

In addition, NSGA-II algorithm is one of the most efficient multiobjective optimization algorithms and has been widely studied to deal with different multiobjective problems. To verify the performance of PICEA-g algorithm employed in this study, NSGA-II algorithm is used to train the WNN model for comparison. The hypervolume results obtained from the WNN model with NSGA-II algorithm are listed in the last column of Table 2.13, which are also average results of five individual runs. It is obvious that the hypervolume results from WNN trained by PICEA-g algorithm are better than those from WNN with NSGA-II algorithm. Thus, we can conclude that PICEA-g algorithm has good performance in the proposed multiobjective interval prediction method.

Table 2.13: Multiobjective comparison results for different models

	WNN	MLP	RBF	WNN-NSGA-II
<i>Victoria</i>				
hypervolume	1.3056	1.3053	1.2993	1.2989
<i>Edmonton</i>				
hypervolume	1.3047	1.3043	1.2899	1.2966

Table 2.14: Training results for different population size

N_{pop}	<i>Victoria</i>		<i>Edmonton</i>	
	time (s)	hypervolume	time (s)	hypervolume
30	36.90	1.3050	37.43	1.3122
40	44.73	1.3086	44.14	1.3169
50	52.36	1.3125	52.17	1.3183

2.3.3.5 Discussion

The proposed multiobjective interval prediction model mainly consists of WNN model and PICEA-g optimization algorithm. For PICEA-g algorithm, the widely used crossover and mutation operators are maintained, and the corresponding parameter values are collected from the reference which can be considered as optimal in this work. It is possible to study different crossover and mutation operators and corresponding parameters to further improve the performance of the forecast method in future. In addition, the influence of the population size N_{pop} is studied. The average training time and average training hypervolume results with different N_{pop} are summarized in Table 2.14, which shows that the training time has a positive correlation with the N_{pop} value. We can also find that all the training time for different population size are less than one minute which shows the computational efficiency of the optimization algorithm. When the training process is finished, the testing time is less than one second. Particularly, the training time is about 45s when N_{pop} is 40 which is much less than the time scale (1 hour) of the dataset. Hence, the proposed forecast model can be used to real-time wind speed forecast. The increasing hypervolume values result from the increasing evenly distributed points in the Pareto front. Therefore, the decision maker needs to select a proper population size to balance the training time and the number of Pareto solutions in practice.

In addition, various feature selection methods, such as mutual information method, recursive feature elimination, and chaotic feature selection based on phase space reconstruction, can be investigated to preprocess the input data which may potentially improve the forecast performance. The correlation analysis method is used in this work for its efficiency and simplicity. To evaluate its effectiveness, another feature selection method, phase space reconstruction, is studied for comparison purpose. The phase space reconstruction technique aims to determine the delay vectors as the input. By

Table 2.15: Hypervolume results for different feature selection methods

N_{pop}	<i>Victoria</i>		<i>Edmonton</i>	
	PSR method	CA method	PSR method	CA method
30	1.2874	1.3050	1.2977	1.3122
40	1.2951	1.3086	1.3048	1.3169
50	1.3017	1.3125	1.3094	1.3183

delay embedding theorem, we need to find two parameters in terms of the embedding dimension and the time delay, which can be obtained by the mutual information method and false nearest method, respectively [155]. In this work, the embedding dimension is 8 and the time delay is 13 for Victoria dataset, and they are 8 and 15 for Edmonton dataset, respectively. Then the delay vectors or the input can be constructed. The number of the hidden nodes in the WNN model is determined with the same method as introduced before. Similarly, the average training hypervolume results based on phase space reconstruction (denoted as PSR) with various N_{pop} are given in Table 2.15, and hypervolume results with correlation analysis (denoted as CA) method are also listed for better comparison. From this table, we can find that the correlation analysis method is effective and sufficient to determine the input for short-term wind speed forecast. For more complex forecast tasks, it is worth studying other advanced feature selection methods, which is left for future work.

Compared with the single-objective interval prediction model, we can get a Pareto front (a set of optimal solutions) from the proposed multiobjective interval prediction model. Among the nondominated optimal solutions, the decision maker can flexibly choose a proper solution according to the demand of reliability and interval width. Each solution corresponds to an interval forecast model. With the choice of a proper model, interval prediction can be implemented with new dataset. In addition, there are different ways to use the prediction intervals in reality. For instance, they can be applied to robust optimization and control problems for power systems integrated with renewable generation [155]. More specifically, in robust optimization problems with box-type uncertainty set, the prediction intervals can be directly used to describe the uncertainty without the assumption of probability distribution, i.e., only the lower and upper bounds are required in robust optimization problems. They can also be processed to get the point forecast values by some convex combination methods, such as the weighted summation method with the obtained lower and upper bounds.

2.4 Summary

Two interval forecast problems for wind energy are studied in this chapter, i.e., direct interval prediction for wind power based on RNN model and multiobjective interval

prediction with WNN model for wind speed.

First, the RNN-based LUBE method is proposed to directly construct optimal PIs for wind power forecast. The RNN model has dynamic features and is suitable for time series forecast. Based on the single-objective problem formulation, a novel comprehensive cost function with a new PI evaluation index is designed to enhance the model training. To optimize the RNN prediction model, the DA algorithm with a new weight update method is introduced to solve the problem. In addition, the delay embedding theorem is applied to reconstruct the chaotic wind power data for better prediction. The numerical results with a real world wind power dataset show that the proposed RNN prediction model can construct better PIs compared with the benchmark models.

Second, multiobjective interval prediction based on WNN model is proposed for short-term wind speed forecast. The novel multiobjective evolutionary algorithm, PICEA-g, is employed to train the WNN prediction model. Two case studies are implemented to testify the performance of the proposed model with real-world hourly wind speed data from Canada, and valid and narrow PIs are obtained. In addition, experimental results show the superiority of the proposed approach compared with other benchmark methods, the performance of PICEA-g algorithm is also verified by a comparison with the popular NSGA-II algorithm. In particular, the quality of PIs from the proposed multiobjective model is better than those from other single-objective forecast models with respect to different PINC settings including 90%, 85%, 80%, 75% and 70%. For multiobjective model comparison, the proposed model also achieves higher average hypervolume results (1.3056 for Victoria dataset and 1.3047 for Edmonton dataset) than other models. Moreover, the training time with PICEA-g algorithm for two case studies are both less than one minute which shows the feasibility of the proposed model.

3

Chance-constrained Energy Management for Islanded Microgrids

Acronyms

CC	Chance-constrained
CCP	Chance Constrained Programming
DG	Distributed Generation
DRO	Distributionally Robust Optimization
ESS	Energy Storage System
SAA	Sample Average Approximation
SDP	Semidefinite Programming
SOCP	Second-order Conic Programming
UC	Unit Commitment

3.1 Introduction

This chapter presents a chance-constrained energy management problem for islanded microgrids with the new uncertainty modeling technique - DRO method.

As an important paradigm of smart grid, the research of microgrid has been widely conducted. With the high penetration of renewable generation, microgrid energy management problem has become more challenging due to the introduction uncertainty. Although traditional SP and RO method have been applied to deal with microgrid energy management with uncertainty, both methods have some drawbacks with respect to computational burden or over-conservativeness. In addition, chance constrained programming (CCP) method is also investigated in microgrid energy management.

In the existing literature, most of the research works focus on grid-connected microgrids and only a few study isolated microgrids. In addition, the influence of renewable generation uncertainty has not been sufficiently studied since the power balance constraint is mostly used in the problem formulation. However, in a real isolated microgrid, the power balance may not always be satisfied with uncertain renewable energy.

Therefore, a chance-constrained (CC) energy management problem for isolated microgrids considering the influence of uncertain renewable energy on power balance is studied in this chapter, and DRO method is proposed to capture the uncertainty of wind power. Specifically, the problem formulation is presented in section 3.2, and section 3.3 proposes the solution method to reformulate the distributionally robust chance constraint to be second-order conic constraints so that the complete problem can be solved as an SOCP problem. Then a case study with real datasets is carried out in 3.4.

3.2 Problem Formulation

A typical isolated microgrid is composed of conventional generators, ESS, renewable generation and different kinds of load. It is usually assumed that the microgrid energy management is controlled by an EMS in a centralized mode. In this section, different components of the studied microgrid system are introduced and the corresponding energy management problem is formulated.

3.2.1 Distributed Generation

Distributed generation (DG) in a microgrid usually includes micro-turbines, diesel generators, fuel cells and renewable generation such as uncertain wind power which is introduced in the next subsection. The conventional generation units are important components of a microgrid and they are dispatchable to meet the load demand. In this study, distributed generators are the main focus which burn fossil fuel to generate electricity. The output power of generators is restricted by the maximum and minimum limits as follows:

$$P_i^{min} \leq P_{i,t} \leq P_i^{max}, \forall i, t. \quad (3.1)$$

In addition, the DG units should satisfy the ramping up/down constraints:

$$P_{i,t+1} - P_{i,t} \leq R_i^{up}, \forall i, t \quad (3.2)$$

$$P_{i,t} - P_{i,t+1} \leq R_i^{dn}, \forall i, t. \quad (3.3)$$

Note that the generators are assumed to have on status over a finite time horizon in the energy management [59]. If the unit commitment problem is considered in an extended

model, then the start-up and shut-down constraints should also be included.

Generally, the operation cost of generators is mainly the fuel consumption cost or generation cost which can be expressed as a quadratic model [69] as follows:

$$CG_{i,t} = (a_i P_{i,t}^2 \Delta t + b_i P_{i,t} + c_i) \Delta t. \quad (3.4)$$

The linear cost model [61] is included in this quadratic model. In order to achieve environmentally friendly microgrid energy management, the emission effect should also be considered. In this study, it is assumed that only diesel generators produce emission and the emission model is also represented as a quadratic function [156], [157]. Then the emission cost model can be expressed as follows:

$$Em_{i,t} = (d_i P_{i,t}^2 \Delta t + e_i P_{i,t} + f_i) \Delta t \quad (3.5)$$

$$CE_{i,t} = c^{emis} * Em_{i,t} \quad (3.6)$$

where c^{emis} is the emission cost coefficient. Note that the diesel generators are common and necessary in isolated microgrids installed in some specific areas. Some more clean renewable sources such as fuel cell may also be included in the microgrids to reduce the emission, and the objective function will change accordingly.

3.2.2 Energy Storage System

In a microgrid integrated with renewable generation, the ESS plays a critical role in mitigating the system uncertainties and maintaining the power balance. Considering a battery ESS, we have the following dynamic model and constraints:

$$E_{j,t+1} = E_{j,t} + \eta_j^{ch} P_{j,t}^{ch} \Delta t - P_{j,t}^{dch} \Delta t / \eta_j^{dch}, \forall j, t \quad (3.7)$$

$$0 \leq P_{j,t}^{ch} \leq P_j^{ch,max}, \forall j, t \quad (3.8)$$

$$0 \leq P_{j,t}^{dch} \leq P_j^{dch,max}, \forall j, t \quad (3.9)$$

$$E_j^{min} \leq E_{j,t} \leq E_j^{max}, E_{j,T} = E_{j,0}, \forall j, t \quad (3.10)$$

where constraint (3.7) represents the dynamics of the stored energy; constraints (3.8) and (3.9) are used to limit the charging and discharging power. In constraint (3.10), the ESS capacity is restrained by a lower and upper bound to avoid overcharging and deep discharging. In addition, the final stored energy is assumed to be equal to its initial energy level. Note that the complementary constraint $P_{j,t}^{ch} P_{j,t}^{dch} = 0$ is usually used to avoid simultaneous charging and discharging which results in a mixed-integer linear programming (MILP) model for ESS in some references [63], [69]. Actually, this constraint is redundant when charging and discharging efficiency are considered as in this work and the MILP model can be exactly relaxed to a linear model to reduce the computational burden [158], [159].

The ESS will degrade with frequent charging and discharging process. Therefore, the ESS degradation cost should be considered in energy management. In this work, a linear model is adopted to calculate the ESS degradation cost [61] for simplicity as follows:

$$CS_{j,t} = c_j^{ess}(\eta_j^{ch} P_{j,t}^{ch} \Delta t + P_{j,t}^{dch} \Delta t / \eta_j^{dch}) \quad (3.11)$$

where c^{ess} is the degradation cost coefficient.

3.2.3 Load Demand

The load demand in a microgrid can be classified into two groups: critical loads and deferrable loads. Critical loads are non-dispatchable and must be satisfied in highest priority such as the hospital load demand. In this study we use L_t to represent the critical load at each time slot.

Unlike critical loads, deferrable loads are dispatchable and can be scheduled according to the real-time power supply and demand. For these kind of loads, using the electrical vehicle as an example, their load only needs to be satisfied over a specified time horizon. Hence, the deferrable load demand model can be expressed as follows [59]:

$$\sum_{t=T_k^a}^{T_k^b} P_{k,t}^{def} = P_k^{def}, \forall k, t \in [T_k^a, T_k^b] \quad (3.12)$$

$$P_k^{def,min} \leq P_{k,t}^{def} \leq P_k^{def,max}, \forall k, t \in [T_k^a, T_k^b] \quad (3.13)$$

$$P_{k,t}^{def} = 0, \forall k, t \notin [T_k^a, T_k^b]. \quad (3.14)$$

3.2.4 Chance Constraint for Power Balance

For power balance constraint, most of the existing research focuses on the strict balance of power generation and load demand. However, in an islanded microgrid, the power demand may not always be satisfied due to the uncertain renewable generation, or the strict power balance will result in high cost. To deal with this problem, CCP can be used to allow the solutions to violate the constraint with no more than a small specified probability, i.e., the constraint should be met with a certain confidence level [160]. Consequently, a chance constraint for power supply and demand is proposed in this study which can be represented as follows:

$$Pr\left\{\sum_{i=1}^{N_{dg}} P_{i,t} + \sum_{j=1}^{N_{ess}} (P_{j,t}^{dch} - P_{j,t}^{ch}) + w_t \geq L_t + \sum_{k=1}^{N_{def}} P_{k,t}^{def}\right\} \geq 1 - \epsilon \quad (3.15)$$

where w_t is the aggregated random renewable power output and the uncertainty set of its probability distribution is introduced in the next section, ϵ is a predefined small probability index.

Based on the above notations, we can get the objective function of the microgrid energy management, i.e., the total operation cost of the microgrid, which includes the generation cost and emission cost of diesel generators and the ESS degradation cost as shown below:

$$C_{tot} = \sum_{t=1}^T \left[\sum_{i=1}^{N_{dg}} (CG_{i,t} + CE_{i,t}) + \sum_j^{N_{ess}} CS_{j,t} \right]. \quad (3.16)$$

Thus, the complete chance constrained microgrid energy management problem is formulated as follows:

$$\min_{\mathbf{x}} \{(3.16) : (3.1) - (3.3), (3.7) - (3.10), (3.12) - (3.15)\}$$

where \mathbf{x} denotes the set of decision variables.

The objective of the proposed chance constraint in the system model is to maximize the system reliability, i.e., the load demand should be satisfied with a high probability. Therefore, the dumping load can be added to absorb the excess power supply in practical system operation to keep the power balance [161]. In case of insufficient energy supply, we can also conduct frequency regulation for the generators or add more spare storage systems to achieve the power balance in real-time operation. In addition, the proposed energy management model is a single-stage one as the unit commitment (UC) problem is not considered here. It is assumed that the UC decision has been done. The integration of UC problem which elicits a two-stage energy management model is left for future research. In practice, the optimal solution can provide the decision maker a preliminary and robust dispatch plan when the uncertainties are unknown and the proposed microgrid model may be applied in some remote islands where real-time dispatch is not so convenient.

3.3 Solution Methodology

The microgrid energy management problem formulated above is difficult to solve directly due to the chance constraint and the uncertain renewable generation. To handle this problem, in this section, a novel ambiguity set is first introduced to describe the uncertain probability distribution of renewable power output. Then, based on this ambiguity set, the DRO method is applied to process the chance constraint and the problem is reformulated as a tractable SOCP problem.

3.3.1 An Ambiguity Set for Wind Power Output

In the power balance chance constraint, the renewable generation w_t is a random variable. In this work, without loss of generality, wind power is considered as the renewable generation. To describe the uncertainty of wind power output, different methods have been studied in the literature, e.g., the polyhedral and ellipsoid uncertain set in

robust optimization method [59], [162], a particular probability distribution in stochastic optimization [63]. Unlike the robust optimization and stochastic optimization method, the DRO method handles the uncertain wind power with an ambiguity set. The ambiguity set is described as the family of all distributions that have the same moment information such as mean, variance and covariance and structural properties [163]. Different kinds of ambiguity sets have been researched on DRO in the literature including Markov ambiguity set [164], Chebyshev ambiguity set [165] and so on.

Among various ambiguity sets, the ambiguity sets with known moment information are widely studied on uncertainty quantification [165], [100] and they have been adopted to handle the random wind power [81]. The typical moment ambiguity set with known mean and variance can be expressed as follows:

$$\mathcal{P}_t^1 = \left\{ \mathbb{P}_t \in \mathcal{P}_t^0(W_t) \left| \begin{array}{l} \mathbb{P}\{w_t \in W_t\} = 1 \\ \mathbb{E}_{\mathbb{P}_t}\{w_t\} = \mu_t \\ \mathbb{E}_{\mathbb{P}_t}\{(w_t - \mu_t)^2\} = \sigma_t^2 \end{array} \right. \right\} \quad (3.17)$$

where μ_t and σ_t^2 represent the mean and variance which can be obtained from historical data. $\mathcal{P}_t^0(W_t)$ denotes the family of all the probability distributions on the support of W_t and \mathbb{P}_t is the probability distribution of w_t .

Although the mean and variance of wind power can be estimated from abundant historical data, their actual values are hard to know in reality and the estimation may not be accurate. In other words, it is difficult to determine the exact moment values. To tackle the uncertain moment information, the ambiguity set with bounded moment such as ellipsoid and conic bound [82] is studied in uncertainty description. Inspired by the polyhedral uncertain set in robust optimization, a box-type ambiguity set is designed to capture the uncertain moment information (mean and variance) in this work [166]. In such a set as shown below, the moments are assumed to lie in a box region specified by upper and lower bounds [167]:

$$\mathcal{P}_t^2 = \left\{ \mathbb{P}_t \in \mathcal{P}_t^0(W_t) \left| \begin{array}{l} \mathbb{P}_t\{w_t \in W_t\} = 1, \mathbb{E}_{\mathbb{P}_t}\{w_t\} = \mu_t \\ \mathbb{E}_{\mathbb{P}_t}\{(w_t - \mu_t)^2\} = \sigma_t^2 \\ \underline{\mu}_t \leq \mu_t \leq \overline{\mu}_t, \underline{\sigma}_t^2 \leq \sigma_t^2 \leq \overline{\sigma}_t^2 \end{array} \right. \right\} \quad (3.18)$$

where the first and second row in (3.18) have the same definition with those in (3.17), and the third row is used to describe the estimated intervals of unknown mean and variance. Note that μ_t and σ_t^2 in (3.18) are just mathematical symbols and they are unknown compared with those in (3.17).

Based on this ambiguity set, we can obtain the distributionally robust (DR) variant of

chance constraint (3.15):

$$\mathbb{P}_t \left\{ \sum_{i=1}^{N_{dg}} P_i^t + \sum_{j=1}^{N_{ess}} (P_{j,t}^{dch} - P_{j,t}^{ch}) + w_t \geq L_t + \sum_{k=1}^{N_{def}} P_{k,t}^{def} \right\} \geq 1 - \epsilon, \forall \mathbb{P}_t \in \mathcal{P}_t^2. \quad (3.19)$$

Then we have the following DR chance constrained (CC) problem:

$$\min_{\mathbf{x}} \{(3.16) : (3.1) - (3.3), (3.7) - (3.10), (3.12) - (3.14), (3.19)\}.$$

3.3.2 Problem Reformulation

Despite the fact that chance constrained problems with moment ambiguity sets have been investigated by a few works, they mainly concentrate on theoretical derivation and the study on DR-CC problem is still limited. To the best of our knowledge, this is the first application of this method to islanded microgrid energy management with uncertain wind power output. In this subsection, the chance constraint is handled with a conservative approximation, i.e., the sufficient condition of the constraint is first derived, which is, then, analyzed and processed based on our ambiguity set.

To solve the DR-CC problem introduced above, the DR chance constraint with worst-case Conditional Value-at-Risk (CVaR) approximation is first tackled and then it can be transformed into a tractable SOCP constraint. For convenience, we consider the DR chance constraint (3.19) in a general form

$$\mathbb{P}_t \{h^0(x_t) + h(x_t)w_t \leq 0\} \geq 1 - \epsilon, \forall \mathbb{P}_t \in \mathcal{P}_t^2, x_t \in \mathbf{x}. \quad (3.20)$$

Note that we drop the index t of decision variables x_t and other auxiliary variables in the following for simplicity. It has been demonstrated that [165], [168]:

$$\sup_{\mathbb{P}_t \in \mathcal{P}_t^2} \mathbb{P}_t\text{-CVaR}_\epsilon(h^0(x) + h(x)w_t) \leq 0 \Rightarrow \inf_{\mathbb{P}_t \in \mathcal{P}_t^2} \mathbb{P}_t \{h^0(x) + h(x)w_t \leq 0\} \geq 1 - \epsilon \quad (3.21)$$

where the CVaR at level ϵ with respect to probability distribution \mathbb{P}_t is defined as follows [169]:

$$\mathbb{P}_t\text{-CVaR}_\epsilon(h^0(x) + h(x)w_t) = \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_t} [(h^0(x) + h(x)w_t - \beta)^+] \right\} \quad (3.22)$$

where $(\theta)^+ = \max\{\theta, 0\}$. Therefore, according to (3.21), we can consider the following conservative approximation which is a sufficient condition to derive (3.20):

$$\sup_{\mathbb{P}_t \in \mathcal{P}_t^2} \mathbb{P}_t\text{-CVaR}_\epsilon(h^0(x) + h(x)w_t) \leq 0. \quad (3.23)$$

Then we can investigate the above worst-case CVaR approximation and the problem solution based on this constraint is also feasible for the original problem. First, the left side of constraint (3.23) can be processed equivalently as follows:

$$\begin{aligned} & \sup_{\mathbb{P}_t \in \mathcal{P}_t^2} \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_t} [(h^0(x) + h(x)w_t - \beta)^+] \right\} \\ &= \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}_t \in \mathcal{P}_t^2} \mathbb{E}_{\mathbb{P}_t} [(h^0(x) + h(x)w_t - \beta)^+] \right\} \end{aligned} \quad (3.24)$$

where the interchange of sup and inf is based on the stochastic saddle point theorem [170]. By introducing another uncertainty set $\mathcal{Q} = \{(\underline{\mu}_t, \underline{\sigma}_t^2) : \underline{\mu}_t \leq \mu_t \leq \bar{\mu}_t, \underline{\sigma}_t^2 \leq \sigma_t^2 \leq \bar{\sigma}_t^2\}$, we can reformulate the inner maximization problem in (3.24) as follows:

$$\sup_{\mathcal{Q}} \sup_{\mathbb{P}_t \in \mathcal{P}_t^1} \mathbb{E}_{\mathbb{P}_t} [(h^0(x) + h(x)w_t - \beta)^+]. \quad (3.25)$$

Further, let $r = h(x)w_t$, then the mean and variance of r are $h(x)\mu_t$ and $(h(x))^2\sigma_t^2$, respectively. Thus the inner maximization problem of (3.25) can be represented equivalently in integral form:

$$\sup_{\xi \in \mathcal{M}} \int_{\mathbb{R}} ((h^0(x) + h(x)w_t - \beta)^+) \xi(dr) \quad (3.26)$$

$$s.t. \int_{\mathbb{R}} \xi(dr) = 1, \int_{\mathbb{R}} r \xi(dr) = h(x)\mu_t \quad (3.27)$$

$$\int_{\mathbb{R}} r^2 \xi(dr) = h^2(x)\sigma_t^2 + (h(x)\mu_t)^2 \quad (3.28)$$

where \mathcal{M} is the cone of nonnegative Borel measures on \mathbb{R} including the decision variable ξ . Based on duality theory and using change of variables, the problem above is transformed into the following equivalent problem [166] (see Appendix A):

$$\inf_{v, \tau, z, s} v + s \quad (3.29a)$$

$$s.t. 4zs \geq \tau^2 + (h(x))^2\sigma_t^2 \quad (3.29b)$$

$$v - h^0(x) + \beta + \tau - h(x)\mu_t - z \geq 0 \quad (3.29c)$$

$$z > 0, v \geq 0. \quad (3.29d)$$

Considering the outer uncertainty set \mathcal{Q} , the constraint (3.25) is thus equivalent to

$$\inf_{v, \tau, z, s} v + s \quad (3.30a)$$

$$s.t. 4zs \geq \tau^2 + \max_{\sigma_t^2} (h(x))^2\sigma_t^2 \quad (3.30b)$$

$$v - h^0(x) + \beta + \tau - \max_{\mu_t} h(x)\mu_t - z \geq 0 \quad (3.30c)$$

$$z > 0, v \geq 0 \quad (3.30d)$$

$$\underline{\mu}_t \leq \mu_t \leq \overline{\mu}_t, \underline{\sigma}_t^2 \leq \sigma_t^2 \leq \overline{\sigma}_t^2. \quad (3.30e)$$

Since $\sigma_t^2 \geq 0$ in (3.30b), we can easily get that $\max(h(x))^2 \sigma_t^2 = (h(x))^2 \overline{\sigma}_t^2$. For the maximization problem in (3.30c), it has the following dual form:

$$\min \eta_1 \overline{\mu}_t - \eta_2 \underline{\mu}_t : \text{s.t. } h(x) = \eta_1 - \eta_2, \eta_1, \eta_2 \geq 0. \quad (3.31)$$

Therefore, combining all the equations above, we can get the equivalent version of constraint (3.23) as follows:

$$\inf_{\beta, v, \tau, z, s, \eta_1, \eta_2} \beta + \frac{1}{\epsilon}(v + s) \leq 0 \quad (3.32a)$$

$$\text{s.t. } 4zs \geq \tau^2 + (h(x))^2 \overline{\sigma}_t^2, h(x) = \eta_1 - \eta_2 \quad (3.32b)$$

$$v - h^0(x) + \beta + \tau - (\eta_1 \overline{\mu}_t - \eta_2 \underline{\mu}_t) - z \geq 0 \quad (3.32c)$$

$$z > 0, v \geq 0, \eta_1 \geq 0, \eta_2 \geq 0. \quad (3.32d)$$

Note that the constraint (3.32b) is a rotated SOCP constraint which can be transformed into a tractable standard SOCP constraint [167]. In summary, the chance constraint (3.19) in the original energy management model is transformed into the SOCP constraint (3.32) for which we only need to determine the auxiliary functions $h^0(x)$ and $h(x)$ with respect to decision variables from (3.19).

Thus, based on the CVaR approximation of the DR chance constraint, the islanded microgrid energy management problem is reformulated as follows:

$$\min_{x, \Theta} \{(3.16) : (3.1) - (3.3), (3.7) - (3.10), (3.12) - (3.14), (3.32)\}.$$

where Θ is the set of auxiliary variables including $\beta, v, \tau, z, s, \eta_1$ and η_2 which are introduced in the above reformulation process. Despite the conservatism of CVaR constraint, this approximation is advantageous since the original problem can be reformulated as a tractable SOCP problem. Note that the CVaR approximation in (3.21) is actually equivalent when the chance constraint function is concave in w_t [165], [166].

3.3.3 Including Unimodality Information

DRO method intends to find the optimal solution of the problem considering the worst-case distribution in the ambiguity set. However, the worst-case distribution which usually consists of some discrete points [171] is rarely encountered in practice. Thus, only considering the moment information in the ambiguity set will lead to a very conservative solution. In this regard, the unimodality information or strengthened supports can be investigated to reduce the conservatism [172]. In this work, the inclusion of α -unimodality is further studied [173], i.e., the unimodality information of random variable is assumed

to be known except for the moment information in the ambiguity set.

The α -unimodality is defined as follows [174]: for any fixed positive α , a random variable ω is said to have an α -unimodal distribution with mode 0 if $q^\alpha \mathbb{E}[g(q\omega)]$ is nondecreasing in $q > 0$ for every bounded, nonnegative, Borel measurable function g on \mathbb{R}^n . Based on the moment and unimodality information, let $\tilde{\mu}_t = \frac{\alpha+1}{\alpha} \mu_t$, $\tilde{S}_t = \frac{\alpha+2}{\alpha} \sigma_t^2$, we have the following inequation according to [173]:

$$\sup_{\mathbb{P}_t \in \mathcal{P}_t^1} \mathbb{E}_{\mathbb{P}_t} [(h^0(x) + h(x)w_t - \beta)^+] \geq \sup_{\mathbb{P}_t \in \mathcal{P}_t^1(\tilde{\mu}_t, \tilde{S}_t)} \mathbb{E}_{\mathbb{P}_t} [(L(w_t) - \beta)^+] \quad (3.33)$$

where $\mathcal{P}_t^1(\tilde{\mu}_t, \tilde{S}_t)$ is defined similarly as \mathcal{P}_t^1 in (3.17) with the mean $\tilde{\mu}_t$ and variance \tilde{S}_t , and $L(w_t) = h^0(x) + (\frac{\alpha}{\alpha+1})h(x)w_t$. Combining (3.33) with (3.24) and (3.25), we can derive the following constraint from (3.23):

$$\inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathcal{Q}} \sup_{\mathbb{P}_t \in \mathcal{P}_t^1(\tilde{\mu}_t, \tilde{S}_t)} \mathbb{E}_{\mathbb{P}_t} [(L(w_t) - \beta)^+] \right\} \leq 0. \quad (3.34)$$

With the same reformulation method introduced in subsection 3.3.2, the equivalent version of (3.34) can be attained similarly as (3.32), given by:

$$\inf_{\beta, v, \tau, z, s, \eta_1, \eta_2} \beta + \frac{1}{\epsilon} (v + s) \leq 0 \quad (3.35a)$$

$$s.t. \quad 4zs \geq \tau^2 + \frac{\alpha+2}{\alpha} (h(x))^2 \overline{\sigma_t^2}, \quad h(x) = \eta_1 - \eta_2 \quad (3.35b)$$

$$v - h^0(x) + \beta + \frac{\alpha}{\alpha+1} \tau - (\eta_1 \overline{\mu}_t - \eta_2 \underline{\mu}_t) - \left(\frac{\alpha}{\alpha+1}\right)^2 z \geq 0 \quad (3.35c)$$

$$z > 0, \quad v \geq 0, \quad \eta_1 \geq 0, \quad \eta_2 \geq 0. \quad (3.35d)$$

Note that α is set to 1 in this study considering the wind power characteristic and the nesting property of α -unimodality [175]. Thus, the DR-CC microgrid energy management problem based on the moment and unimodality assumption is reformulated as follows:

$$\min_{x, \Theta} \{(3.16) : (3.1) - (3.3), (3.7) - (3.10), (3.12) - (3.14), (3.35)\}.$$

The inclusion of unimodality information is expected to induce a less conservative solution theoretically.

3.4 Case Study

In this section, a case study is conducted to evaluate the proposed DRO method in solving the CC islanded microgrid energy management. The studied microgrid configuration

Table 3.2: Parameters of conventional generators

Unit	P_i^{min} (kW)	P_i^{max} (kW)	R_i^{up}/R_i^{down} (kW)	a_i (\$/kWh ²)	b_i (\$/kWh)	c_i (\$/h)	$d_i(\times 10^{-4})$ (kg/kWh ²)	$e_i(\times 10^{-4})$ (kg/kWh)	$f_i(\times 10^{-4})$ (kg/h)
G1	10	150	30	0.02	0.05	0.1	6.49	-5.55	4.09
G2	8	135	25	0.1	0.04	0.14	5.64	-6.05	2.54
G3	15	280	40	0.01	0.02	0.02	3.38	-3.55	5.33

Table 3.3: Parameters of ESS

E^{min} (kWh)	E^{max} (kWh)	$P^{ch,max}$ (kW)	$P^{dch,max}$ (kW)	η^{ch}	η^{dch}
40	180	100	100	0.95	0.95

and relevant datasets are first described. Then the simulation results and discussion are presented. The microgrid energy management is implemented over a finite time horizon (e.g., $T=24$ hours) in this study and the time step is set to be 1 hour. All the experiments are performed in MATLAB with the modeling tool YALMIP [176] and CPLEX 12.71 solver on a desktop with an Intel Core i7-6700 CPU 3.40 GHz and 8 GB of RAM.

3.4.1 Description of Microgrid

In this work, a microgrid composed of three conventional generators, an ESS, a wind turbine, a critical load and a deferrable load [61] is considered. For the three generators, their parameters are given in Table 3.2 which are collected and modified from [59], [156]. The same emission coefficients are used here and the emission cost coefficient is 1 \$/kg. In addition, an ESS with storage capacity of 200 kWh is deployed in the microgrid whose parameters are summarized in Table 3.3. The initial and final energy levels of ESS are set to be half of its capacity and the degradation cost coefficient is 0.0035 \$/kWh [61]. Note that the studied microgrid system is a general test system which is developed based on previous literature. More cases or test systems may be studied in the future with the proposed solution method, and the results are expected to change accordingly.

As for wind power, we only need the upper and lower bounds of the mean and variance. Based on the estimated mean and variance taken from [100], the upper and lower bounds can be obtained by deviating 10% from the rated values. The mean values of wind power are illustrated in Fig. 3.1 as an example. Moreover, the critical load demand from [73] is approximately scaled and used as shown in Fig. 3.1. Note that only the real power is considered here. The deferrable load is assumed to have a total demand of 100 kWh which needs to be satisfied between 13th and 18th time slot. The minimum and maximum serving rates for deferrable load are set to be 10 kW and 50 kW, respectively. The chance constraint confidence level is set as $1 - \epsilon = 95\%$.

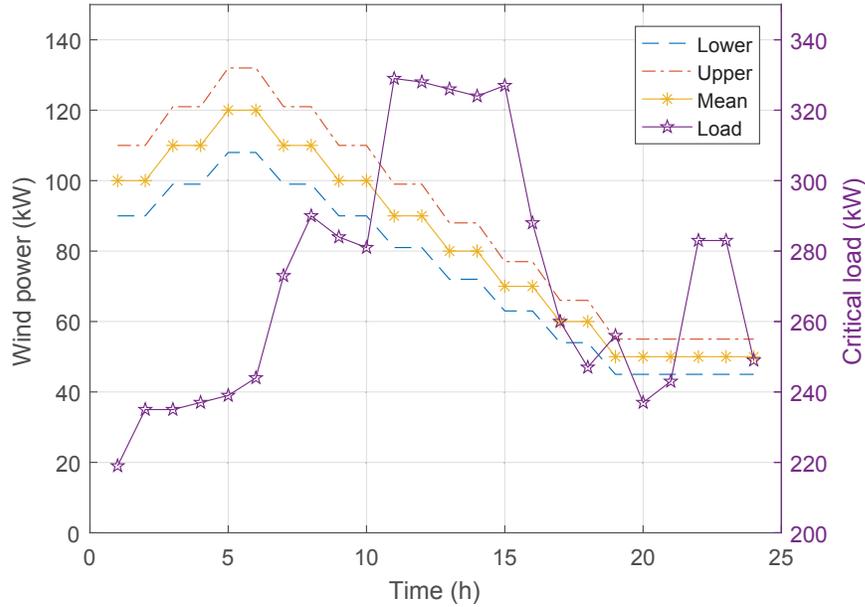


Figure 3.1: Wind power and critical load.

3.4.2 Simulation Results

With the parameters and datasets introduced above, the DR-CC microgrid energy management problem is solved in this subsection. The output power of conventional generators is first analyzed. The optimal energy schedule of the three generators is shown in Fig. 3.2. As can be seen from this figure, the output of three generators has great difference: the unit G3 has the largest output, followed by the unit G1, while the unit G2 produces the least output power. This phenomenon is consistent with the generation cost as we can find that unit G3 has the smallest generation cost coefficients. However, the generation cost coefficients of unit G2 are quite large. Additionally, the emission cost only takes a small proportion compared with the generation cost which will be introduced later.

ESS plays a significant role in microgrid energy management which helps achieve peak load shifting. The energy schedule of ESS including the stored energy level, the charging and discharging power is shown in Fig. 3.3. As we can see from this figure, the ESS charges in the first few hours when the load demand is low, it stays unchanged for several hours when the maximum capacity is reached. As the load increases later, the generators produce more power and the ESS starts to discharge. It can also be found that the change of ESS state is influenced by the load change most of the time. In addition, the energy schedule of deferrable load (e.g., EV) is also presented in Fig. 3.3. The deferrable load is served almost at a uniform rate over the specified time interval except the sudden increase in the 17th and 18th period. The total microgrid energy management cost in this case is \$9847.5.

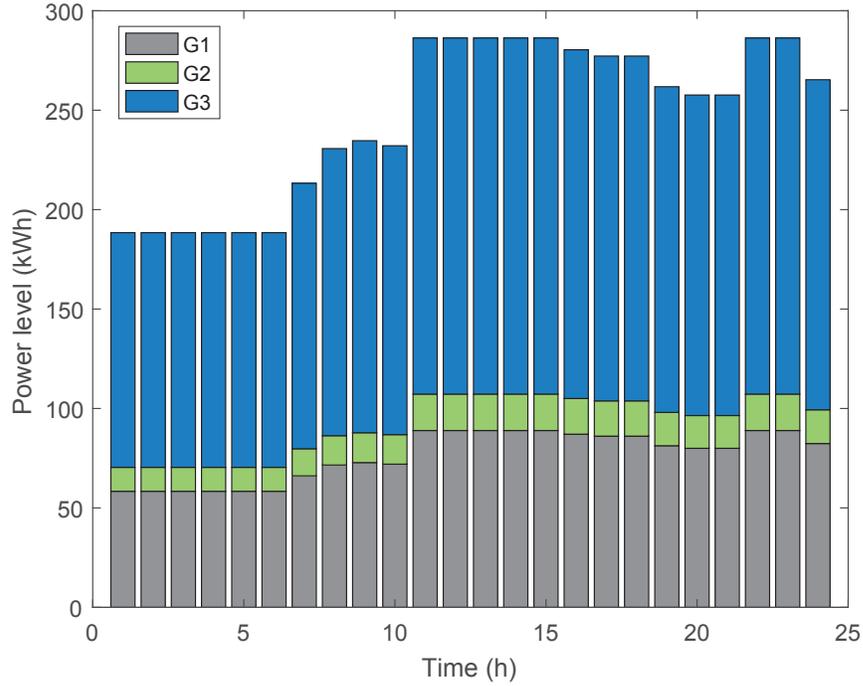


Figure 3.2: Optimal power schedule of conventional generators.

For the method considering the unimodality information, the energy management has a similar dispatch solution which has been omitted here. However, the total system cost with this method is \$8381.5 which substantiates the improved conservatism by incorporating the unimodality information.

3.4.3 Discussion

1) *Reliability and robustness of our method:* The DRO method solves the CC islanded microgrid energy management with a worst-case CVaR approximation. To validate the effectiveness of our solutions, we use Monte Carlo simulation method to test the reliability and robustness of our approach. Based on the estimated mean and variance values of wind power, a million scenarios are randomly generated by assuming a normal distribution. With these wind power samples, our energy management solutions are checked. The percentage of the scenarios that satisfy the power balance constraint is up to 99.99934%, which is higher than the setting confidence level 95%. In addition, this probability value ensures that the chance constraints violation probability in the total scheduling horizon is very low. Similarly, the method with unimodality information can also be checked and the percentage is about 99.9919%. Therefore, this result further verifies that our method is reliable and robust against the unknown probability distributions sharing the same moment information.

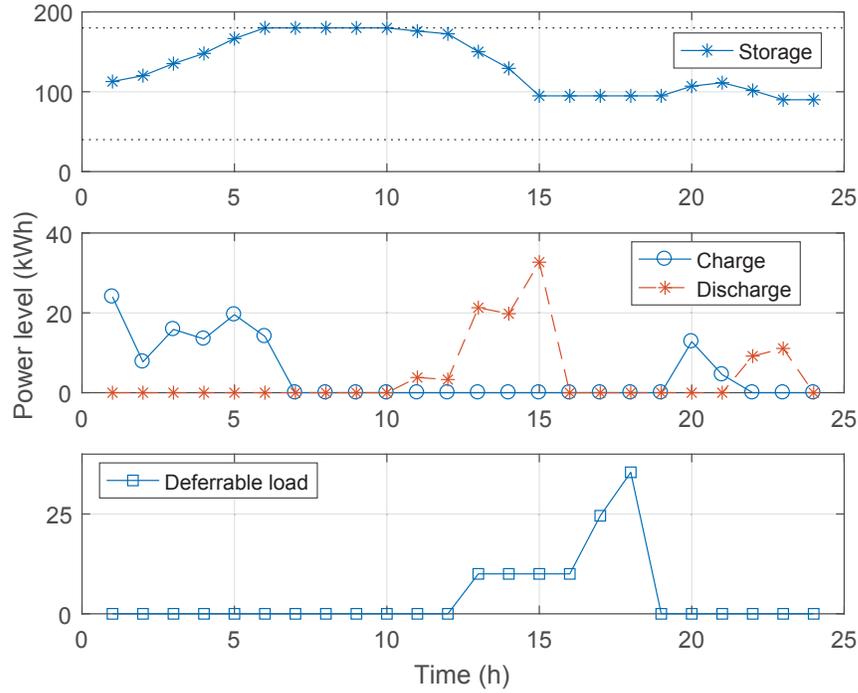
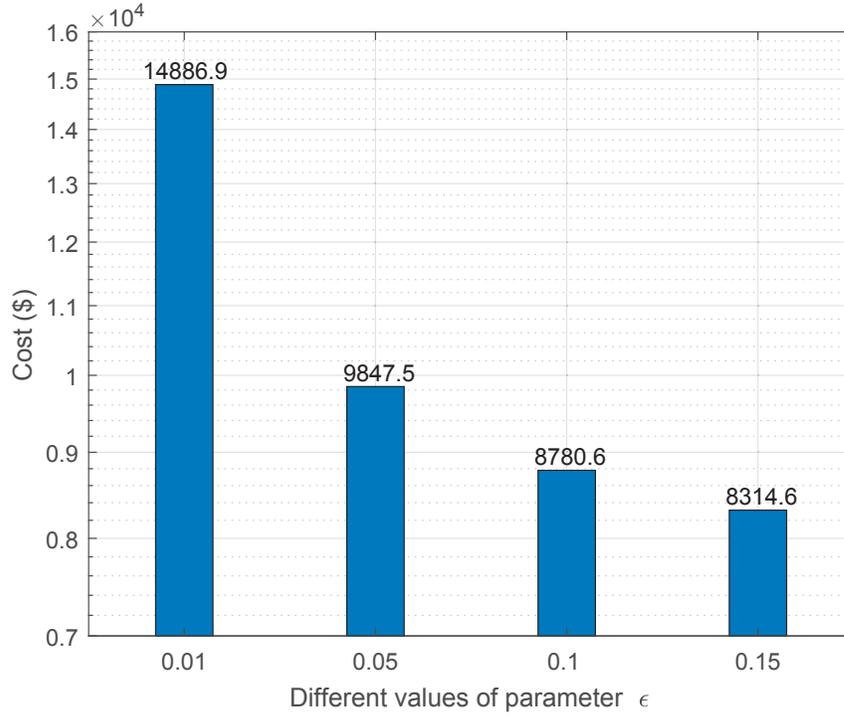


Figure 3.3: Optimal energy schedule of ESS and deferrable load.

2) *Influence of different parameter ϵ and interval size*: Various confidence levels will result in various solutions in microgrid energy management. Moreover, the uncertainty of microgrid system can be controlled by adjusting the confidence level or the parameter ϵ . Although the confidence level $1 - \epsilon = 95\%$ is mostly studied, the influence of some other parameter settings is investigated in this study. The total system cost with different parameter ϵ , including 0.01, 0.05, 0.1 and 0.15, are summarized in Fig. 3.4 and the corresponding energy management solutions are omitted for simplicity.

As can be seen from this figure, the smaller the ϵ value, the higher the total system cost. A small ϵ value represents a high confidence level. In other words, we have to pay more to achieve a more reliable system. In addition, comparing the difference between the neighbouring system cost, we can find that a dramatic decrease occurs when the parameter ϵ increases from 0.01 to 0.05 which means that the marginal cost becomes larger as the confidence level increases. Thus, we have to select a proper confidence level in practice to avoid the extremely high cost caused by a strict reliability constraint.

The critical feature of the proposed ambiguity set is the inclusion of the interval for uncertain mean and variance. To investigate the impacts of the interval on the simulation results, the interval size is enlarged gradually. In this case study, the intervals for mean

Figure 3.4: System cost with different ϵ .

and variance are attained as follows: $(1 \pm \lambda_{\mu})\mu_t$ and $(1 \pm \lambda_{var})\sigma_t^2$, where the deviation coefficients λ_{μ} and λ_{var} are increased from 0.1 to 0.4 with a step size of 0.1. The corresponding results of total cost are plotted in Fig. 3.5 with interpolation, from which we can see that the system total cost increases as the intervals expand. Also, the intervals for mean values have a larger impact on the cost in this case caused by the larger nominal values. Therefore, to reduce the total cost, the interval in the ambiguity set should be shortened in practice. In this regard, different data-driven techniques can be used to estimate the interval based on historical dataset, such as the popular direct interval forecast methods [31].

3) *Comparison with other methods*: In this part, the proposed method denoted as M_1 is compared with the DRO method based on known moment information (e.g., mean and variance), sample average approximation (SAA) method and stochastic optimization with normal distribution (SND) in solving the CC microgrid energy management problem. The method with known moment information is denoted as M_2 , where the ambiguity set \mathcal{P}_t^1 is used instead of \mathcal{P}_t^2 and the other constraints remain unchanged. In addition, the inclusion of unimodality information in M_1 and M_2 is also studied, denoted as M_1^{uni} and M_2^{uni} , respectively. In these methods, the confidence level is set as 95%. Note that the CC problem is usually transformed into a semidefinite problem (SDP) with the method M_2 [165]. The SDP problem can be solved directly by off-the-shelf solvers or we can use

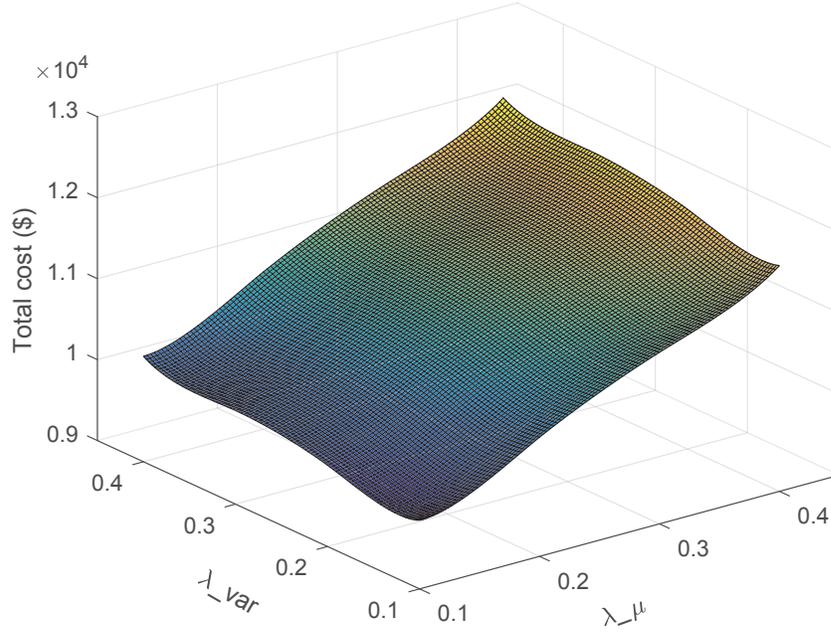


Figure 3.5: System cost with different interval size.

the same method in Section 3.3 to further reformulate it as a tractable SOCP problem.

SAA is an effective method to cope with CC problems and a number of theoretical research can be found in previous literature [177], [178]. The basic idea is to approximate the true distribution in chance constraint with an empirical distribution obtained from Monte Carlo sampling technique. For the chance constraint $Pr\{h^0(x) + h(x)w_t > 0\} \leq \epsilon$ in this work, it can be approximated with SAA method as follows:

$$N_s^{-1} \sum_{l=1}^{N_s} \mathbb{I}_{(0,\infty)}(h^0(x) + h(x)w_t^l) \leq \gamma \quad (3.36)$$

where $\mathbb{I}_{(0,\infty)}$ is the indicator function of $(0, \infty)$, i.e., $\mathbb{I}(y) = 1$ if $y > 0$, otherwise, $\mathbb{I}(y) = 0$; N_s is the number of samples and γ is the risk level of the SAA chance constraint; w_t^l represents the sample of the random variable. According to [177], the SAA chance constraint can be replaced with mixed-integer constraints, thus the SAA problem is a mixed-integer quadratic programming (MIQP) problem. In this case study, N_s and γ are set to 500 and 0.05, respectively, and the samples are generated from a normal distribution with nominal mean and variance. More details about the SAA method can be found in relevant references.

For the SND method, it is assumed that the uncertain wind power follows a normal distribution with deterministic mean and variance. With this assumption, the original

Table 3.4: Comparison of system cost with different methods

Method	Generation (\$)	Emission (\$)	ESS (\$)	Total (\$)
M_1	9553.9	292.9	0.7	9847.5
M_1^{uni}	9101.8	278.9	0.8	9381.5
M_2	8816.8	270.0	0.7	9087.5
M_2^{uni}	8403.0	257.1	0.7	8660.8
SAA	22542.6	697.9	0.8	23241.3
SND	6978.3	212.9	0.8	7192.0

chance constraint can be transformed into a deterministic constraint as follows:

$$\{Pr\{h^0(x) + h(x)w_t \leq 0\} \geq 1 - \epsilon\} = \{h(x)\mu_t + h^0(x) + |h(x)|\sigma_t z_\epsilon \leq 0\} \quad (3.37)$$

where $z_\epsilon = \Phi^{-1}(1 - \epsilon)$ is the $(1 - \epsilon)$ quantile of standard normal distribution. In this case, the mean and variance are also set to the nominal values introduced before.

The system cost results of these different methods are summarized in Table 3.4. It can be seen that the cost of our method (M_1) is higher than that of M_2 which implies that our method with the ambiguity set \mathcal{P}_t^2 is more conservative. However, our method M_1 is more reliable and robust than M_2 whose CC satisfaction percentage is 99.909% calculated from the method discussed above. This comparison gives us an intuitive effect of the method M_1 and the result is rational since M_1 assumes that less information is known about the uncertain wind power compared with M_2 . Also, the inclusion of unimodality information produces a less conservative solution for both methods as expected. Additionally, we can also find that the generation cost accounts for a considerable proportion of the total system cost compared with the emission cost and ESS degradation cost. This large difference is mainly caused by the setting of different cost coefficients. This verifies the analysis of different output power of conventional generators in the simulation results.

As the SAA method is based on Monte Carlo sampling, the system cost may vary for each run. Hence, the simulation is repeated for ten times and the average cost is calculated as given in Table 3.4. As can be seen, the result of SAA is more conservative and its performance is not as good as M_1 despite the fact that the samples are generated from a normal distribution. For SND method, we can see that the system cost is lower due to the more certain information about the uncertain wind power in the assumption.

3.5 Summary

In this chapter, distributionally robust CC energy management is studied for isolated microgrids. The CC energy management model is first designed for an isolated microgrid which consists of conventional generators, ESS, wind turbines and various load demand.

In addition to the common generation cost in the objective function, the emission cost and ESS degradation cost are also considered in this model. The uncertainty of wind power is captured by a novel ambiguity set in this work, based on which the individual chance constraint can be tackled with the DRO method and the microgrid energy management problem is reformulated as a tractable SOCP problem. The proposed method has been analyzed through a case study and the simulation results show its effectiveness and reliability. Moreover, the comparison with the approach with known moment information validates the robustness of the proposed method, which is more applicable in practice. The comparison with SAA and stochastic optimization method also reveals the advantage of the proposed method.

4

Data-driven Distributionally Robust UC and ED

Acronyms

ARUC	Adjustable Robust Unit Commitment
CVaR	Conditional Value-at-risk
C&CG	Column-and-constraint Generation
DDRC	Data-driven Distributionally Robust Chance-constrained
DRED	Distributionally Robust Economic Dispatch
DRUC	Distributionally Robust Unit Commitment
ED	Economic Dispatch
DRO	Distributionally Robust Optimization
MILP	Mixed-integer Linear Programming
RUC	Robust Unit Commitment
SED	Stochastic Economic Dispatch
SUC	Stochastic Unit Commitment
UC	Unit Commitment

4.1 Introduction

UC and ED are two basic and important problems in power system operation which have been the research focus for a long time. In recent years, the deployment and utilization of renewable energy sources has increased significantly in power system operation, which also brings many new challenges to the reliability and security of power systems. Consequently, it is necessary to investigate new uncertainty modeling methods for UC and ED problems with uncertain renewable generation.

In this chapter, a data-driven distributionally robust chance-constrained (DDRC) UC problem with uncertainty is studied which is formulated as a two-stage model. The commitment decision and base-case dispatch plan are determined in the first stage with a chance constraint considering possible power imbalance, and the expected worst-case re-dispatch cost is minimized in the second stage. Compared with the popular stochastic programming or robust methods in previous literature, DRO method is studied in this work, and the uncertainty of wind power distribution is captured by a distance-based ambiguity set, more specifically, a set with the form of L_1 norm which can be constructed from historical data. Based on the ambiguity set, the proposed complex DDRC UC problem can be reformulated into a tractable optimization problem, thus solved by some existing decomposition algorithms such as the column-and-constraint generation (C&CG) algorithm.

In addition, new types of reserves have been considered to compensate the uncertainty of renewable generation which also makes the co-optimization of energy and reserve dispatch important. Similarly, DRO method is studied in this chapter for a data-driven two-stage energy and reserve dispatch problem. The proposed two-stage model minimizes the generation and reserve cost with forecasted wind power in the first stage and minimizes the expected re-dispatch cost considering the worst-case probability distribution in the second stage. An ambiguity set based on L_∞ norm is investigated in this problem which is different from those studied in previous literature. Based on the proposed ambiguity set, the second-stage worst-case expectation is reformulated into a combination of the conditional value-at-risk (CVaR) and expected cost with respect to a reference distribution, thus the original two-stage DRO problem becomes a two-stage stochastic linear program problem which can be readily solved.

4.2 Data-driven Distributionally Robust Chance-constrained UC

This section aims to study the chance-constrained UC problem with a data-driven DRO method. In subsection 4.2.1, the problem formulation is introduced including the UC model and ambiguity set. Detailed solution methodology is presented in 4.2.2. Then case studies are conducted based on the IEEE 6-bus test system and modified IEEE 118-bus system and a practical-scale 319-bus system to validate the performance of the proposed approach in subsection 4.2.3.

4.2.1 Problem Formulation

In this subsection, the DDRC two-stage UC problem is first formulated which includes various constraints and objective function. The ambiguity set and its construction are then introduced to capture the uncertain distribution of renewable generation (i.e., wind

power).

4.2.1.1 UC Mathematical Model

Two-stage UC models are widely studied in previous literature, and most of them focus on commitment decision in the first stage and optimal recourse in the second stage. For example, two-stage robust UC problems are studied in [179] [180], and two-stage UC problems considering distributional uncertainty are investigated in [80] [181]. In these works, the here-and-now decision variables are usually commitment decision and wait-and-see decision variables are recourse action. By contrast, a new DDRC two-stage UC model is developed in this work which considers the traditional UC model with a chance constraint in the first stage and attains the best corrective actions by minimizing the expected re-dispatch cost in the second stage. In addition, two-stage energy and reserve dispatch problems are also widely studied [102] [182] which considers the base-case dispatch plan in the first stage and re-dispatch in the second stage, while UC decisions are not covered in these works. From this aspect, the proposed UC model can be regarded as a combination of traditional two-stage UC problem with the energy and reserve dispatch problem, and it is solved with the new DRO method instead of the previous stochastic or robust optimization methods. The detailed formulation of the proposed model is as follows:

$$\min \sum_t \sum_i [SU_i u_{it} + SD_i v_{it} + F_i(x_{it})] + \max_{P \in \mathcal{D}} EP[Q(y, u, v, x, \xi)] \quad (4.1)$$

$$s.t. -y_{i(t-1)} + y_{it} - y_{ih} \leq 0, \forall t \in \mathcal{T}, \forall i \in \mathcal{I}, 1 \leq h - (t-1) \leq T_i^{up} \quad (4.2)$$

$$y_{i(t-1)} - y_{it} + y_{ih} \leq 1, \forall t \in \mathcal{T}, \forall i \in \mathcal{I}, 1 \leq h - (t-1) \leq T_i^{dn} \quad (4.3)$$

$$-y_{i(t-1)} + y_{it} - u_{it} \leq 0, \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \quad (4.4)$$

$$y_{i(t-1)} - y_{it} - v_{it} \leq 0, \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \quad (4.5)$$

$$y_{it}, u_{it}, v_{it} \in \{0, 1\}, \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \quad (4.6)$$

$$\underline{x}_i y_{it} \leq x_{it} \leq \bar{x}_i y_{it}, \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \quad (4.7)$$

$$x_{it} - x_{i(t-1)} \leq \overline{RU}_i u_{it} + RU_i y_{i(t-1)}, \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \quad (4.8)$$

$$x_{i(t-1)} - x_{it} \leq \overline{RD}_i v_{it} + RD_i y_{it}, \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \quad (4.9)$$

$$-L_l \leq \sum_{b \in \mathcal{B}} K_l^b \left(\sum_i x_{it}^b + w_t^b(\xi) - d_t^b \right) \leq L_l, \forall t \in \mathcal{T}, l \in \mathcal{G} \quad (4.10)$$

$$Pr(-\delta \leq \sum_i x_{it} + \sum_b w_t^b(\xi) - \sum_b d_t^b \leq \delta) \geq 1 - \epsilon_1, \forall t \in \mathcal{T} \quad (4.11)$$

where the objective function in (4.1) contains the start-up cost, shut-down cost, fuel cost and worst-case expected penalty cost. Constraints (4.2) and (4.3) represent the minimum up-time and minimum down-time constraints, respectively. Constraints (4.4) and (4.5) restrict the start-up and shut-down operation, respectively. Constraint (4.6) lists the

binary variables representing the generators' statuses of on/off, start-up and shut-down. Constraint (4.7) represents the generation capacity limits. Constraints (4.8) and (4.9) enforce the ramp-up and ramp-down rates, respectively. Constraint (4.10) denotes the power transmission line capacity limits which is from DC power flow model. Note that the variable x_{it}^b represents the output of the i th generator located at bus b at time t . Constraint (4.11) defines the chance constraint for possible power imbalance, and the small violation probability should be less than a predefined risk level.

After determining the commitment decision and base-case dispatch plan, the system operators conduct re-dispatch strategy by considering all uncertainty realizations in the second stage. Discrete scenarios are often used to replace the continuous distribution to solve the difficult numerical computation. In this work, it is assumed that the uncertain parameter ξ has a finite support, i.e., there are a finite number of realizations (e.g., scenarios $\xi_1, \xi_2, \dots, \xi_N$) for the uncertain wind power output [90]. However, the true probability distribution is unknown here and is described by the ambiguity set.

In formulation (4.1), the operational risk for the second stage problem is considered which also represents the expected penalty cost or re-dispatch cost [95]. This cost is caused by load curtailment or over-generation of the system with the reveal of uncertain wind power. Note that wind power curtailment is not considered since finite scenarios are assumed in this work as mentioned above. In other words, we need to re-adjust the generation or consider load shedding with possible worst-case wind power in the second stage. Additionally, we should note that both system-level and nodal-level uncertainty modeling should be investigated to identify the real worst case. Specifically, we have the following formulation for the second stage problem:

$$Q(y, u, v, x, \xi) = \min \sum_t [\pi_t^{gen} \sum_i (r_{it}^u + r_{it}^d) + \pi_t^{ls} \sum_b d_t^{ls,b}] \quad (4.12)$$

$$s.t. \sum_i (r_{it}^u - r_{it}^d) + \sum_b d_t^{ls,b} = \sum_b d_t^b - \sum_i x_{it} - \sum_b w_t^b(\xi), \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \quad (4.13)$$

$$x_{it} + r_{it}^u \leq \bar{x}_i y_{it}, \forall t \in \mathcal{T}, i \in \mathcal{I} \quad (4.14)$$

$$x_{it} - r_{it}^d \geq \underline{x}_i y_{it}, \forall t \in \mathcal{T}, i \in \mathcal{I} \quad (4.15)$$

$$0 \leq r_{it}^u \leq RU_i, \forall t \in \mathcal{T}, i \in \mathcal{I} \quad (4.16)$$

$$0 \leq r_{it}^d \leq RD_i, \forall t \in \mathcal{T}, i \in \mathcal{I} \quad (4.17)$$

$$\sum_{b \in \mathcal{B}} K_l^b \left(\sum_i x_{it}^b + w_t^b(\xi) - d_t^b + \Delta p_t^b \right) \leq L_l, \forall t \in \mathcal{T}, l \in \mathcal{G} \quad (4.18)$$

$$\sum_{b \in \mathcal{B}} K_l^b \left(\sum_i x_{it}^b + w_t^b(\xi) - d_t^b + \Delta p_t^b \right) \geq -L_l, \forall t \in \mathcal{T}, l \in \mathcal{G} \quad (4.19)$$

$$\Delta p_t^b = \sum_i (r_{it}^u - r_{it}^d) + d_t^{ls,b}, d_t^{ls,b} \geq 0, \forall t \in \mathcal{T} \quad (4.20)$$

where the objective function in (4.12) includes the possible re-dispatch cost of generators and load shedding cost. Constraint (4.13) denotes the system power balance with re-dispatch action. The re-dispatch amount of each unit should be limited by the available generator capacity as given in (4.14)-(4.15) and the corresponding ramp rate given in (4.16)-(4.17). The set of constraints (4.18) and (4.19) show the transmission line flow considering the adjusted dispatch action as given in (4.20). Note that the re-dispatch variables in (4.20) correspond to the generator units in (4.19) and the symbol b is omitted here for consistency. In addition, it is assumed that all generators can flexibly adjust their output in the re-dispatch process with corresponding ramping rate in this work [71].

In the objective function, the fuel cost function of generation units $F_i(x_{it})$ is typically a non-decreasing quadratic function. For practical and computational purposes, the quadratic fuel cost function is usually approximated by a piece-wise linear function. In this work, the following piece-wise linear model is used for fuel cost calculation [183] [184]:

$$0 \leq \hat{x}_{it}^k \leq \Delta x_i^k y_{it}, \forall k = 1, \dots, K_0 \quad (4.21a)$$

$$\Delta x_i^k = \frac{\bar{x}_i - \underline{x}_i}{K_0} \quad (4.21b)$$

$$x_{i,ini}^k = (k-1)\Delta x_i^k + \underline{x}_i \quad (4.21c)$$

$$x_{i,fin}^k = \Delta x_i^k + x_{i,ini}^k \quad (4.21d)$$

$$C_{i,ini}^k = a_i(x_{i,ini}^k)^2 + e_i x_{i,ini}^k + c_i \quad (4.21e)$$

$$C_{i,fin}^k = a_i(x_{i,fin}^k)^2 + e_i x_{i,fin}^k + c_i \quad (4.21f)$$

$$s_i^k = \frac{C_{i,fin}^k - C_{i,ini}^k}{\Delta x_i^k} \quad (4.21g)$$

$$x_{it} = \underline{x}_i y_{it} + \sum_k \hat{x}_{it}^k \quad (4.21h)$$

$$F_i(x_{it}) = a_i(\underline{x}_i)^2 + e_i \underline{x}_i + c_i y_{it} + \sum_k s_i^k \hat{x}_{it}^k \quad (4.21i)$$

where new variable \hat{x}_{it}^k is introduced, and its relationship with the decision variable x_{it} is described in (4.21h). In addition, K_0 is the number of pieces, and s_i^k is the slope of each linear piece. The coefficients a_i , e_i and c_i are dependent on the physical characteristic of the generators.

4.2.1.2 Ambiguity Set Construction

As discussed above, the true probability distribution of wind power is unknown and ambiguous in practice. However, we can get partial information about the true distribution from available historical data and construct the ambiguity sets to capture the uncertainty of distribution. In this work, a distance-based ambiguity set is studied which has the following form:

$$\mathcal{D} = \{P \in \mathcal{P} : \text{dist}(P, P_0) \leq \theta\} \quad (4.22)$$

where \mathcal{P} is the set of all possible distributions, P_0 and θ are the nominal distribution and divergence tolerance level, respectively. The $\text{dist}()$ function denotes a distance measure between two distributions, such as the KL divergence [88]. Since discrete distribution for wind power is considered in this study, the L_1 -norm distance is adopted to construct the ambiguity set. In addition, the nominal distribution derived from historical data tends to converge to the true distribution under L_1 -norm as the data size increases [90]. Consequently, the ambiguity set used in this work, denoted by \mathcal{D}_1 , can be expressed as follows:

$$\mathcal{D}_1 = \{p \in [0, 1]^N : \sum_{n=1}^N |p^n - \hat{p}^n| \leq \theta, \sum_{n=1}^N p^n = 1\} \quad (4.23)$$

where p^n and \hat{p}^n are the true probability and nominal probability respectively corresponding to index n , and N is the number of scenarios. Note that the ambiguity set \mathcal{D}_1 is a specific set compared with the general form \mathcal{D} , and the true probability in this set is unknown which can be described with the nominal probability estimated from historical data.

To construct the set \mathcal{D}_1 , a critical step is the determination of nominal distribution and tolerance level. For the nominal or reference distribution, we can derive it with nonparametric estimation method in a data-driven manner. Specifically, assuming that there are A historical data samples available in total, we can estimate the reference distribution with a histogram. For example, according to the number of scenarios, we can construct a histogram with N bins. Count the number of data samples in each bin, say, A_1, A_2, \dots, A_N and $A = \sum_{n=1}^N A_n$, then we can use the frequency A_n/A in each bin as the nominal probability \hat{p}^n . For simplicity, the nominal distribution is denoted as $P_0 = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_N)$. Note that the nominal distribution from the histogram is only an estimation of the true distribution and it may be obtained from other methods.

Based on the size of available historical data, we can also define proper tolerance level (i.e., θ) to construct effective ambiguity set. Following the above histogram approach, we can determine the tolerance level according to the Proposition 8 in [185] as follows:

$$\theta = \sqrt{\chi_{N-1, 1-\tilde{\alpha}}^2 / A} \quad (4.24)$$

where $1 - \tilde{\alpha}$ is the confidence level that the data-driven ambiguity set \mathcal{D}_1 with nominal distribution P_0 and θ contains the unknown real distribution. From (4.24), we can find that the value of θ is mainly determined by the confidence level and historical data size. Since the value of θ decreases as the size of historical data increases, the true distribution becomes much closer to the reference distribution. When the data size goes to infinity, θ will be zero. In addition, since $\sum_{n=1}^N |p^n - \hat{p}^n| \leq \sum_{n=1}^N (p^n + \hat{p}^n) = 2$, we can obtain the θ value limit which should fall into $[0, 2]$ interval. The effect of θ value will be further analyzed in the numerical experiments.

4.2.2 Solution Methodology

In this subsection, the solution methodology is introduced which first reformulates the original UC problem including transformation of the chance constraints and objective function. With the proposed reformulation technique, the original complex UC problem becomes tractable, and the conservativeness can be controlled flexibly in a data-driven manner. Then, based on the problem structure, a decomposition algorithm (i.e., C&CG algorithm) is introduced to solve the problem.

4.2.2.1 Problem Reformulation

To solve the UC problem, we need to focus on two critical points in the problem formulation. One is the data-driven chance constraint for power imbalance, and the other is the two-level objective function. Since the distribution of wind power is ambiguous and defined within a distance-based ambiguity set, worst-case distribution should be considered to ensure the chance constraint. In other words, the chance constraint in (4.11) can be recast as follows:

$$\min_{P \in \mathcal{D}_1} Pr(-\delta \leq \sum_i x_{it} + \sum_b w_t^b(\xi) - \sum_b d_t^b \leq \delta) \geq 1 - \epsilon_1, \forall t \in \mathcal{T}. \quad (4.25)$$

Based on the studied discrete distribution, the above inequality can be further reformulated as [91]

$$\min_{P \in \mathcal{D}_1} \sum_{n=1}^N p_t^n \cdot \mathbb{I}_{[-\delta, \delta]}(\sum_i x_{it} + \sum_b w_t^b(\xi^n) - \sum_b d_t^b) \geq 1 - \epsilon_1, \forall t \in \mathcal{T} \quad (4.26)$$

where $\mathbb{I}_{[-\delta, \delta]}(\sum_i x_{it} + \sum_b w_t^b(\xi^n) - \sum_b d_t^b)$ represents an indicator function. It equals 1 if $-\delta \leq \sum_i x_{it} + \sum_b w_t^b(\xi^n) - \sum_b d_t^b \leq \delta$, otherwise it is 0. Note that similar idea can also be used to deal with robust chance constraint with other distance based ambiguity set such as the KL divergence based set and CVaR method may be investigated to derive a conservative constraint [186].

To simplify the notation, we can introduce binary variables z_t^n to replace the indicator function, i.e., $z_t^n = \mathbb{I}_{[-\delta, \delta]}(\cdot)$. Then, the above constraint (4.26) can be reformulated using big-M method as follows:

$$-\delta - (1 - z_t^n)M \leq \sum_i x_{it} + \sum_b w_t^b(\xi) - \sum_b d_t^b \leq \delta + (1 - z_t^n)M, \forall t \in \mathcal{T}, \forall n \quad (4.27)$$

$$\min_{p_t^n} \sum_{n=1}^N p_t^n z_t^n \geq 1 - \epsilon_1 \quad (4.28)$$

$$\sum_{n=1}^N |p_t^n - \hat{p}_t^n| \leq \theta \quad (4.29)$$

$$\sum_{n=1}^N p_t^n = 1, \forall t \in \mathcal{T}, \forall n \quad (4.30)$$

$$p_t^n \geq 0, \forall t \in \mathcal{T}, \forall n \quad (4.31)$$

where constraints (4.29)-(4.31) are derived from the predefined ambiguity set \mathcal{D}_1 . For constraint (4.29), we can introduce an auxiliary variable q_t^n to eliminate the absolute operation, i.e., let $q_t^n = |p_t^n - \hat{p}_t^n|$, which leads to the following equivalent formulas [90]:

$$\sum_{n=1}^N q_t^n \leq \theta \quad (4.32)$$

$$q_t^n \geq p_t^n - \hat{p}_t^n \quad (4.33)$$

$$q_t^n \geq \hat{p}_t^n - p_t^n. \quad (4.34)$$

Considering the minimization operation in constraint (4.28), we can try to transform the minimization into maximization by duality theory which helps remove the operation in the inequality. Based on the constraints (4.32)-(4.34), (4.30)-(4.31) and the minimization objective, the dual results can be deduced as follows:

$$\max_{\alpha_t, \beta_t^n, \gamma_t^n, \lambda_t} -\alpha_t \theta + \sum_{n=1}^N (-\beta_t^n \hat{p}_t^n + \gamma_t^n \hat{p}_t^n) + \lambda_t \quad (4.35)$$

$$-\beta_t^n + \gamma_t^n + \lambda_t \leq z_t^n, \forall t \in \mathcal{T}, \forall n \quad (4.36)$$

$$-\alpha_t + \beta_t^n + \gamma_t^n \leq 0, \forall t \in \mathcal{T}, \forall n \quad (4.37)$$

$$\alpha_t, \beta_t^n, \gamma_t^n \geq 0, \lambda_t \text{ unrestricted}, \forall t \in \mathcal{T}, \forall n \quad (4.38)$$

where $\alpha_t, \beta_t^n, \gamma_t^n$ and λ_t are corresponding dual variables.

Then the constraints (4.28)-(4.31) can be replaced with the following constraints:

$$-\alpha_t \theta + \sum_{n=1}^N (-\beta_t^n \hat{p}_t^n + \gamma_t^n \hat{p}_t^n) + \lambda_t \geq 1 - \epsilon_1, \forall t \in \mathcal{T} \quad (4.39)$$

$$\text{Constraints (4.36) - (4.38)}. \quad (4.40)$$

In addition to the data-driven chance constraint, the objective function in the proposed UC model also involves the uncertain distribution which hinders the optimization of the problem. To reformulate the objective function, we need to focus on the second-level objective, i.e., the worst-case expected penalty cost $\max_{P \in \mathcal{D}_1} E_P[Q(\mathbf{x}, \xi)]$. Note that \mathbf{x} is used to represent the decision vector for notation brevity. According to Theorem 1 in [185], we can get an equivalent reformulation with the ambiguity set \mathcal{D}_1 as follows:

$$\max_{P \in \mathcal{D}_1} E_P[Q(\mathbf{x}, \xi)] = (1 - \frac{\theta}{2}) \text{CVaR}_{\theta/2}^{P_0}[Q(\mathbf{x}, \xi)] + \frac{\theta}{2} \max_{\xi} Q(\mathbf{x}, \xi) \quad (4.41)$$

where $\text{CVaR}_{\theta/2}^{P_0}[Q(\mathbf{x}, \xi)]$ denotes the conditional value-at-risk of $Q(\mathbf{x}, \xi)$ with respect to the nominal distribution P_0 with confidence level $\theta/2$. In addition, the CVaR is defined as below [187]:

$$\text{CVaR}_{\theta/2}^{P_0} = \min_{\phi} \phi + \frac{1}{1 - \theta/2} E_{P_0}[Q(\mathbf{x}, \xi) - \phi]^+ \quad (4.42)$$

where ϕ is a new free variable and $[Q(\mathbf{x}, \xi) - \phi]^+ = \max\{Q(\mathbf{x}, \xi) - \phi, 0\}$.

By substituting the CVaR, the worst-case expectation objective can be further written as follows:

$$\max_{P \in \mathcal{D}_1} E_P[Q(\mathbf{x}, \xi)] = \min_{\phi} \{(1 - \theta/2)\phi + E_{P_0}[Q(\mathbf{x}, \xi) - \phi]^+\} + \frac{\theta}{2} \max_{\xi} Q(\mathbf{x}, \xi). \quad (4.43)$$

In the above formulation, $E_{P_0}[\cdot]^+$ can be obtained based on the estimated nominal distribution P_0 . For $\max_{\xi} Q(\mathbf{x}, \xi)$, it is actually a max-min function. Thus, the final reformulated objective function including the commitment cost has a min-max-min form. Then the reformulated problem can be regarded as a common two-stage robust problem which can be solved by a decomposition algorithm.

4.2.2.2 C&CG Decomposition Algorithm

As discussed above, the reformulated problem is a two-stage optimization problem that can be solved in a decomposition framework. In this study, the C&CG algorithm [188] is investigated to solve the problem which creates a master problem and subproblem. Given a unit commitment decision and base-case dispatch plan, the worst-case uncertainty is captured in the subproblem. Meanwhile, new variables and constraints are generated in the subproblem and fed back to the master problem. The algorithm iterates until all uncertainties can be guarded against.

Combining (4.1) and (4.43), we can acquire the following objective function:

$$\begin{aligned} \min \sum_t \sum_i [SU_i u_{it} + SD_i v_{it} + F_i(x_{it})] + (1 - \theta/2)\phi \\ + E_{P_0}[Q(\mathbf{x}, \xi) - \phi]^+ + \frac{\theta}{2} \max_{\xi} \min_{r_{it}^u, r_{it}^d, d_t^{ls,b}} \sum_t \left[\pi_t^{gen} \sum_i (r_{it}^u + r_{it}^d) + \pi_t^{ls} \sum_b d_t^{ls,b} \right]. \end{aligned} \quad (4.44)$$

Based on this objective, we can decompose the reformulated problem with the related constraints, and the master problem (MP) is as follows:

$$\begin{aligned} \min \sum_t \sum_i [SU_i u_{it} + SD_i v_{it} + F_i(x_{it})] + (1 - \theta/2)\phi + E_{P_0}[Q(\mathbf{x}, \xi) - \phi]^+ + \eta \\ \text{s.t. Constraints (4.2) - (4.9), (4.21) and (4.27),} \\ \text{Constraints (4.10), (4.13) - (4.20), } \forall n, \\ \text{Constraints (4.36) - (4.39), Optimality cuts,} \end{aligned}$$

where η represents the optimal value of subproblem and the optimality cuts are derived from subproblem. Note that the CVaR or $E_{P_0}[\cdot]^+$ term in the objective function can be solved as a scenario-based stochastic optimization problem [102]. Moreover, all the constraints involved in MP are linear with continuous or integer variables. Therefore, the MP is a mixed-integer linear programming (MILP) problem that can be solved by off-the-shelf solvers.

To formulate the subproblem (SP1), the second-stage problem with constraints (4.13)-(4.20) is first dualized to eliminate the inner minimization in (4.44). For the second stage problem, we need to find the worst-case scenario in the finite support set of uncertain parameter ξ . Actually, the finite support set in this work can be extended to be the common interval uncertainty set, and the worst-case scenario is usually the extreme point of this convex set. Thus, we can define the following adjustable uncertainty set for the second stage problem:

$$\mathcal{U} = \left\{ w_t(\xi) = \bar{w}_t^b + \hat{w}_t^{b+} \rho_{bt}^+ - \hat{w}_t^{b-} \rho_{bt}^-, \rho_{bt}^+ + \rho_{bt}^- \leq 1, \right. \\ \left. \left[\sum_b \sum_t (\rho_{bt}^+ + \rho_{bt}^-) \right] / \Gamma_{\max} \leq \Gamma, (\rho_{bt}^+, \rho_{bt}^-) \in \{0, 1\} \right\} \quad (4.45)$$

where \bar{w}_t^b is the forecasted mean value of wind power, \hat{w}_t^{b+} and \hat{w}_t^{b-} are the corresponding deviation from the upper bound and lower bound in the finite support set, ρ_{bt}^+ and ρ_{bt}^- are auxiliary binary variables. In addition, the normalized budget of uncertainty Γ is used here [71]. Note that Γ is set to 1 in the proposed DDRC UC model to find the worst case scenario of the second stage problem. Then the subproblem can be written as below:

$$f_{sp} = (\theta/2) \max \sum_t \left\{ \mu_t^1 \left(\sum_b d_t^b - \sum_i x_{it} - \sum_b \bar{w}_t^b \right) \right. \\ + \sum_b \hat{w}_t^{b+} \sigma_{bt}^{1+} + \sum_b \hat{w}_t^{b-} \sigma_{bt}^{1-} + \sum_i \mu_{it}^2 (-\bar{x}_i y_{it} + x_{it}) \\ + \sum_i \mu_{it}^3 (x_i y_{it} - x_{it}) + \sum_i (-\mu_{it}^4 R U_i - \mu_{it}^5 R D_i) \\ \left. + \sum_l \mu_{tl}^6 (-L_l + L_{tl}^0) + \sum_l \mu_{tl}^7 (-L_l - L_{tl}^0) + \sum_l \sum_b \hat{w}_t^{b+} \sigma_{btl}^{2+} + \sum_l \sum_b \hat{w}_t^{b-} \sigma_{btl}^{2-} \right\} \quad (4.46)$$

$$\mu_t^1 - \mu_{it}^2 - \mu_{it}^4 + \sum_l K_l^b (-\mu_{tl}^6 + \mu_{tl}^7) \leq \pi_t^{gen}, \forall t, \forall i \quad (4.47)$$

$$-\mu_t^1 - \mu_{it}^3 - \mu_{it}^5 + \sum_l K_l^b (\mu_{tl}^6 - \mu_{tl}^7) \leq \pi_t^{gen}, \forall t, \forall i \quad (4.48)$$

$$\mu_t^1 + \sum_l K_l^b (-\mu_{tl}^6 + \mu_{tl}^7) \leq \pi_t^{ls}, \forall t, \forall i \quad (4.49)$$

$$\sigma_{bt}^{1+} \leq M \rho_{bt}^+, \sigma_{bt}^{1+} \leq -\mu_t^1 + M(1 - \rho_{bt}^+) \quad (4.50)$$

$$\sigma_{bt}^{1-} \leq M \rho_{bt}^-, \sigma_{bt}^{1-} \leq \mu_t^1 + M(1 - \rho_{bt}^-) \quad (4.51)$$

$$\sigma_{btl}^{2+} \leq M\rho_{bt}^+, \sigma_{btl}^{2+} \leq (\mu_{tl}^6 - \mu_{tl}^7)K_l^b + M(1 - \rho_{bt}^+) \quad (4.52)$$

$$\sigma_{btl}^{2-} \leq M\rho_{bt}^-, \sigma_{btl}^{2-} \leq (-\mu_{tl}^6 + \mu_{tl}^7)K_l^b + M(1 - \rho_{bt}^-) \quad (4.53)$$

$$\rho_{bt}^+ + \rho_{bt}^- \leq 1, (\rho_{bt}^+, \rho_{bt}^-) \in \{0, 1\}, \forall t, \forall b \quad (4.54)$$

$$\sum_b \sum_t (\rho_{bt}^+ + \rho_{bt}^-) / \Gamma_{\max} \leq \Gamma \quad (4.55)$$

$$(\mu_{it}^2, \mu_{it}^3, \mu_{it}^4, \mu_{it}^5, \mu_{it}^6, \mu_{it}^7) \geq 0 \quad (4.56)$$

$$L_{it}^0 = \sum_b K_l^b (\sum_i x_{it}^b + \bar{w}_t^b - d_t^b) \quad (4.57)$$

where (4.46) is the dual objective function and (4.47)-(4.57) are corresponding constraints.

In the objective function, there are some bilinear terms due to the introduction of dual variables which are given below:

$$- \sum_b \hat{w}_t^{b+} \mu_t^1 \rho_{bt}^+ + \sum_b \hat{w}_t^{b-} \mu_t^1 \rho_{bt}^- \quad (4.58)$$

$$\sum_l \sum_b \hat{w}_t^{b+} \rho_{bt}^+ (\mu_{tl}^6 - \mu_{tl}^7) K_l^b - \sum_l \sum_b \hat{w}_t^{b-} \rho_{bt}^- (\mu_{tl}^6 - \mu_{tl}^7) K_l^b. \quad (4.59)$$

These bilinear terms have been linearized using big-M technique which introduces the auxiliary variables σ_{bt}^{1+} , σ_{bt}^{1-} and σ_{btl}^{2+} , σ_{btl}^{2-} along with corresponding constraints (4.50)-(4.53). Then we can solve the MP and SP1 iteratively until the algorithm terminates. Specifically, if $f_{sp} \leq \eta$, the algorithm terminates and the optimal decisions are obtained, otherwise, the following cut is fed back to the MP for next iteration:

$$\eta \geq (\theta/2) \sum_t \left[\pi_t^{gen} \sum_i (r_{it}^u + r_{it}^d) + \pi_t^{ls} \sum_b d_t^{ls,b} \right]. \quad (4.60)$$

More details about the algorithm can be found in related references [91] [95] [188].

4.2.3 Numerical Results

To validate the performance of the proposed approach, case studies based on IEEE test systems and a practical system are conducted in this subsection. First, the IEEE 6-bus test system with a wind farm is studied to illustrate the effectiveness of the proposed model and approach. Then a modified IEEE 118-bus test system is investigated for the scalability and potential practical application of the proposed approach. In addition, a practical 319-bus system with a large number of uncertainty sources is also studied to further verify the scalability of the proposed approach. Related simulation experiments are implemented in MATLAB environment with YALMIP toolbox [176] and GUROBI solver on a desktop which has an Intel Core i7-6700 CPU 3.40 GHz and 8 GB of RAM.

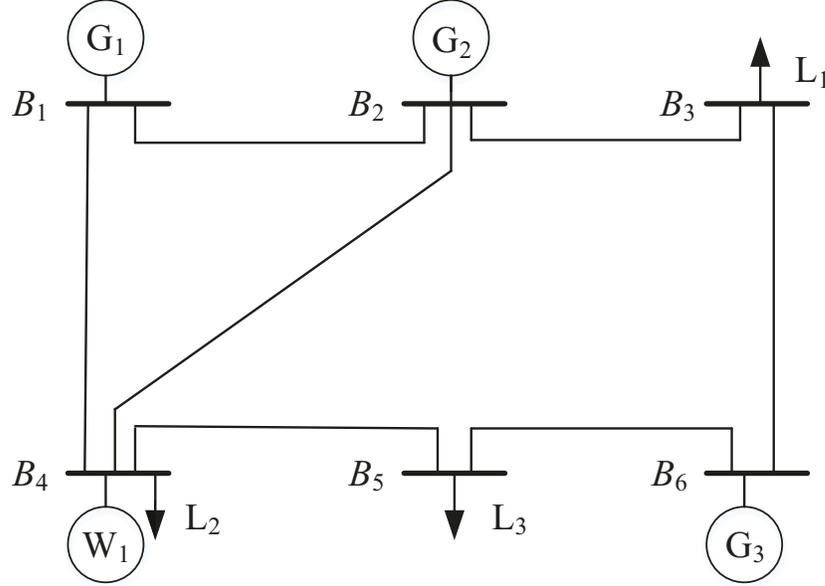


Figure 4.1: Structure of IEEE 6-bus test system.

4.2.3.1 IEEE 6-bus Test System

The studied IEEE 6-bus test system is illustrated in Fig. 4.1. Detailed parameters about this system can be found in [189]. In addition, a wind farm is connected to the system at bus 4 which is the source of uncertainty. For wind power data, we can generate the historical data from a certain assumed distribution such as the multivariate normal distribution [90]. In this work, the forecasted mean wind power is acquired from the IESO website [190] which is properly scaled and shown in Fig. 4.2, and the variance is 0.3 of the mean. Then the historical data can be generated by Monte Carlo simulation. Note that the historical data can be directly collected in practice. In addition, the number of bins is set to be 5, $K_0 = 5$ for the piecewise linear fuel cost function, the penalty re-dispatch and load shedding cost coefficient π_t^{gen} and π_t^{ls} are 50 \$/MW and 100 \$/MW, respectively, and the time horizon is 24 hours.

- **Influence of historical data size**

For the proposed data-driven UC problem, the influence of historical data size is first investigated. With different data size, different nominal distributions and thus the ambiguity sets can be generated. In this case, the confidence level in the ambiguity set is set to be $1 - \tilde{\alpha} = 95\%$ and let the data size A vary from 50 to 5000. The corresponding results including total system cost and tolerance level θ are summarized in Table 4.2. From this table, we can see that θ value decreases with the increment of historical data which means that the ambiguity set \mathcal{D}_1 shrinks and the problem becomes less conservative. Thus, the total system cost decreases with

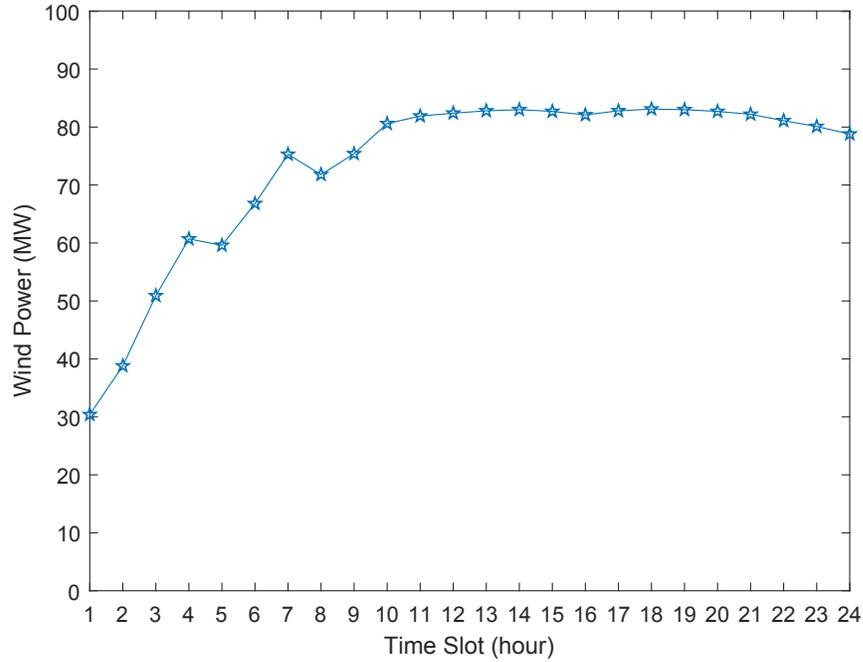


Figure 4.2: Mean wind power output.

Table 4.2: Influence of data size on system cost

Data size (A)	Total cost ($\times 10^4 \$$)	θ
50	4.9131	0.4356
100	4.8527	0.3080
500	4.7475	0.1378
1000	4.7231	0.0974
2000	4.7031	0.0689
5000	4.6850	0.0436

the rise of data size as shown in the table. This result verifies the importance of data for reducing the uncertainty, i.e., more information can be attained with more data available. Actually, θ will become 0 and the set \mathcal{D}_1 degrades to a singleton (i.e., the nominal distribution) as the data size goes to infinity. In addition, the average time for the simulation experiment is about 13.27s.

- **Influence of confidence level in ambiguity set**

To construct the ambiguity set in a data-driven manner, we also need to pay attention to the effects of confidence level ($1 - \tilde{\alpha}$) in addition to the data size. In this subsection, the data size is fixed to be 1000 and the variation of total system cost is studied with different confidence level ranging from 0.6 to 0.95. Similarly, the corresponding system cost and the θ values are reported in Table 4.3. As can be seen from this table, θ becomes larger as the confidence level increases which also results in a larger ambiguity set. A larger ambiguity set can cover the true distribution with a higher

Table 4.3: Influence of confidence level on system cost

Confidence level	Total cost ($\times 10^4$ \$)	θ
0.6	4.6943	0.0636
0.7	4.7014	0.0698
0.8	4.7080	0.0774
0.9	4.7127	0.0882
0.95	4.7231	0.0974

chance, and this reliability is achieved by increasing the total cost, i.e., the problem becomes more conservative.

- **Comparison with other methods**

One of the significant advantages of the proposed data-driven solution methodology is that the conservativeness of the problem can be adjusted depending on the amount of available data. As discussed above, the data size has an effect on the θ value which determines the ambiguity set and conservativeness. With the proposed problem reformulation method, we can easily compare the proposed distributionally robust UC problem with the common stochastic UC (SUC) problem and Robust UC (RUC) problem by adjusting the θ value. From (4.41), we can derive that the worst-case expected cost becomes the worst-case cost when θ is set to be 2 or CVaR when θ is 0 which converges to the expectation [185]. In other words, the corresponding two-stage SUC or RUC problem can be obtained by setting θ to be 0 or 2, and they are used for comparison purpose here. In addition, adjustable robust optimization is also a popular method in solving UC problems with uncertainties [71] [76]. Therefore, the adjustable robust UC (ARUC) problem is also considered as a benchmark model here. Note that the ARUC model in this work is derived from adjusting the normalized budget of uncertainty Γ in (4.45), i.e., Γ is set to 0.2, 0.5 and 0.8, respectively. Thus, the ARUC here is actually a data-driven model considering the distributional uncertainty. Taking $A=1000$ as an example, we can compare the proposed distributionally robust UC (DRUC) problem with SUC, RUC and ARUC problem, and the result is shown in Fig. 4.3. Note that the results of SUC can be regarded as optimal with known probability distribution of wind power. From this figure, it can be seen that the conservativeness of the proposed UC problem is between those of SUC and RUC, which also validates the flexibility of the data-driven method in controlling the conservativeness. In addition, the ARUC problem, with a lower cost, is less conservative than the DRUC problem since it is derived based on DRUC model, and the conservatism of ARUC model decreases as the parameter Γ becomes smaller.

To better show the benefits of the proposed approach, an out-of-sample assessment

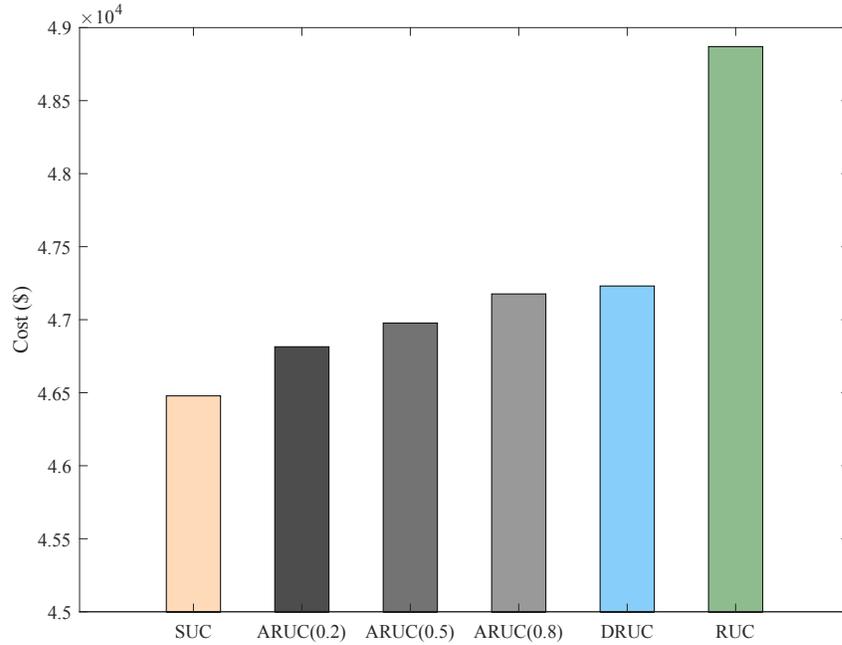


Figure 4.3: Comparison with other methods for 6-bus system.

of the commitment and base-case dispatch decisions obtained from different models mentioned above was implemented. Out-of-sample assessment is widely used to compare the performance of the solution in related references [80] [89]. More specifically, we first solved the first-stage UC problem and obtained the corresponding decisions. Then, we fixed the first-stage decisions, and the second-stage re-dispatch problem was solved with 300 randomly generated scenarios from the ambiguity set [102]. By calculating the average second-stage cost, we can compare the out-of-sample performance of different models. The boxplot result of the second-stage simulated cost is given in Fig. 4.4. As shown in this figure, the proposed DRUC model has the lowest average (median) second-stage cost which represents better performance. For the RUC model, the higher cost is caused by the great down re-dispatch and there is almost no load shedding for this case. Note that the second-stage cost is related to the cost coefficients, and the difference may become more significant by setting a larger penalty cost coefficients.

4.2.3.2 IEEE 118-bus Test System

In this subsection, a case study with the modified 118-bus test system [191] is conducted to verify the scalability and potential application of the proposed approach for large systems. This test system has 54 generation units, 186 transmission lines and 91 loads. In this case, three wind farms are connected to the system at buses 10, 30 and 50. With the same parameter settings and analysis method, we can obtain the corresponding simulation

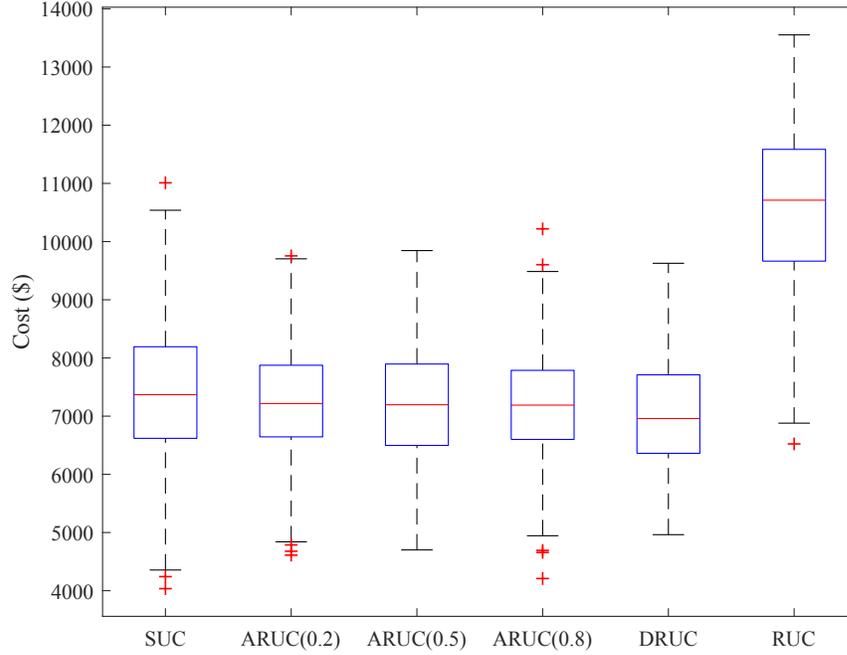


Figure 4.4: Out-of-sample assessment result for 6-bus system.

Table 4.4: Influence of data size for 118-bus system

Data size (A)	Total cost (\$)	θ
50	1150931.7	0.4356
100	1148830.2	0.3080
500	1146150.4	0.1378
1000	1145678.2	0.0974
2000	1145239.2	0.0689
5000	1144773.4	0.0436

results as given in Table 4.4, 4.5 and Fig. 4.5, respectively. From these results, similar conclusions can be attained which are omitted here. Additionally, the out-of-sample assessment for this system is also carried out, and the boxplot result is shown in Fig. 4.6. From this figure, we can also see the benefit of the proposed approach by comparing the average cost. This proves the effectiveness of the proposed approach for large-scale systems. In addition, the average time for the experiment with this system is about 1147.63s which is also acceptable in practice.

4.2.3.3 Practical-scale Power System

To further evaluate the performance of the proposed approach in a practical-scale power system, a real-world provincial 319-bus system located in Northeast China is studied in this subsection [192]. There are 65 units and 431 branches in this system, and the detailed data can be found in [88] [192]. In order to model a large number of uncertainty sources, 10

Table 4.5: Influence of confidence level for 118-bus system

Confidence level	Total cost (\$)	θ
0.6	1145084.8	0.0636
0.7	1145179.4	0.0698
0.8	1145233.9	0.0774
0.9	1145449.2	0.0882
0.95	1145678.2	0.0974

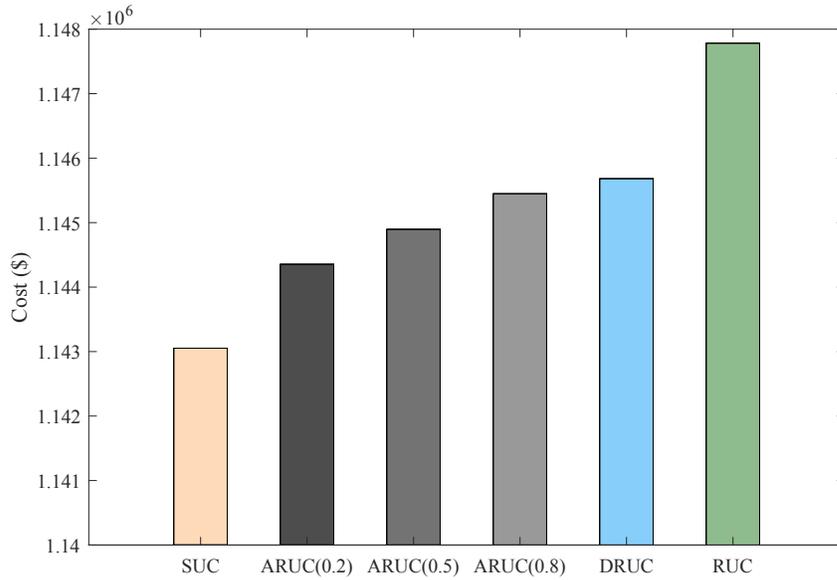


Figure 4.5: Comparison with other methods for 118-bus system.

wind farms are considered here which are assumed to share the same support set for wind power data. For simplicity, only the performance comparison of different models was conducted with this practical system, and the influence of the data size and confidence level in the ambiguity set is omitted. The corresponding experiment results are shown in Fig. 4.7 and Fig. 4.8. As can be seen from these results, the cost comparison has a similar trend with those in previous case studies. Therefore, we can acquire similar conclusions for this practical-scale power system which shows the scalability of the proposed approach. In addition, due to the scale increase of this practical system, the resulting model is also a large-scale complex model with many variables and constraints, and the average time for this experiment is about 5754.52s. Actually, this simulation time is used to obtain the day-ahead commitment and base-case dispatch plan, and for the second-stage problem or real-time re-dispatch, the average simulation time is only about 3.77s, which is quite fast in practice. Consequently, the proposed approach is applicable for large-scale practical system. In addition, it is also worth studying the proposed method with larger test systems for future research where faster algorithms may be investigated to deal with the long computation time.

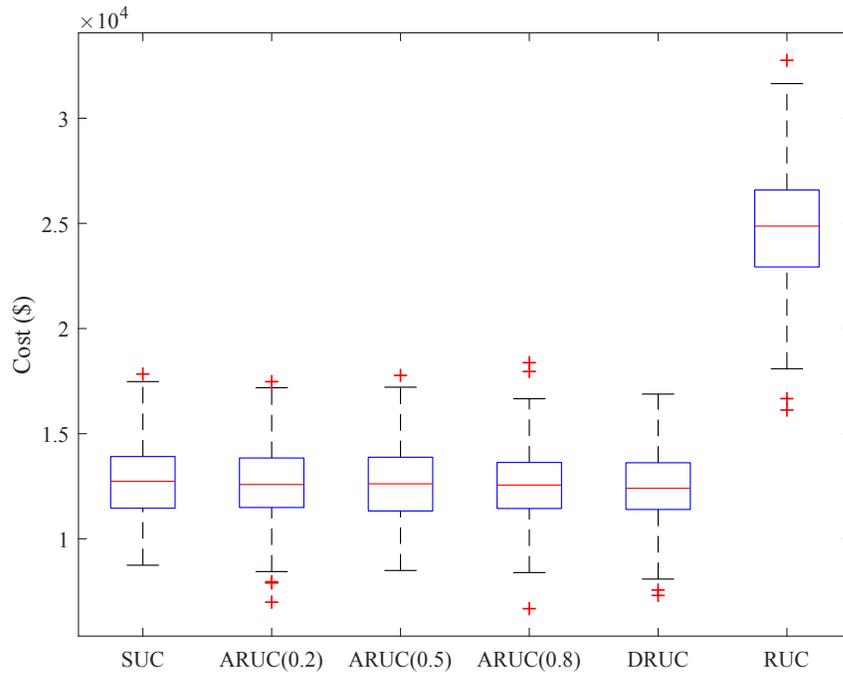


Figure 4.6: Out-of-sample assessment result for 118-bus system.

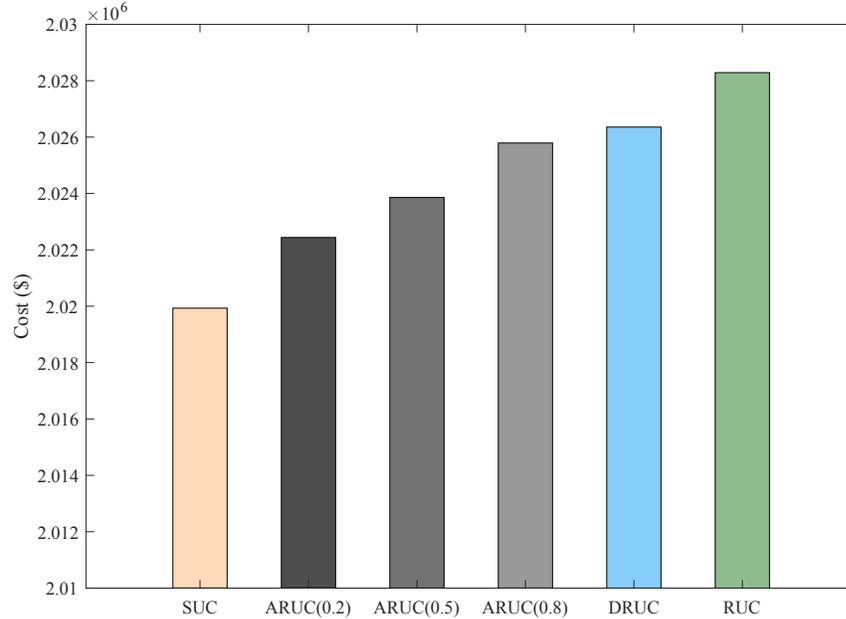


Figure 4.7: Comparison with other methods for 319-bus system.

4.3 Energy and Reserve Dispatch Using Data-driven DRO

Generally, ED problem will become the research focus when UC decisions are determined. Therefore, a similar data-driven distributionally robust (DR) energy and reserve dispatch

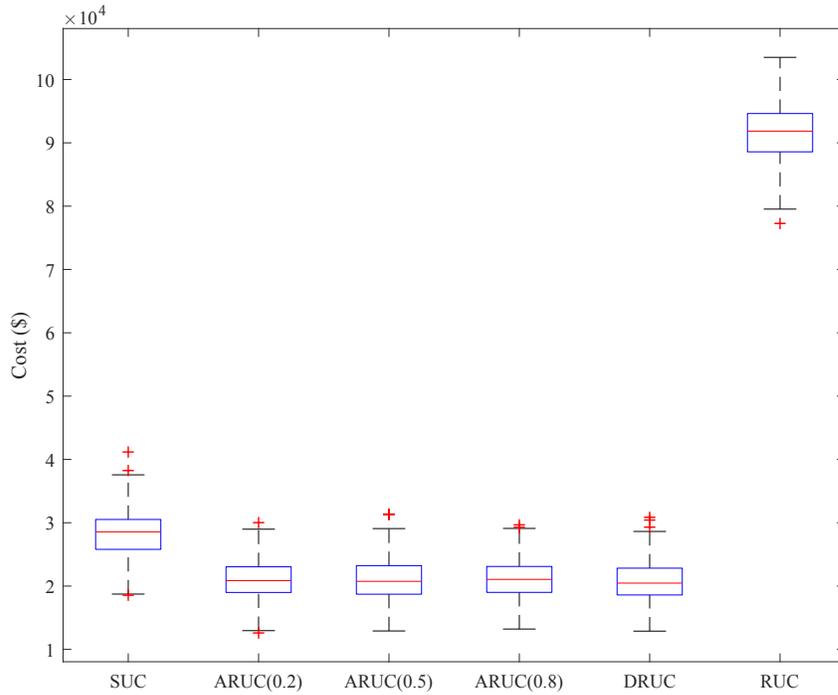


Figure 4.8: Out-of-sample assessment result for 319-bus system.

problem is studied in this section. The proposed two-stage distributionally robust model is formulated and a new different ambiguity set based on L_∞ norm is designed in subsection 4.3.1. Subsection 4.3.2 presents the detailed solution methodology based on reformulation technique, and case studies are carried out to verify the effectiveness of the proposed method in subsection 4.3.3.

4.3.1 Problem Formulation

In this subsection, the studied two-stage DR energy and reserve dispatch model with uncertain renewable generation is first formulated. Without loss of generality, wind power is considered as the renewable generation here. Then, the ambiguity set construction is introduced.

4.3.1.1 Two-stage Dispatch Model

Different from the previous stochastic or robust economic dispatch problem [193], a DR dispatch problem is studied in this work. As mentioned above, the proposed two-stage dispatch model determines the generation and reserve with forecasted wind power in the first stage and finds re-dispatch decisions against any uncertainty realization in the second stage. The first-stage decision can be seen as a base-case dispatch plan. As done in the references [95] [102], an hourly-ahead dispatch problem is also considered in this work. Mathematically, the two-stage DR energy and reserve dispatch model can be expressed as

follows:

$$\min_{\mathbf{p}^g, \mathbf{r}^u, \mathbf{r}^d} \left[\sum_i F_i(p_i^g) + d_i^u r_i^u + d_i^d r_i^d \right] + \max_{P \in \mathcal{D}} E_P[Q(\mathbf{p}^g, \mathbf{r}^u, \mathbf{r}^d, \tilde{\boldsymbol{\xi}})] \quad (4.61)$$

$$\text{s.t.} \quad \sum_i^{N_g} p_i^g + \sum_j^{N_w} w_j^f = \sum_k^{N_L} p_k^L \quad (4.62)$$

$$p_i^{\min} \leq p_i^g - r_i^d, p_i^g + r_i^u \leq p_i^{\max}, \forall i \quad (4.63)$$

$$-L_l \leq \sum_i^{N_g} \pi_{il}^g p_i^g + \sum_j^{N_w} \pi_{jl}^w w_j^f - \sum_k^{N_L} \pi_{kl}^L p_k^L \leq L_l, \forall l \quad (4.64)$$

$$0 \leq r_i^d \leq RD_i, 0 \leq r_i^u \leq RU_i, \forall i \quad (4.65)$$

$$F_i(p_i^g) \geq a_i^m p_i^g + b_i^m, \forall i, \forall m \quad (4.66)$$

where the objective function in (4.61) includes the fuel cost, spinning reserve cost and second-stage worst-case expected cost. $Q(\cdot)$ represents the second-stage cost and \mathcal{D} is the ambiguity set which will be introduced later. Constraint (4.62) is the power balance equation. Constraint (4.63) limits the generation output. Constraint (4.64) ensures the DC power flow limit. Constraint (4.65) limits the up and down reserve capacities. Constraint (4.66) is the piecewise linear approximation with m pieces of the common quadratic fuel cost function.

In the second stage, when the forecast errors $\tilde{\boldsymbol{\xi}}$ of wind generation are revealed, we can adopt corrective actions by re-dispatching. In this case, the real wind power and generation output can be expressed as below:

$$\tilde{w}_j = w_j^f + \tilde{\xi}_j, \forall j \quad (4.67)$$

$$\tilde{p}_i^g = p_i^g + p_i^{gu} - p_i^{gd}, \forall i \quad (4.68)$$

Furthermore, the second-stage problem can be stated as follows:

$$Q(\mathbf{p}^g, \mathbf{r}^u, \mathbf{r}^d, \tilde{\boldsymbol{\xi}}) = \min \sum_i^{N_g} (c_i^u p_i^{gu} + c_i^d p_i^{gd}) + \sum_j^{N_w} c_j^w w_j^c + \sum_k^{N_L} c_k^{Lc} p_k^{Lc} \quad (4.69)$$

$$\sum_i^{N_g} \tilde{p}_i^g + \sum_j^{N_w} (\tilde{w}_j - w_j^c) = \sum_k^{N_L} (p_k^L - p_k^{Lc}) \quad (4.70)$$

$$-L_l \leq \sum_i^{N_g} \pi_{il}^g \tilde{p}_i^g + \sum_j^{N_w} \pi_{jl}^w (\tilde{w}_j - w_j^c) - \sum_k^{N_L} \pi_{kl}^L (p_k^L - p_k^{Lc}) \leq L_l, \forall l \quad (4.71)$$

$$0 \leq p_i^{gu} \leq r_i^u, 0 \leq p_i^{gd} \leq r_i^d, \forall i \quad (4.72)$$

$$0 \leq p_k^{Lc} \leq p_k^L, 0 \leq w_j^c \leq w_j^f, \forall k, \forall j \quad (4.73)$$

where the second-stage cost in (4.69) includes the re-dispatch cost of generators, wind power curtailment cost and load shedding cost. Constraint (4.70) limits the power

balance in the second stage problem. Constraint (4.71) is the transmission line flow limit. Constraint(4.72) ensures that the generator re-dispatch capability is limited by the first-stage reserve capacity. Constraint (4.73) restricts the load shedding and wind power curtailment.

In this study, the uncertainty source is the wind power generation, which is demonstrated by $\tilde{\xi}$ in above formulations. In addition, it is assumed that the uncertain wind power has a finite support [90], i.e., the number of possible realizations $\tilde{\xi}^s$ is finite (e.g., $\tilde{\xi}^1, \tilde{\xi}^2, \dots, \tilde{\xi}^S$). However, the true probability distribution is unknown and is restricted by the constraints in the ambiguity set \mathcal{D} .

4.3.1.2 Ambiguity Set Construction

As mentioned above, the true probability distribution of wind power is unknown. In this study, a distance-based ambiguity set is designed to constrain the true distribution, i.e., a distance measure between two distributions is used to describe the ambiguity set. Specifically, the general ambiguity set can be defined as: $\mathcal{D} = \{P \in \mathcal{P} : dist(P||P_0) \leq \theta\}$, where \mathcal{P} is the set of all distributions, $dist(\cdot)$ is a distance measure between the true distribution P and a reference distribution P_0 , and θ is the tolerance level. In this work, the L_∞ norm is adopted as the distance function which has not been studied for energy and reserve dispatch problem before. Accordingly, the studied ambiguity set is given as follows:

$$\mathcal{D}_\infty = \{P \in \mathcal{P} : \|P - P_0\|_\infty \leq \theta\}. \quad (4.74)$$

The advantage of using the L_∞ norm is that the convergence between the true distribution and reference distribution can be guaranteed as the data size goes to infinity.

In above ambiguity set, the reference distribution can be derived with the histogram method from historical data [90] [91]. For example, it is assumed that there are N data samples in total, and the sample space can be partitioned into S bins. Then, we can count the frequency of samples in each bin (e.g., N^s) and the reference distribution can be obtained as $P_0 = (f_0^1, f_0^2, \dots, f_0^S)$ with the element $f^s = N^s/N$. In addition, the tolerance level θ can be adjusted to control the conservatism. According to the Proposition 9 in [185], θ can be determined as follows:

$$\theta = (z_{\alpha/2}/\sqrt{N}) \max_{s=1, \dots, S} \sqrt{f_0^s(1 - f_0^s)} \quad (4.75)$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)$ th percentile of standard normal distribution. In this case, the ambiguity set contains the true distribution with a $(1 - \alpha)$ confidence level.

4.3.2 Solution Methodology

For better description, the two-stage DR dispatch model is first expressed in a compact form as follows:

$$\min \mathbf{c}^\top \mathbf{x} + \max_{P \in \mathcal{D}_\infty} E_P[Q(\mathbf{x}, \tilde{\boldsymbol{\xi}})] \quad (4.76)$$

$$\text{s.t. } Q(\mathbf{x}, \tilde{\boldsymbol{\xi}}) = \min \mathbf{q}^\top \mathbf{y} \quad (4.77)$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{h} \quad (4.78)$$

$$\mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{y} + \mathbf{D}\tilde{\boldsymbol{\xi}} \leq \mathbf{d} \quad (4.79)$$

$$\mathbf{E}\mathbf{x} + \mathbf{F}\mathbf{y} \leq \mathbf{e} \quad (4.80)$$

$$\mathbf{x} \geq 0, \mathbf{y} \geq 0 \quad (4.81)$$

where \mathbf{x} represents the first-stage decision variables including p_i^g, r_i^u and r_i^d , \mathbf{y} is the second-stage variables including $p_i^{gu}, p_i^{gd}, w_j^c$ and p_k^{Lc} . In addition, constraint (4.78) represents the first-stage constraints, (4.79) includes (4.70)-(4.71), and (4.80) includes (4.72)-(4.73).

To solve the above complex two-stage DR dispatch problem, a general idea is to reformulate the worst-case second-stage expected cost, and this is also the basic idea of our solution method. For the ambiguity set introduced in (4.74), we can write it equivalently as $f_0^s - \theta \leq f^s \leq f_0^s + \theta$, for all $s = 1, 2, \dots, S$. Furthermore, let $f_l^s = f_0^s - \theta$ and $f_u^s = f_0^s + \theta$ by introducing two auxiliary variables f_l^s and f_u^s , we can get the following equivalent L_∞ norm based ambiguity set:

$$\mathcal{D}_\infty = \{f \geq 0 : f_l^s \leq f^s \leq f_u^s, \forall s, \sum_{s=1}^S f^s = 1\} \quad (4.82)$$

where f_l and f_u are the lower and upper bound of the unknown probability distribution, respectively. In addition, it is assumed that $P_l := \sum_{s=1}^S f_l^s \in (0, 1)$ and $P_u := \sum_{s=1}^S f_u^s > 1$ to avert trivial cases.

Based on the above ambiguity set, we can induce two new probability measures from f_l and $f_u - f_l$, respectively, which are given below:

$$\mathbb{P}_l = \sum_{s \in \{1, \dots, S\}} \frac{f_l^s}{P_l} \quad (4.83)$$

$$\mathbb{P}_{u-l} = \sum_{s \in \{1, \dots, S\}} \frac{f_u^s - f_l^s}{P_u - P_l}. \quad (4.84)$$

With these two probability measures, we can reformulate the worst-case second-stage expectation in (4.76) to a convex combination of an expectation and the CVaR [185] as

follows:

$$\max_{P \in \mathcal{D}_\infty} E_P[Q(\mathbf{x}, \tilde{\boldsymbol{\xi}})] = P_l E_{\mathbb{P}_l}[Q(\mathbf{x}, \tilde{\boldsymbol{\xi}})] + (1 - P_l) \text{CVaR}_{(P_u - P_l)/(P_u - P_l)}^{\mathbb{P}_{u-l}}[Q(\mathbf{x}, \tilde{\boldsymbol{\xi}})]. \quad (4.85)$$

Considering the definition of CVaR [169], the right-hand second term of (4.85) can be further written as below:

$$(1 - P_l) \text{CVaR} = \min_{\phi \in \mathbb{R}} (1 - P_l)\phi + (P_u - P_l) E_{\mathbb{P}_{u-l}}[Q(\mathbf{x}, \tilde{\boldsymbol{\xi}}) - \phi]^+ \quad (4.86)$$

where $[\tau]^+ = \max\{\tau, 0\}$.

Combining (4.85) and (4.76), we can find that the original two-stage DR problem is transformed into a two-stage stochastic linear program problem. In addition, since there is only min operator in this two-stage problem, the problem can be readily solved by available solvers without decomposition. More specifically, we need to solve a scenario-based stochastic problem for the second-stage cost and CVaR term in (4.86) [102], which can be described as follows:

$$\min_{\phi, \beta_s, \mathbf{y}_s} (1 - P_l)\phi + (P_u - P_l) E_{\mathbb{P}_{u-l}}(\beta_s) \quad (4.87)$$

$$\beta_s \geq 0, \beta_s \geq \mathbf{q}^\top \mathbf{y}_s - \phi, \forall s \quad (4.88)$$

$$\mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{y}_s + \mathbf{D}\tilde{\boldsymbol{\xi}}_s \leq \mathbf{d}, \forall s \quad (4.89)$$

$$\mathbf{E}\mathbf{x} + \mathbf{F}\mathbf{y}_s \leq \mathbf{e}, \mathbf{y}_s \geq 0, \forall s \quad (4.90)$$

where β_s is the auxiliary variable for scenario s . Different from the common two-stage DR or robust problem solved by decomposition method such as Benders decomposition and column and constraint generation method, the proposed two-stage model is reformulated to a two-stage stochastic linear program problem which can be solved directly based on scenarios.

4.3.3 Case Studies

In this subsection, case studies are carried out based on IEEE 6-bus test system and 118-bus test system to verify the effectiveness of the proposed method. The problem is programmed in Matlab with YALMIP toolbox and solved by GUROBI solver. All the simulation experiments are implemented on a Windows-based PC with an Intel Core i7-6700 CPU 3.40 GHz and 8 GB of RAM.

4.3.3.1 IEEE 6-bus System

The structure of IEEE 6-bus system is illustrated before, and there are 3 generators connected to buses 1, 2 and 6, seven transmission lines and three loads in this test system. Detailed data about this system can be found in [189]. The generation units are assumed

Table 4.6: Effects of data size for 6-bus system

Data size	Cost (\$)	θ	SED cost (\$)
100	7969.51	0.0821	7744.02
500	7837.29	0.0367	7744.02
1000	7807.93	0.0258	7744.02
2000	7777.29	0.0183	7744.02
5000	7768.88	0.0116	7744.02

to be on for dispatch problem. For the cost coefficients of each unit, d_i^u and d_i^d in the first stage are assumed to be 10% of the first-order coefficients of the quadratic fuel cost function, c_i^u and c_i^d are equal to the per-unit production cost at maximum output [81]. The cost coefficients of wind power curtailment and load shedding are 100 \$/MW and 200 \$/MW, respectively. Three loads at buses 3, 4 and 5 have a demand of 100MW, 100MW and 150MW, respectively.

In addition, it is assumed that there are two wind farms connected to buses 4 and 6, and each has a forecasted output of 50 MW [95]. To generate historical data of wind power, normal distribution is used where the mean equals to the forecasted value and the variance is 10% of the mean. Note that historical data can be collected directly in practice. In addition, the number of bins is set to be 5 [90]. Then we can estimate the reference distribution with the histogram method. For the piecewise linear function, the number of pieces is 3 in this study.

- **Effects of historical data size**

Based on the above parameter settings, the value of data in influencing the conservatism of the problem is first validated. The parameter α in (4.75) is set to be 0.05 and the data size ranges from 100 to 5000. The numerical results of distributionally robust ED (DRED) are summarized in Table 4.6. In addition, the cost result with 50000 data size is also given as a benchmark which can be seen as the stochastic economic dispatch (SED) problem with perfect information, i.e., this case is used to estimate the problem with true distribution. From this table, we can see that the total average cost decreases as the data size increases, thus the problem becomes less conservative. Considering the change of θ value, it can also be derived that the ambiguity set shrinks with more data information involved, and the DRED problem is expected to converge to the SED problem with perfect information when the data size goes to infinity. In summary, the results show the value of additional data in controlling the conservatism of the problem.

- **Influence of confidence level**

In this subsection, the influence of the confidence level in ambiguity set on the conservatism of the problems is tested. From (4.75), we can see that the confidence

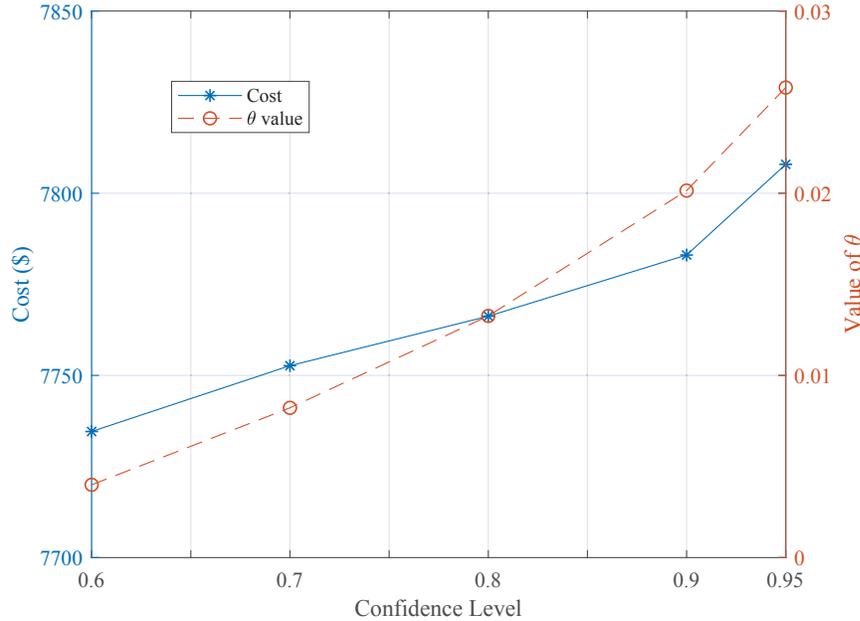


Figure 4.9: Influence of confidence level for 6-bus system.

level will influence the θ value which indicates the size of the ambiguity set. In this case study, the number of data is fixed as 1000 and the value of confidence level ($1 - \alpha$) is adjusted from 0.6 to 0.95. The results of corresponding cost and θ value are illustrated in Fig. 4.9. As can be seen from this figure, both the total cost and the θ value increase with the increment of confidence level. This is reasonable since higher confidence level means a higher chance that the ambiguity set contains the true distribution. Therefore, θ should become larger to enlarge the ambiguity set when confidence level increases, which also results in a more conservative problem.

4.3.3.2 IEEE 118-bus System

In this subsection, the performance of the proposed method is further tested with a larger system, a modified IEEE 118-bus test system, which can also be used to verify the potential application of the approach in practice. The detailed data about this system can be found in [191] and the total load demand is 4720 MW. Related parameters such as the cost coefficients are the same with those in 6-bus system case study. In addition, it is assumed that six wind farms are connected to the system at buses 12, 17, 49, 59, 80 and 92, and each has a 100 MW forecasted output [95]. The wind farms are assumed to have the same support set for simplicity, and the historical data are also generated from normal distribution.

With a similar simulation experiment, the effectiveness of the proposed method is also validated by checking the influence of historical data size and confidence level. The

Table 4.7: Effects of data size for 118-bus system

Data size	Cost (\$)	θ	SED cost (\$)
100	105278.57	0.0821	102405.63
500	103803.66	0.0367	102405.63
1000	103397.21	0.0258	102405.63
2000	103003.38	0.0183	102405.63
5000	102771.85	0.0116	102405.63

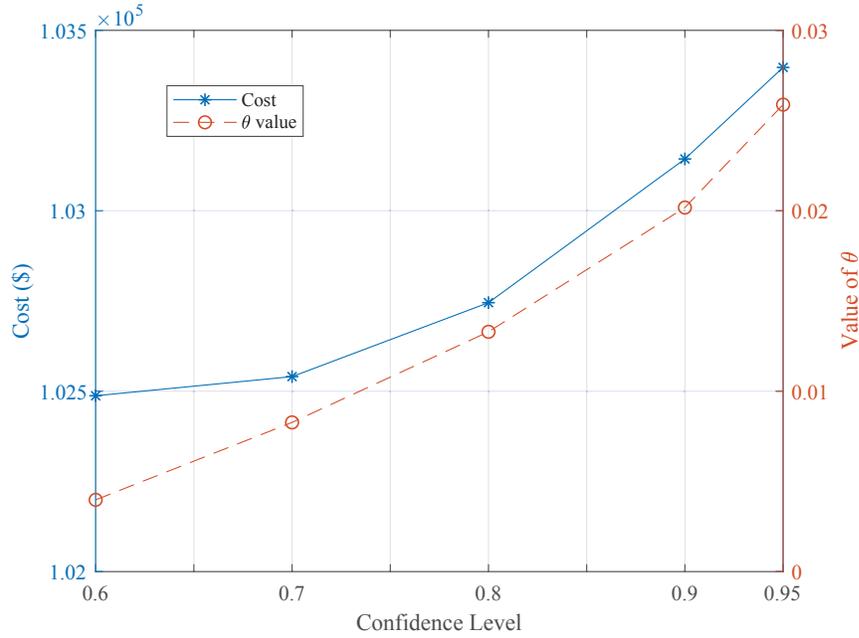


Figure 4.10: Influence of confidence level for 118-bus system.

simulation results are given in Table 4.7 and Fig. 4.10, respectively. From analyzing these results, we can derive the same conclusions as before which are omitted. The computation time for this case study is only several seconds, and this also confirms the potential application of the proposed method on a practical power system.

In addition, the proposed method is tested with more data sets, i.e., Weibull distribution and log-normal distribution are also used to generate historical data except for the above normal distribution. For these various distributions, we study the influence of the data size similarly and compare the obtained average total costs with the cost of SED with perfect information. More specifically, the gap between the DRED cost (e.g., z_{dr}) and optimal SED cost (z^*) is calculated which is defined as $(z_{dr} - z^*)/z^*$. The results are illustrated in Fig. 4.11. As can be seen from this figure, the cost of DRED problem converges to that of perfect SED problem with the increase of historical data size under different distributional settings. Consequently, the conservatism of the DRED problem decreases with more data information.

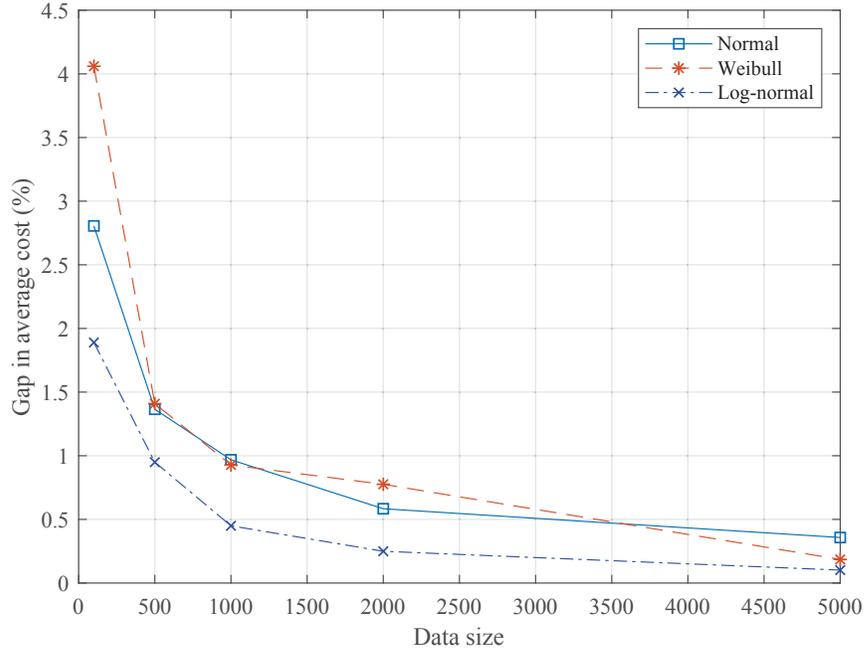


Figure 4.11: Influence of data size under different distributions.

4.4 Summary

Data-driven distributionally robust chance-constrained UC and energy and reserve dispatch problem are investigated in this chapter.

First, a two-stage data-driven distributionally robust chance-constrained unit commitment problem is studied which determines the commitment decision and basic dispatch plan in the first stage, and considers the worstcase expected power imbalance or re-dispatch cost in the second stage. The chance constraint is used to restrain the possible energy imbalance. Different from the moment-based ambiguity set, a distance-based ambiguity set is constructed to capture the uncertainty of wind power distribution, and this set can be derived in a data-driven environment. Numerical results show that the system cost decreases with more available historical data, for example, the cost decreases from $\$4.9131 \times 10^4$ with data size 50 to $\$4.6850 \times 10^4$ with data size 5000 for IEEE 6-bus test system, and that the conservativeness of the problem can be controlled by tuning the data size and confidence level in the ambiguity set. In addition, the effectiveness and flexibility of the proposed data-driven approach is also verified by the comparison with SUC, ARUC and RUC problems, for example, the total cost of the proposed DRUC problem with data size 1000 is $\$4.7231 \times 10^4$ for IEEE 6-bus test system which is between the cost of SUC problem ($\$4.6478 \times 10^4$) and RUC problem ($\4.8870×10^4).

In addition, a two-stage data-driven distributionally robust energy and reserve

dispatch problem is also studied, and the two-stage model considers the base-case dispatch plan and reserve in the first stage and worst-case re-dispatch in the second stage. Unlike the common moment-based ambiguity set, a new distance-based ambiguity set, i.e., L_∞ norm based set, is designed which has not been studied for such problem before. Based on this new ambiguity set, the second-stage worst-case expected cost is reformulated into a combination of CVaR and an expectation with respect to a reference distribution, which makes the proposed two-stage distributionally robust model become a two-stage stochastic optimization problem. The results of simulation experiments validate the effectiveness of the proposed approach, especially the value of data in controlling the conservatism of the problem. Moreover, it is also shown that the distributionally robust problem converges to the stochastic problem with perfect information with the historical data size goes to infinity.

5

Multistage Energy Management for Microgrids

Acronyms

CCHP	Combined Cooling, Heating and Power
COP	Coefficient of Performance
C&CG	Column-and-constraint Generation
DG	Distributed Generation
ESS	Energy Storage System
DRO	Distributionally Robust Optimization
LDR	Linear Decision Rule
MISOCP	Mixed-integer Second-order Conic Programming
MT	Micro Turbine
RO	Robust Optimization
RDDP	Robust Dual Dynamic Programming
SDDP	Stochastic Dual Dynamic Programming
TSS	Thermal Storage System
UC	Unit Commitment

5.1 Introduction

Microgrids play a significant role in future smart grid, and they can operate in either islanded mode or grid-connected mode. Compared with the islanded microgrids, grid-connected microgrids can transact electricity with the main grid. In addition, almost all of the energy management models for microgrids in existing literature belong to single-stage or two-stage models which ignore the non-anticipativity of uncertainty. Therefore, multistage energy management for grid-connected microgrids with uncertainty is investigated to overcome this problem in this chapter, and RO and DRO uncertainty modeling techniques are studied, respectively.

A new multistage robust energy management problem is proposed first for grid-connected microgrids which considers the uncertainty from renewable generation and load demand. Compared with the common single-stage or two-stage model, multistage models with uncertainty are more complicated and computationally difficult. To address the computational difficulty, a novel decomposition method similar to SDDP method, i.e., the robust dual dynamic programming method (RDDP), is proposed to solve the multistage model. This method decomposes the multistage problem into small-stage problems and tries to approximate the unknown cost-to-go function with a lower and upper bound. The effectiveness of the proposed approach is validated by a case study with real datasets.

Among different microgrids, one interesting kind is the combined cooling, heating and power (CCHP) based microgrid which can supply electric and thermal power simultaneously. As a leading power generation method with high efficiency and good environmental benefits, CCHP microgrids have also attracted wide attention. However, distributionally robust energy management for CCHP microgrids has seldom been reported. In this chapter, a multi-period energy management model for CCHP microgrids is studied with DRO technique, and the non-anticipativity of uncertainty is considered. The proposed multi-period model is included in a two-stage framework, and here multi-period instead of multi-stage is used to make a difference. In addition, a new second-order conic representable ambiguity is designed with DRO method, and linear decision rule (LDR) method is investigated to help solve the distributionally robust multi-period energy management problem.

5.2 Multistage Robust Energy Management for Microgrids

This section intends to study the multistage robust energy management problem for grid-connected microgrids. In subsection 5.2.1, the multistage robust energy management model is formulated including the system model and a description of multistage robust optimization problem, followed by the solution methodology based on RDDP method in subsection 5.2.2. The case study and corresponding simulation results are provided in subsection 5.2.3.

5.2.1 Problem Formulation

Generally, a microgrid has two operation modes, i.e., grid-connected mode and islanded mode. In grid-connected mode, the microgrid can exchange power with the main grid to maintain the power balance. In case of main grid fault, it can switch to the islanded mode. Multistage energy management of a grid-connected microgrid with uncertainty is studied

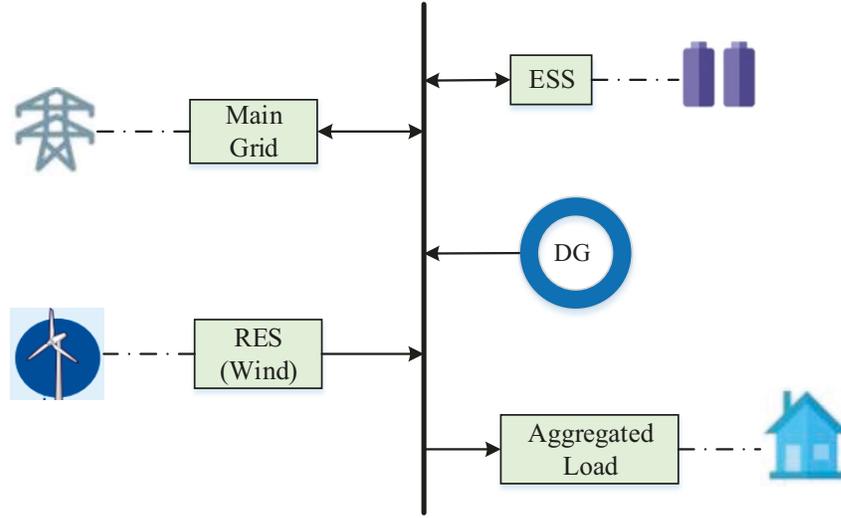


Figure 5.1: Microgrid system model.

in this work, which can also be extended to an islanded microgrid by setting zero power exchange. To formulate the problem, the microgrid system model and uncertainty set considered are first introduced in this subsection, then the multistage robust optimization problem is presented.

5.2.1.1 System Model

In general, a microgrid is composed of a number of conventional generators, renewable generators (or RESs), ESSs and a group of loads. The schematic diagram of a typical grid-connected microgrid is shown in Fig. 5.1 [194]. In this model, all units are connected to one bus, only the nodal power balance is considered and there are no transmission line flows. In addition, it is assumed that the considered distribution lines have sufficiently large capacity and line loss is neglected [60] [61] [194]. The objective of microgrid energy management is to minimize the total system cost by determining the output of different DERs, the charging or discharging schedule of ESSs and the amount of electricity purchased from or sold to the main grid.

There are different kinds of conventional generators in a microgrid such as micro-turbines, DG units, and fuel cells. In this work, DG units are considered as the conventional generators, and the fuel cost is formulated as a linear function at each time period t , which can also be approximated by a piece-wise linear function [60]:

$$C_{it}^{dg} = (a_i^{dg} P_{it} + b_i^{dg}) \Delta t, \forall i, t \quad (5.1)$$

where P_{it} is the output of i th DG unit at time t and Δt is the duration of a time slot (e.g., 1h). At each time period, the output of DG units should satisfy the lower and upper limits

and corresponding ramp-up/ramp-down constraints as follows:

$$\underline{P}_i \leq P_{it} \leq \overline{P}_i, \forall i, t \quad (5.2)$$

$$P_{it} - P_{i(t-1)} \leq R_i^{up}, \forall i, t \quad (5.3)$$

$$P_{i(t-1)} - P_{it} \leq R_i^{dn}, \forall i, t. \quad (5.4)$$

The day-ahead schedule of DG units is assumed to be finished and the DG units have on status in this work. Therefore, the startup and shutdown costs are also neglected here.

To alleviate the negative impact of renewable generation uncertainty, distributed ESSs such as batteries and flywheels can be utilized in microgrid. The battery storage system is selected in this work. Let E_{jt} denote the stored energy of ESS unit j at the end of time t , then the energy balance of ESS in Δt time slot can be expressed as follows [195]:

$$E_{jt} = E_{j(t-1)} + \eta_j^{e+} P_{jt}^{ch} \Delta t - P_{jt}^{dch} \Delta t / \eta_j^{e-}, \forall j, t \quad (5.5)$$

where η_j^{e+} and η_j^{e-} represent the charging and discharging rates, respectively, P_{jt}^{ch} and P_{jt}^{dch} are charging and discharging power at time period t for unit j , respectively. In addition, the charging and discharging power should satisfy the following constraints:

$$0 \leq P_{jt}^{ch} \leq \overline{P}_j^{ch}, 0 \leq P_{jt}^{dch} \leq \overline{P}_j^{dch}, \forall j, t \quad (5.6)$$

where \overline{P}_j^{ch} and \overline{P}_j^{dch} are maximum charging and discharging limits, respectively. Note that the complementary constraint to avoid simultaneous charging and discharging is not required here as the charging/discharging efficiency is considered [158]. Since each ESS unit has a capacity limit, the following constraint is used to restrict its stored energy:

$$\underline{E}_j \leq E_{jt} \leq \overline{E}_j, \forall j, t \quad (5.7)$$

where \underline{E}_j and \overline{E}_j are minimum and maximum stored energy for unit j . In addition, for periodic use of ESS, the final stored energy level at time T should be equal to the known initial capacity value as follows:

$$E_{jT} = E_{j0}. \quad (5.8)$$

Considering the influence of frequent charging and discharging on storage lifetime, a linear function model is used to describe the ESS degradation or maintenance cost [61]:

$$C_{jt}^{ess} = c_j^{ess} (\eta_j^{e+} P_{jt}^{ch} + P_{jt}^{dch} / \eta_j^{e-}) \Delta t, \forall j, t \quad (5.9)$$

where c_j^{ess} denotes the degradation cost coefficient of unit j .

For a grid-connected microgrid, it can trade energy with the external market, i.e., purchase electricity from the main grid or sell the excess power to the main grid. Let

c_t^{buy} and c_t^{sell} denote the purchase price and sale price at time period t in electricity market, respectively, then the power exchange cost with the main grid can be expressed as follows:

$$C_t^{grid} = (c_t^{buy} P_t^{buy} - c_t^{sell} P_t^{sell}) \Delta t \quad (5.10)$$

where P_t^{buy} and P_t^{sell} represent the amount of power purchased from or sold to the external market, respectively. It is usually considered that $c_t^{buy} > c_t^{sell}$ which is reasonable to avert trading arbitrage. Otherwise we can gain profits from buying and selling power in the market [196]. In addition, the trading power should satisfy the following constraints:

$$0 \leq P_t^{buy} \leq \bar{P}_t^{buy}, 0 \leq P_t^{sell} \leq \bar{P}_t^{sell} \quad (5.11)$$

where \bar{P}_t^{buy} and \bar{P}_t^{sell} are the maximum limits of power purchase and sale, respectively.

Renewable energy generation is an important part of microgrids. Without loss of generality, wind power is considered as the renewable source in this work. Although there are different kinds of demands in a microgrid, aggregated load is considered for simplicity. Taking all the power supply and demand into account, we have the following power balance constraint at each time period:

$$\sum_i P_{it} + \sum_j (P_{jt}^{dch} - P_{jt}^{ch}) + P_t^{buy} - P_t^{sell} + w_t + P_t^{loss} = P_t^{load} \quad (5.12)$$

where w_t and P_t^{load} represent the aggregated wind power output and load demand, respectively, P_t^{loss} is an auxiliary variable to balance the load shedding. Wind power and load demand are sources of uncertainty in this work, which can be expressed by a new random vector $\xi_t = (w_t, P_t^{load})$ for convenience.

In robust optimization problems, the true distribution of random variable is usually unknown and uncertainty set is used to describe the uncertainty. In this work, the polyhedral interval uncertainty set is used for uncertain wind power and load demand as given below:

$$U_\xi = [\xi_t : \underline{\xi}_t \leq \xi_t \leq \bar{\xi}_t] \quad (5.13)$$

where $\underline{\xi}_t$ and $\bar{\xi}_t$ are the lower and upper bounds of the variable. The uncertainty set introduced above is bounded and stage-wise rectangularity is assumed. To obtain such uncertainty sets, several methods can be used as shown in the literature. For example, we can use the popular interval prediction method to estimate the uncertainty set [31]. In addition, the uncertainty set can also be attained by point forecast methods which construct the interval by a point forecast value and corresponding confidence level [60]. Note that the random variable is assumed to be deterministic in the first stage of a multistage problem.

5.2.1.2 Multistage Robust Optimization Problem

Based on the microgrid equipment models introduced above, we can define the decision variable x_t and the objective function of a single stage problem as follows:

$$c_t^\top x_t = \sum_i C_{it}^{dg} \Delta t + \sum_i C_{jt}^{ess} + C_t^{grid} + c^{loss} P_t^{loss} \quad (5.14)$$

$$x_t = [P_{it}, P_{jt}^{ch}, P_{jt}^{dch}, E_{jt}, P_t^{buy}, P_t^{sell}, P_t^{loss}] \quad (5.15)$$

where c^{loss} is the corresponding penalty cost coefficient. Note that $x_t = x_t(\xi_t)$ where ξ_t is omitted for notational simplicity. For a multistage model, the decisions are made sequentially with the gradual disclosure of uncertainty.

Combining the previous constraints and objective functions, we can write the multistage robust problem as below:

$$\begin{aligned} \min_{x_1} c_1^\top x_1 + [\max_{\xi_2} \min_{x_2} c_2^\top x_2 + [\dots + \max_{\xi_T} \min_{x_T} c_T^\top x_T] \dots] \\ \text{s.t. (5.2) - (5.8), (5.11) - (5.12).} \end{aligned} \quad (5.16)$$

By transforming the constraints into a matrix form at each stage, we can also write the multistage problem in a compact form:

$$\min_{\substack{A_1 x_1 \geq b_1 \\ x_1 \geq 0}} c_1^\top x_1 + [\max_{\xi_2} \min_{\substack{B_2 x_1 + A_2 x_2 \geq b_2 \\ x_2 \geq 0}} c_2^\top x_2 + [\dots + \max_{\xi_T} \min_{\substack{B_T x_{T-1} + A_T x_T \geq b_T \\ x_T \geq 0}} c_T^\top x_T] \dots] \quad (5.17)$$

where A_t and B_t are corresponding coefficient matrices derived from the constraints, and ξ_t is omitted in $b_t(\xi_t)$ which means that b_t is influenced by the random parameter ξ_t . ξ_1 is deterministic in the first stage. The problem (5.17) is also called the nested problem formulation, which is very difficult or intractable to solve for even a small number of stages.

To solve this problem, we can decompose it into stage problems and write it in a dynamic programming form as follows:

$$\begin{aligned} \min c_1^\top x_1 + Q_2(x_1) \\ \text{s.t. } A_1 x_1 \geq b_1 \\ x_1 \geq 0 \end{aligned} \quad (5.18)$$

where Q_2 represents the worst-case future cost-to-go function in the first stage problem and the stage- t worst-case cost-to-go function Q_t is defined as $Q_t(x_{t-1}) = \max_{\xi_t} \{S_t(x_{t-1}; \xi_t)\}$ with the inner problem $S_t(x_{t-1}; \xi_t)$ defined as follows:

$$\begin{aligned} S_t(x_{t-1}; \xi_t) = \min c_t^\top x_t + Q_{t+1}(x_t) \\ \text{s.t. } B_t x_{t-1} + A_t x_t \geq b_t \\ x_t \geq 0, t = 2, \dots, T \end{aligned} \quad (5.19)$$

and at the last stage, we have $Q_{T+1}(x_T) = 0$ or it can be some known convex polyhedral function [197]. Thus, the nested multistage robust problem can be replaced by the dynamic programming equations, i.e., the stage-1 problem and stage- t problem as shown in (5.18) and (5.19). The solution methodology is also developed based on the dynamic programming form.

5.2.2 Solution Methodology

In multistage stochastic programming problem, we need to identify the expected future cost function to solve the problem and the SDDP method is thus developed with an outer approximation method for the expected future cost. By contrast, it is the worst-case future cost that needs to be considered in multistage robust problem, and this difference hinders the direct application of SDDP method. To solve the multistage robust energy management problem introduced above in this work, a robust version of dual dynamic programming method is proposed which is also called robust dual dynamic programming (RDDP) method [198]. This method has a similar framework with the popular SDDP method which also consists of a forward pass and a backward pass procedure. Generally, the forward pass is used to search for the worst-case realizations of random variables and corresponding optimal recourse decisions which will generate upper and lower bounds of the worst-case cost-to-go function at each stage, while the backward pass aims to refine the obtained upper and lower bounds.

Compared with the SDDP method, the RDDP method introduces an upper bound to approximate the worst-case cost-to-go function in addition to the common lower bound. Based on the assumption of a convex affine worst-case cost-to-go function, the upper bound is obtained by inner approximation method [199], and the lower bound is attained by outer approximation which will be discussed in detail below.

5.2.2.1 Upper Bound Problem

Recall that the multistage robust problem in (5.19), the worst-case cost-to-go function Q_{t+1} in this problem is unknown at current stage which makes the evaluation of Q_t difficult. A good idea is to find the approximation of this function so that the robust optimal solution can be obtained. Therefore, in this work, we approximate each worst-case cost-to-go function $Q_t, t = 2, \dots, T$ by iteratively constructing its lower and upper bounds which satisfy $\underline{Q}_t(x_{t-1}) \leq Q_t(x_{t-1}) \leq \bar{Q}_t(x_{t-1})$. To get the upper bound $\bar{Q}_t(x_{t-1})$, we can define and solve the following upper bound problem:

$$\begin{aligned} \bar{S}_t(x_{t-1}; \xi_t) &= \max_{\xi_t} \min_{x_t} c_t^\top x_t + \bar{Q}_{t+1}(x_t) \\ \text{s.t. } & B_t x_{t-1} + A_t x_t \geq b_t(\xi_t) \\ & x_t \geq 0, \quad t = 2, \dots, T \end{aligned} \quad (5.20)$$

where $\bar{Q}_{t+1}(x_t)$ is the upper bound of $Q_{t+1}(x_t)$. Generally, we have $\bar{S}_t(x_{t-1}; \xi_t) \geq Q_t(x_{t-1})$ since the upper bound $\bar{Q}_{t+1}(x_t)$ is considered in $\bar{S}_t(x_{t-1}; \xi_t)$. Likewise, if $\bar{Q}_{t+1}(x_t) = Q_{t+1}(x_t)$, then the optimal value of $\bar{S}_t(x_{t-1}; \xi_t)$ would be the same with that of $Q_t(x_{t-1})$. Hence, the optimal solution of problem $\bar{S}_t(x_{t-1}; \xi_t)$ can be used to approximate $Q_t(x_{t-1})$ from above.

It is obvious that the problem \bar{S}_t is a common max-min problem which can be solved by dual method or vertex enumeration method in condition that $\bar{Q}_{t+1}(x_t)$ is a convex piecewise affine function. By dualizing the inner minimization problem, the problem can be transformed into a single level problem and solved by mixed-integer linear programming method. Another method is the vertex enumeration method [198]. Since it has been demonstrated that the inner minimization problem in (5.20) is convex in the parameter ξ_t , the optimal solution is obtained at a certain extreme point of the uncertainty set U_ξ . Thus, we can try to identify the extreme points of the set U_ξ and solve a finite number of linear programming problems. The vertex enumeration method can be much easier when the number of extreme points is small in the problem or the parallel computation technique can be used. Considering the problem scale and the interval uncertainty set, the vertex enumeration method will be used to solve the upper bound problem in this work.

As the explicit expression of $\bar{Q}_{t+1}(x_t)$ is unknown, we use the inner approximation method [199] for it in this work so that we can solve the upper bound problem approximately. Assuming that J_t points $(\underline{x}_t^j, \bar{S}_{t+1}^j(\underline{x}_t^j))$, $j = 1, \dots, J_t$ have been collected for the upper bound \bar{Q}_{t+1} of stage $t + 1$ worst-case cost-to-go function, then we can approximate the upper bound function with the lower convex envelop or convex hull of these points, and the upper bound problem in (5.20) can be expressed approximately as follows:

$$\begin{aligned} \bar{S}_t(x_{t-1}; \xi_t) &= \max_{\xi_t} \min_{x_t} c_t^\top x_t + \sum_{j=1}^{J_t} \lambda^j \bar{S}_{t+1}^j \\ &\text{s.t. } B_t x_{t-1} + A_t x_t \geq b_t(\xi_t) \\ &\quad x_t = \sum_{j=1}^{J_t} \lambda^j x_t^j, \quad \sum_{j=1}^{J_t} \lambda^j = 1 \\ &\quad \lambda^j \geq 0, \quad x_t \geq 0, \quad t = 2, \dots, T. \end{aligned} \tag{5.21}$$

where λ^j is the auxiliary variable. Note that the critical point of using inner approximation is the convex assumption of the worst-case cost-to-go function and the inner approximation property can be inherited from stage to stage.

5.2.2.2 Lower Bound Problem

To obtain the lower bound $\underline{Q}_t(x_t)$, we can solve the following lower bound problem with any one fixed uncertainty realization $\tilde{\xi}_t$:

$$\begin{aligned} \underline{S}_t(x_{t-1}; \tilde{\xi}_t) = \min & c_t^\top x_t + \underline{Q}_{t+1}(x_t) \\ \text{s.t.} & B_t x_{t-1} + A_t x_t \geq b_t(\tilde{\xi}_t) \\ & x_t \geq 0, \quad t = 2, \dots, T \end{aligned} \quad (5.22)$$

where $\underline{Q}_{t+1}(x_t)$ is the lower bound of $Q_{t+1}(x_t)$. Compared with $Q_t(x_{t-1})$, only one realization of the random variable and the lower bound \underline{Q}_{t+1} of stage $(t+1)$ worst-case cost-to-go function are considered here, thus the optimal value of this problem can bound $Q_t(x_{t-1})$ from below. With a convex and piecewise affine function $\underline{Q}_{t+1}(x_t)$, the lower bound problem can be solved as a linear programming problem, and we can find the lower bound of $Q_t(x_{t-1})$ by the supporting hyperplanes or outer approximation of \underline{S}_t during the iteration process. More specifically, we can define the approximate problem by replacing $\underline{Q}_{t+1}(x_t)$ with an auxiliary variable θ_{t+1}^a constrained by a set of cutting planes [197] as follows:

$$\begin{aligned} \underline{S}_t(x_{t-1}; \tilde{\xi}_t) = \min & c_t^\top x_t + \theta_{t+1}^a \\ \text{s.t.} & B_t x_{t-1} + A_t x_t \geq b_t(\tilde{\xi}_t), [\pi_t(\tilde{\xi}_t)] \\ & \theta_{t+1}^a + \pi_{t+1,k}^\top B_{t+1} x_t \geq \bar{g}_{t+1,k}, \quad k = 1, \dots, K \\ & x_t \geq 0, \quad t = 2, \dots, T \end{aligned} \quad (5.23)$$

where the variable $\pi_t(\tilde{\xi}_t)$ is the Lagrange multiplier vector corresponding to the constraint and $\bar{g}_{t+1,k}$ is the intercept of k th cut defined as follows:

$$\bar{g}_{t+1} = \underline{S}_{t+1}(x_t; \tilde{\xi}_{t+1}) + \pi_{t+1}^\top B_{t+1} x_t \quad (5.24)$$

where x_t is the optimal solution of problem $\underline{S}_{t+1}(x_t; \tilde{\xi}_{t+1})$. The cuts can also be generated and expressed by the cut calculation algorithm in [200] which has the same principle of outer approximation method.

5.2.2.3 RDDP Method

The upper bound and lower bound problems introduced above are the main components of the proposed RDDP method. They are frequently constructed and refined in the forward pass and backward pass during the iteration process until we find the optimal worst-case cost-to-go function. Based on this idea, the procedure of RDDP method can be summarized as follows:

1) *Initialization*: Set the lower bound and $\underline{Q}_t(x_{t-1}) = -M$ and the upper bound $\bar{Q}_t(x_{t-1}) = +M$, for $t = 2, \dots, T$, where M is a large enough number and it can also be set to ∞ . Note that this is for the first iteration before the update of lower and upper bounds.

In addition, set $\underline{Q}_{T+1}(x_T) = \overline{Q}_{T+1}(x_T) = 0$ for the last stage.

2) *First stage problem*: Solve the lower bound problem for the first stage as follows with a deterministic realization of the random variable and store the optimal solution \underline{x}_1 . If the difference between the upper bound \overline{Q}_2 and the lower bound \underline{Q}_2 is less than a small predefined parameter or the maximum iteration is reached, then terminate the algorithm, else go to step 3).

$$\begin{aligned} \min \quad & c_1^\top x_1 + \underline{Q}_2(x_1) \\ \text{s.t.} \quad & A_1 x_1 \geq b_1 \\ & x_1 \geq 0. \end{aligned} \quad (5.25)$$

3) *Forward pass*: For $t = 2, \dots, T$, solve the upper bound problem $\overline{\mathcal{S}}_t(\underline{x}_{t-1}; \xi_t)$ in (5.20) and get the optimal solution $\overline{\xi}_t^{fw}$. Then based on the worst-case uncertainty realization $\overline{\xi}_t^{fw}$, solve the lower bound problem $\underline{\mathcal{S}}_t(\underline{x}_{t-1}; \overline{\xi}_t^{fw})$ in (5.22) and store the optimal solution \underline{x}_t .

4) *Backward pass*: For $t = T, \dots, 2$, solve the upper bound problem $\overline{\mathcal{S}}_t(\underline{x}_{t-1}; \xi_t)$ based on \underline{x}_{t-1} and obtain the new optimal solution $\overline{\xi}_t^{bw}$. If $\overline{\mathcal{S}}_t(\underline{x}_{t-1}) < \overline{Q}_t(\underline{x}_{t-1})$, then collect the point $(\underline{x}_{t-1}, \overline{\mathcal{S}}_t(\underline{x}_{t-1}))$ to update the upper bound approximation in (5.21). Based on $\overline{\xi}_t^{bw}$ and \underline{x}_{t-1} , we can solve the lower bound problem $\underline{\mathcal{S}}_t(\underline{x}_{t-1}; \overline{\xi}_t^{bw})$, and let π_t be the dual optimal solution. If $\underline{\mathcal{S}}_t(\underline{x}_{t-1}; \overline{\xi}_t^{bw}) > \underline{Q}_t(\underline{x}_{t-1})$, then update the lower bound problem in (5.23) by adding the following cut into the previous stage problem:

$$\theta_t^a + \pi_t^\top B_t x_{t-1} \geq \overline{g}_t \quad (5.26)$$

$$\overline{g}_t = \underline{\mathcal{S}}_t(\underline{x}_{t-1}; \overline{\xi}_t^{bw}) + \pi_t^\top B_t \underline{x}_{t-1}. \quad (5.27)$$

This lower bound update method has the same idea with that in the common SDDP method, which iteratively adds Benders' cuts to approximate a convex and piecewise affine function. After completing step 4), go back to step 2).

To better illustrate the RDDP solution process for multistage robust energy management, a schematic diagram based on the above procedure is given in Fig. 5.2. The idea is to express the multistage problem in a nested formulation. Take the first stage problem as an example, the uncertainty is revealed in the first stage, and different cost-to-go functions are used to capture the uncertain parameters in the following stages. As shown in this figure, we need to obtain the estimated $\underline{Q}_2(x_1)$ function for the first stage problem. Similarly, the corresponding cost-to-go function $\underline{Q}_{t+1}(x_t)$ should also be estimated at stage t . These cost-to-go functions can be approximately attained by solving the upper bound problem and lower bound problem iteratively. With the above RDDP method, we can solve the proposed multistage robust energy management problem, more

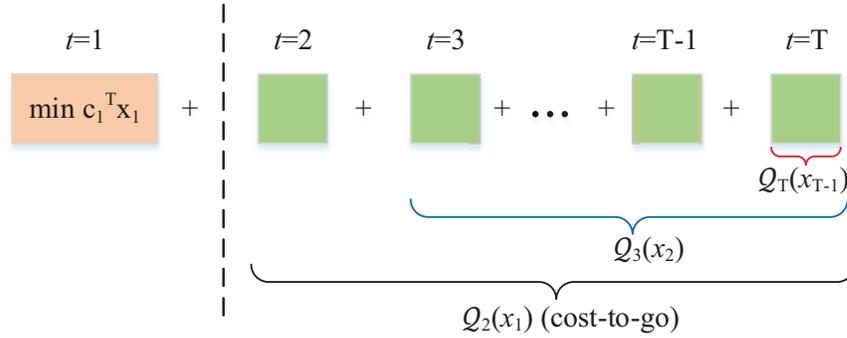


Figure 5.2: Schematic diagram of solution process.

specifically, we can find the worst-case scenario and get the robust optimal solution. The case study and numerical results will be presented in next subsection.

5.2.3 Case Study

In order to validate the effectiveness of the proposed multistage robust model and solution methodology, a case study for a grid-connected microgrid with real-world data is conducted in this subsection. First, a description of the studied microgrid and related parameters and data is given. Then the simulation results of the robust optimal solution from RDDP method are presented. The comparison between the proposed method and other methods is also analyzed.

5.2.3.1 Microgrid Description and Data

In this work, a grid-connected microgrid composed of three diesel generators, an energy storage system, a wind turbine and uncertain loads is studied. Without loss of generality, wind power is considered as the renewable generation. The time horizon for the energy management problem is set to 24 hours, i.e., the number of stages T is 24, and the time step Δt is 1 hour. For three diesel generators, the involved parameters are collected from related references [60], [59] and summarized in Table 5.2. The considered storage system has a maximum capacity of 90 kWh, and the minimum storage level is set to be 20 kWh to avoid overdischarging. The limits for charging and discharging rate are both set to 50 kWh, and the same charging and discharging efficiency is used which is 0.95. In addition, the initial energy storage level is half of the maximum capacity, and the degradation cost coefficient of the storage system is set to 0.0035 \$/kWh [61].

Renewable generation and load demand are considered to be uncertain factors. As mentioned before, the uncertainty set for these random variables can be obtained by different methods. Here the point forecast values are used to generate the interval uncertainty set. The point forecast data for wind power and load are collected from the

Table 5.2: Parameters for the generators

Unit	\underline{P}_i (kW)	\bar{P}_i (kW)	R_i^{up}/R_i^{dn} (kW)	a_i^{dg} (\$/kWh)	b_i^{dg} (\$/h)
G1	10	50	30	0.13	30
G2	8	45	25	0.2	50
G3	15	70	40	0.25	80

IESO website [201] which are properly scaled. Based on the point forecast values, we create the interval sets by setting up a certain deviance (e.g., 10%) from the nominal values. The generated intervals and the point forecast values are given in Fig. 5.3. Note that the interval bounds are assumed to be known in this work. If the bounds are also uncertain, we can try to estimate them using some popular interval prediction methods [31]. In addition, it can be a potential future research topic to consider the bounds as decision variables.

As the microgrid operates in grid-connected mode, we also need to consider the exchange power with the main grid and the market electricity price. The maximum power exchange including the purchase and sale is set to 100 kWh in this work, and the electricity prices are acquired from the NYISO website [202]. The appropriately scaled day-ahead electricity price is used as the purchase price from the market, and the electricity sale price is set to 80% of the purchase price for simplicity [60], which are both presented in Fig. 5.4. In electricity market, the electricity price can be determined by static or dynamic pricing scheme. Since the renewable generation and load are usually main factors of uncertainty in a microgrid system, the static prices are used in this work which includes fixed prices and time-of-use prices [194]. For the effect of uncertain prices, it will lead to uncertain coefficients in the objective function, and this may be studied in future work. In addition, a linear penalty cost function is used for possible load shedding and the unit penalty cost is equal to 10 \$/kWh in this study [61].

5.2.3.2 Simulation Results

Based on the above relevant parameters and data, the multistage robust energy management problem can be solved with the RDDP method. All the experiments are implemented in MATLAB environment with Gurobi solver on a desktop with an Intel Core i7-6700 CPU 3.40 GHz and 8 GB of RAM.

The convergence of the algorithm is first analyzed. In this experiment, the maximum iteration is used as the termination criterion. The evolution of the lower and upper bounds for the worst-case cost-to-go function value in the first stage problem is shown in Fig. 5.5. As can be seen from this figure, the lower and upper bounds almost converge to

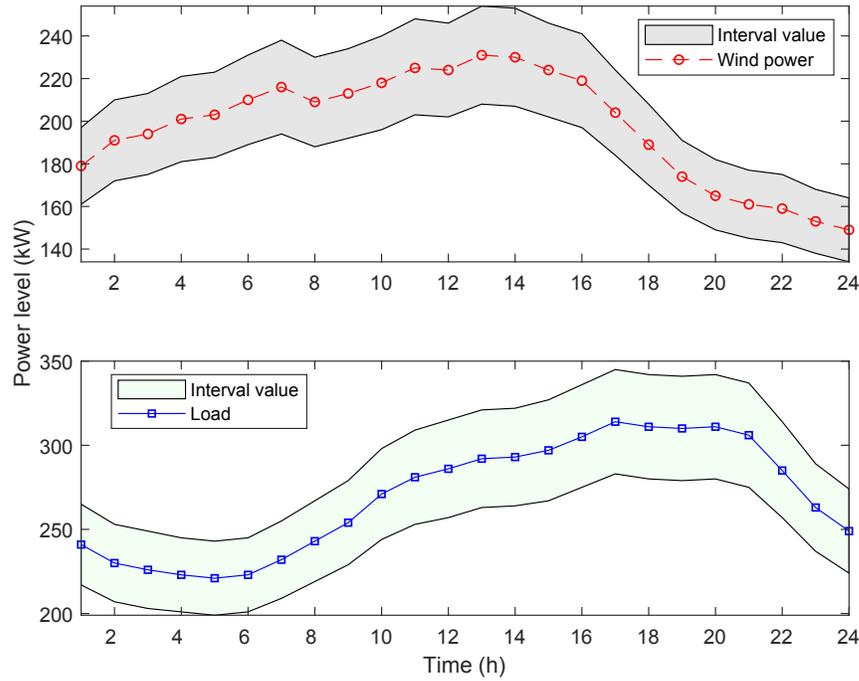


Figure 5.3: Wind power and load data.

the same value and their difference is very small (0.44%). Thus the obtained solution can be considered as the optimal robust solution when the algorithm terminates. Note that we can also set a small positive parameter ϵ_1 (e.g., 10^{-2}) to terminate the method, which may need less iterations. The execution time for solving this problem with this method is about 255s which is acceptable in practice. For the proposed multistage robust problem, the critical point is that we need to determine the unknown cost-to-go functions in this method. It is supposed that the approximate optimal cost-to-go functions are obtained when the algorithm terminates. Accordingly, with the estimated cost-to-go functions, optimal decisions or implementable policies can be attained which can guard against any uncertainty realization in the uncertainty set.

After the termination of the method, the optimal robust solution including the schedule of DG units, ESS and the electricity exchange with the market can be obtained corresponding to the worst-case realizations of uncertainty. The optimal schedule of three DG units is given in Fig. 5.6. From this figure, we can find out that G1 always has the maximum output since the second time period. For G2 and G3, they both have the minimum output at the first few time slots, then the output increases gradually until the maximum output is reached which corresponds to the change of load demand. Taking the generation cost function into account, we can see that G1 always maintains the maximum output due to its lowest cost, while the other two units have larger generation cost and only increase their output when the load becomes larger.

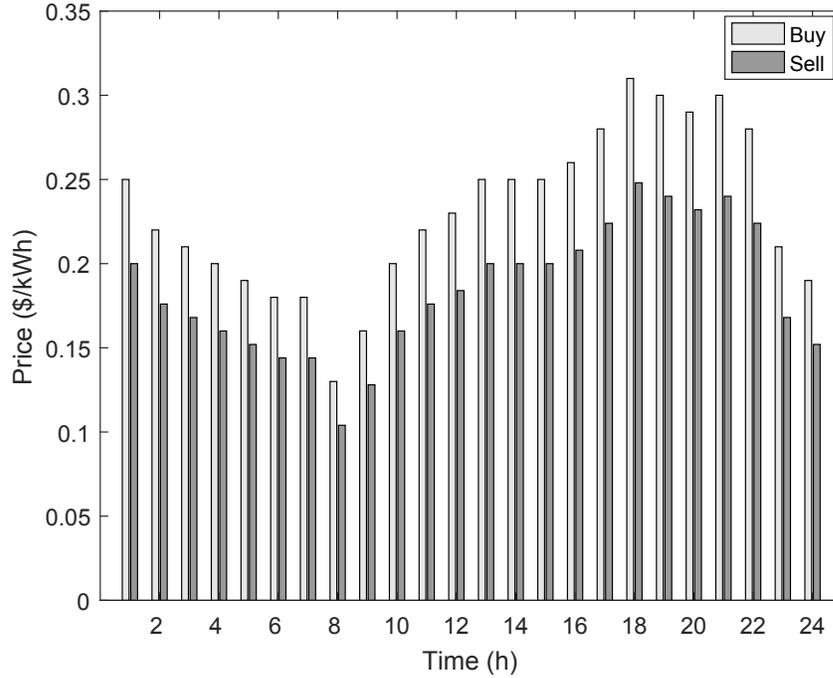


Figure 5.4: Electricity exchange price.

Similarly, we can also attain the optimal schedule of the ESS and the power exchange with the main grid as shown in Fig. 5.7. Note that the blue line in this figure represents the storage level of ESS at each time slot. As shown in this figure, the ESS first discharges to the minimum value considering the initial storage level (half of the maximum capacity), then it charges almost to the maximum capacity and stay unchanged for some time periods. When the load demand increases at later time stages, the ESS starts to discharge. For the electricity exchange between the microgrid and main grid, it is influenced by the electricity prices. For example, from time period 13 to 22 when the electricity purchase price is greater than 0.25, there is no electricity purchase unless the load demand is large enough and cannot be satisfied by the cheaper DG units and ESS. When the electricity purchase price decreases, more power will be purchased from the main grid to substitute the generator output, i.e, the power supply from the main grid would be cheaper than the generation of unit G3. In this case, the total cost for the robust optimal solution is 4356.17 \$. Note that the above optimal robust solution is obtained with the given uncertainty set. If the uncertainty set is adjusted, the corresponding optimal solution will also change.

5.2.3.3 Comparison with Other Methods

In this subsection, the proposed method is compared with some other methods to verify its effectiveness and advantage. Since the objective of multistage robust energy management

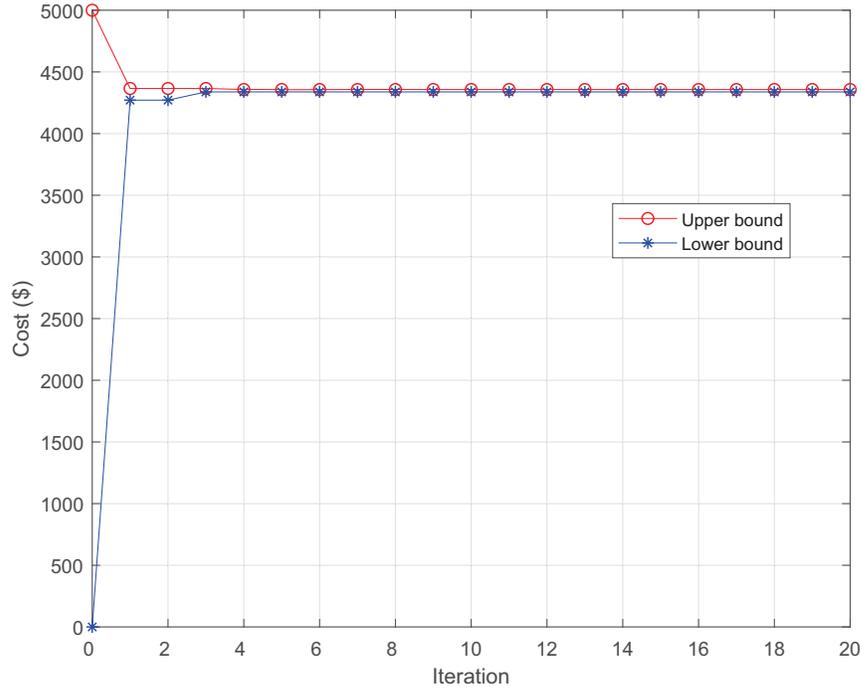


Figure 5.5: Evolution of lower and upper bounds.

problem is to find the robust optimal solution, we can focus on the comparison of solutions from different methods regarding the system reliability. The comparison methods include SDDP method, sample average approximation (SAA) method and the deterministic method with perfect information.

SDDP method is usually used to solve the multistage stochastic programming problems as shown in (5.28) and only the outer approximation for the expected future cost-to-go function is considered. It has a similar framework with the proposed method which mainly consists of a forward pass to generate a statistical upper bound and a backward pass to refine the lower bound. There are different convergence criteria for this method, for instance, the method terminates when the difference between the lower bound and the upper bound is very small, and here we use the improved convergence criterion reported in [203]. The uncertain wind power and load are assumed to follow uniform distribution in the interval set. Note that the uniform distribution here is used to generate some scenarios from the interval set for SDDP method. Some other distributions such as Weibull distribution or normal distribution can also be utilized. In addition, IGDT method [204] may also be investigated to deal with the uncertainty in multistage microgrid energy management problem. SAA is a common method to solve stochastic programming problems. In this work, SAA is actually used to solve a two-stage robust problem as we generate scenarios from the interval uncertainty set and the uncertainties are all known

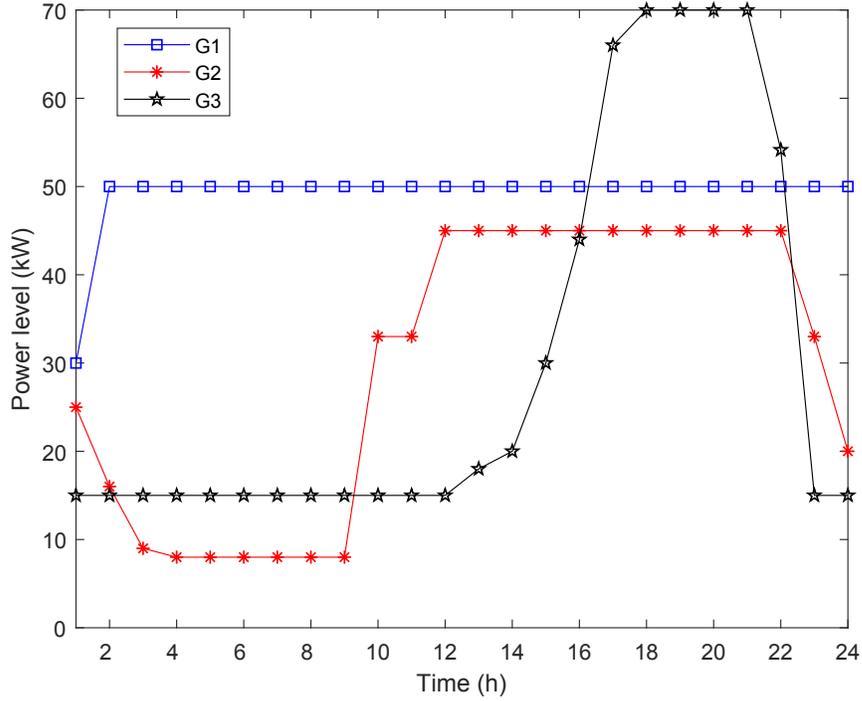


Figure 5.6: Power generation level of DG units.

with the scenario. Both SDDP and SAA method belong to stochastic approach, and we will check them with the worst-case scenario for system reliability concern. In addition, a deterministic two-stage problem with perfect information, i.e., the worst-case scenario, is also studied, and the cost function at each stage are all known, and the problem can be solved without the approximation of future cost-to-go function.

$$\min_{\substack{A_1 x_1 \geq b_1 \\ x_1 \geq 0}} c_1^\top x_1 + \mathbb{E} \left[\min_{\substack{B_2 x_1 + A_2 x_2 \geq b_2 \\ x_2 \geq 0}} c_2^\top x_2 + [\dots + \mathbb{E} [\min_{\substack{B_T x_{T-1} + A_T x_T \geq b_T \\ x_T \geq 0}} c_T^\top x_T] \dots] \right] \quad (5.28)$$

The total cost comparison of the solutions from different methods is given in Fig. 5.8. From this figure, we can see that the total cost of RDDP method for the worst-case scenario is much lower than that of SDDP method and SAA method which validates the advantage of the robust method. Moreover, it is very close to the cost of the deterministic problem with perfect information. Note that the cost of SDDP method is the average value of five simulation experiments and the number of scenarios is 100 for SAA method. For better comparison, the corresponding detailed costs of these methods are listed in Table 5.3. As can be seen from this table, with larger generation cost and exchange cost, both the RDDP and deterministic method have no penalty cost, while SDDP and SAA method generate very large penalty cost for the unexpected load shedding in the worst-case scenario. Note that the negative exchange cost represents the profit from selling excess power. Therefore, the RDDP method can achieve robust optimal energy management by introducing more generator output and electricity purchase from the main grid. Also, the comparison

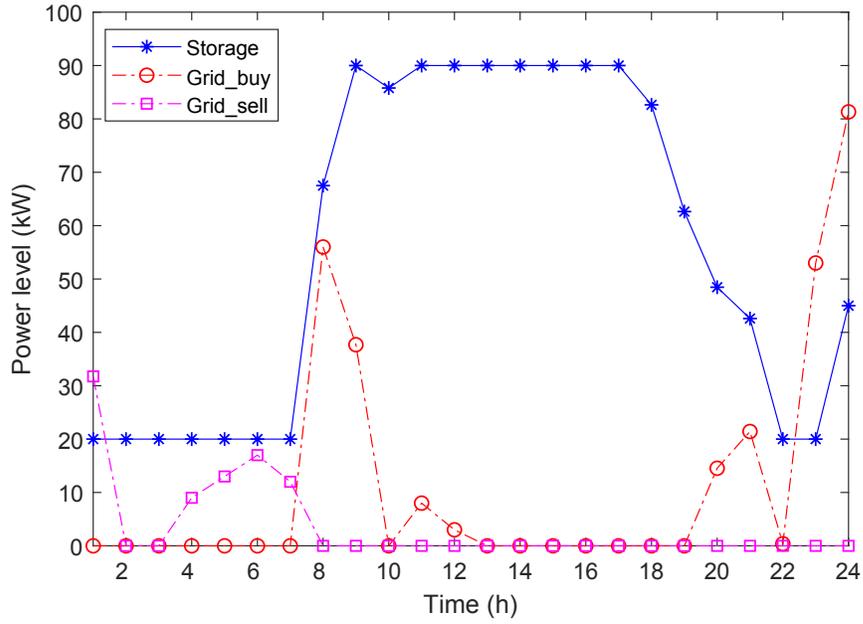


Figure 5.7: Schedule of storage and power exchange.

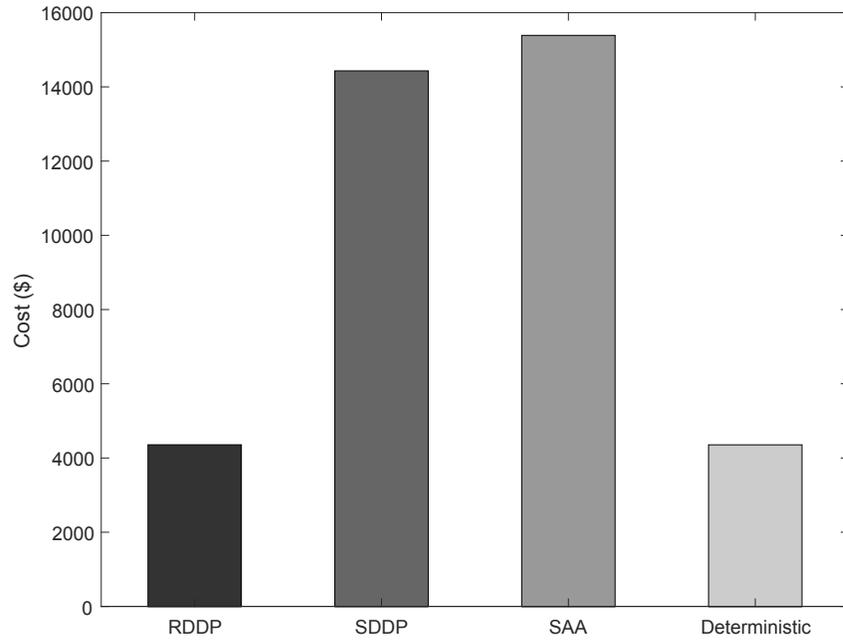


Figure 5.8: The total cost of different methods.

results show the advantage of the proposed multistage model, i.e., the system is more reliable in the worst-case realizations of uncertainty.

Although the above analysis focuses on the comparison of solution methods, the comparison of system models may also be investigated. As mentioned before, microgrid

Table 5.3: Detailed cost comparison for different methods

Method	RDDP	SDDP	SAA	Deterministic
Generation cost	4316.34	4220.14	4209.79	4313.80
Storage cost	0.68	0.65	0.67	0.67
Exchange cost	39.15	-68.74	-82.92	39.97
Penalty cost	0	10279.70	11260.35	0
Total cost	4356.17	14431.75	15387.90	4354.44

system has been widely studied in the literature, and most of the system models have a similar structure. In this work, the microgrid system model is developed based on previous literature. However, the energy management problem is studied in a multistage perspective which is a significant difference compared with the previous single stage or two-stage problems. In addition, the distributed microgrid system is the main focus here, and it is also worth studying the proposed approach with a large realistic system in the future, e.g., the economic dispatch problem in transmission system.

5.3 Multi-period Distributionally Robust Energy Management for CCHP Microgrids

In this section, multi-period energy management for CCHP microgrids including UC scheduling is investigated with the DRO uncertainty modeling technique instead of RO method. The proposed CCHP microgrid system model and two-stage multi-period formulation are introduced in subsection 5.3.1. In subsection 5.3.2, the new ambiguity set is demonstrated and the solution methodology based on LDR is derived. Case studies are carried out to show the results and performance of the proposed approach in subsection 5.3.3.

5.3.1 Problem Formulation

Microgrids can be operated in grid-connected or islanded mode. In this work, the energy management of a grid-connected CCHP microgrid is studied which usually consists of conventional generation units such as micro turbines (MTs), renewable generation, storage systems, electric load and thermal load. The basic structure and energy flows of a CCHP microgrid are depicted in Fig. 5.9. As shown in this figure, there are two energy flows: electricity energy and thermal energy to satisfy the electric load and thermal load respectively. Since the cooling load can be met by transforming some amount of heat energy with the absorption chiller, we combine the heat and cooling load together here [118]. For the energy supply, MTs can generate electric energy and heat energy simultaneously, and the microgrid can exchange energy with the main grid. Considering the coupling nature of MTs, the gas furnace and thermal storage system are introduced to flexibly supply enough heat power. In this work, a two-stage

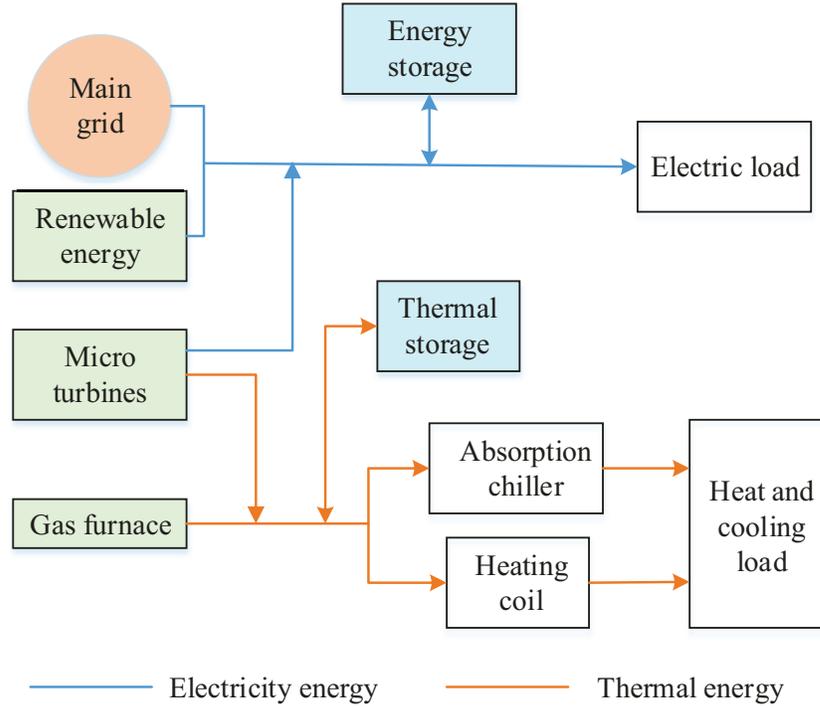


Figure 5.9: The scheme of a grid-connected CCHP microgrid.

multi-period energy management model is studied for CCHP microgrids, specifically, the day-head scheduling of MTs is investigated in the first stage and multi-period dispatch considering non-anticipativity is studied in the second stage. In this subsection, the general components models and constraints are introduced first, then the multi-period problem formulation is given. The detailed models of the studied microgrid are formulated in the following subsections.

5.3.1.1 CCHP Units

In a CCHP microgrid, the most important component is the CCHP unit which can significantly improve the overall energy utilization efficiency by fully utilizing the waste heat from generating electricity. For CCHP unit in the studied system, MTs are used which can directly generate power with natural gas. Specifically, the electric and heat output power of MTs are expressed as follows [112]:

$$P_{it} = G_{it}Q^G\eta_i^{MT}, \forall i, t \quad (5.29)$$

$$h_{it}^{MT} = G_{it}Q^G(1 - \eta_{i,loss}^{MT} - \eta_i^{MT}) \quad (5.30)$$

where $G_{i,t}$ is the natural gas consumption of unit i at time t , Q^G is the heat value of gas, η_i^{MT} is the efficiency coefficient and $\eta_{i,loss}^{MT}$ is the loss coefficient of unit i .

As conventional generation units, the start-up/shut-down operation, minimum on time and off time limits should be considered in the day-ahead stage [61]. The related constraints are given below:

$$-y_{i(t-1)} + y_{it} - y_{is} \leq 0, \forall i, t, 1 \leq s - (t - 1) \leq UT_i \quad (5.31)$$

$$y_{i(t-1)} - y_{it} + y_{is} \leq 1, \forall i, t, 1 \leq s - (t - 1) \leq DT_i \quad (5.32)$$

$$y_{it} - y_{i(t-1)} = y_{it}^+ - y_{it}^-, y_{it}^+ + y_{it}^- \leq 1, \forall i, t \quad (5.33)$$

$$y_{it}, y_{it}^+, y_{it}^- \in \{0, 1\}, \forall i, t \quad (5.34)$$

where $y_{it}, y_{it}^+, y_{it}^-$ are binary variables representing the on/off status, start-up and shut-down operation of MTs, respectively.

In addition, we also have the generation capacity limit, ramping up/down restriction for MTs which are represented as below:

$$y_{it} \underline{P}_i \leq P_{it} \leq y_{it} \bar{P}_i, \forall i, t \quad (5.35)$$

$$P_{it} - P_{i(t-1)} \leq \bar{P}_i y_{it}^+ + R_i^{up} y_{i(t-1)}, \forall i, t \quad (5.36)$$

$$P_{i(t-1)} - P_{it} \leq \bar{P}_i y_{it}^- + R_i^{dn} y_{it}, \forall i, t. \quad (5.37)$$

Considering the coupling electric and heat output of MTs, other heating devices such as gas furnace can be introduced to supply sufficient heat power in this system. Gas furnace generates heat power by combusting natural gas, and the output constraints are as follows:

$$h_{gt}^{GF} = G_{gt}^{GF} Q^G \eta_g^{GF}, \forall g, t \quad (5.38)$$

$$\underline{h}_g^{GF} \leq h_{gt}^{GF} \leq \bar{h}_g^{GF}, \forall g, t \quad (5.39)$$

where G_{gt}^{GF} is the gas consumption of unit g at time t , and η_g^{GF} is the output efficiency. Note that the heat power can be transformed into cooling power via absorption chillers to meet the cooling load.

5.3.1.2 Storage Systems

Energy storage systems can play an important role in energy supply of a microgrid. Both electric energy storage system (ESS) and thermal storage system (TSS) can be used in a CCHP microgrid. For ESS system, it should satisfy the charging and discharging restrictions at each time slot, and the storage level should be limited by the lower and upper bounds in the whole horizon which can be expressed below:

$$\underline{r}_j^{e+} \leq r_{jt}^{e+} \leq \bar{r}_j^{e+}, \underline{r}_j^{e-} \leq r_{jt}^{e-} \leq \bar{r}_j^{e-}, \forall j, t \quad (5.40)$$

$$\underline{E}_j^e \leq E_{j0}^e + \sum_{\tau \in [1:t]} (r_{j\tau}^{e+} \eta_j^{e+} - r_{j\tau}^{e-} / \eta_j^{e-}) \Delta t \leq \bar{E}_j^e, \forall j, t \quad (5.41)$$

where E_{j0}^e denotes the initial energy storage level, η_j^{e+} and η_j^{e-} are charging and discharging efficiency, respectively. Similarly, the energy storage dynamics and related constraints are represented as follows:

$$\underline{r}_m^{q+} \leq r_{mt}^{q+} \leq \bar{r}_m^{q+}, \underline{r}_m^{q-} \leq r_{mt}^{q-} \leq \bar{r}_m^{q-}, \forall m, t \quad (5.42)$$

$$\underline{E}_m^q \leq E_{m0}^q + \sum_{\tau \in [1:t]} (r_{m\tau}^{q+} \eta_m^{q+} - r_{m\tau}^{q-} / \eta_m^{q-}) \Delta t \leq \bar{E}_m^q, \forall m, t. \quad (5.43)$$

5.3.1.3 Load Balance and Objective

As discussed above, there are electric energy flow and thermal energy flow in the CCHP microgrid system which are used to satisfy the corresponding loads. Based on the heat power from MTs, gas furnace and TSS, the thermal load balance can be expressed as follows:

$$\sum_{i=1}^{N^{MT}} h_{it}^{MT} + \sum_{g=1}^{N^{GF}} h_{gt}^{GF} + \sum_{m=1}^{N^q} (r_{mt}^{q-} - r_{mt}^{q+}) = p_t^{HC} + p_t^{AC}, \forall t \quad (5.44)$$

$$0 \leq p_t^{HC} \leq \bar{p}^{HC}, 0 \leq p_t^{AC} \leq \bar{p}^{AC}, \forall t \quad (5.45)$$

where p_t^{HC} and p_t^{AC} are the heat power supplied to the heating coil and absorption chiller [118], respectively, as shown in Fig. 5.9. The final energy conveyed to the demand is limited by the device's coefficient of performance (COP), and the conversion relation is given below:

$$q_t^{HE} = p_t^{HC} \Delta t \cdot \text{COP}^H, q_t^{CE} = p_t^{AC} \Delta t \cdot \text{COP}^C, \forall t. \quad (5.46)$$

CCHP microgrid system is usually used in smart buildings such as residential house or commercial buildings. To make the proposed CCHP microgrid model more realistic, the temperature dependent thermal load is considered in this work. In addition, the dynamics of the heating and cooling system such as air conditioning system can be described by setting the indoor temperature. Specifically, the thermal load can be modeled by the thermodynamic equation which is dependent on the indoor temperature setpoint and environmental temperature as follows:

$$(q_t^{HE} - q_t^{CE}) / \Delta t = c^{air} (d\theta^{in} / dt) - (\theta_t^{am} - \theta_t^{in}) / R^{tr}, \forall t \quad (5.47)$$

where c^{air} is a coefficient representing the air specific heat capacity, θ^{in} and θ^{am} are the indoor and ambient temperature, respectively, and R^{tr} is the thermal resistance of building envelop.

Considering the inertia of thermal energy, the indoor temperature actually alters slowly and it can be regarded as a constant within each time slot (e.g., one hour). Therefore, the thermodynamic equation in (5.47) can be transformed into a discrete state model as below:

$$(q_t^{HE} - q_t^{CE}) / \Delta t = c^{air} (\theta_t^{in} - \theta_{t-1}^{in}) / \Delta t - (\theta_t^{am} - \theta_t^{in}) / R^{tr}, \forall t. \quad (5.48)$$

With this model, we can control the thermal load by setting different indoor temperatures. As the comfortable indoor temperature usually has a certain range, we also have the following constraint:

$$\underline{\theta}_t^{in} \leq \theta_t^{in} \leq \bar{\theta}_t^{in}, \forall t. \quad (5.49)$$

Without loss of generality, wind power is considered as the renewable generation in this work. Combining the output power of MTs, ESS and the electricity exchange with the main grid, the constraints about electric load balance are expressed as follows:

$$\sum_{i=1}^{N^{MT}} P_{it} + \sum_{j=1}^{N^e} (r_{jt}^{e-} - r_{jt}^{e+}) + \sum_{n=1}^{N^w} w_{nt} + P_t^{buy} - P_t^{sell} = P_t^{load}, \forall t \quad (5.50)$$

$$0 \leq P_t^{buy} \leq \bar{P}^{buy}, 0 \leq P_t^{sell} \leq \bar{P}^{sell}, \forall t \quad (5.51)$$

where w_{nt} represents the uncertain wind power, and its uncertainty modeling with DRO method is introduced in the next subsection. For the main grid, we can purchase electricity from or sell excess power to it with the power flow limit.

The proposed energy management for CCHP microgrids is formulated in a two-stage framework, therefore, we need to consider the day-ahead scheduling cost in the first stage and dispatch or recourse cost in the second stage for the objective function. In particular, the first-stage cost includes start-up, shut-down and no load cost of MTs, the second-stage cost consists of fuel cost, degradation cost of ESS and TSS, and the electricity transaction cost with main grid. Mathematically, the objective function is represented as follows:

$$C^{tot} = \sum_t \sum_i (SU_i y_{it}^+ + SD_i y_{it}^- + c_i^{NL} y_{it}) + \max_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[Q(\mathbf{y}, \mathbf{w})] \quad (5.52)$$

$$Q(\mathbf{y}, \mathbf{w}) = \sum_{t=1}^T \left\{ c^{gas} \left(\sum_{i=1}^{N^{MT}} G_{it}^{MT} + \sum_{g=1}^{N^{GF}} G_{it}^{GF} \right) \right. \\ \left. + \sum_j c_j^{ess} (r_{jt}^{e+} \eta_j^{e+} + r_{jt}^{e-} / \eta_j^{e-}) + \sum_m c_m^q (r_{mt}^{q+} \eta_m^{q+} + r_{mt}^{q-} / \eta_m^{q-}) + (c_t^{buy} P_t^{buy} - c_t^{sell} P_t^{sell}) \right\} \quad (5.53)$$

where the set \mathcal{D} is the ambiguity set for uncertain wind power \mathbf{w} , $Q(\cdot)$ is the second-stage operational cost, and \mathbf{y} represents the first-stage decision variables. In addition, a linear degradation cost is used here for storage systems [61].

5.3.1.4 Multi-period Formulation

For notational conciseness, the general two-stage CCHP microgrid model introduced above can be written in a compact matrix formulation as follows:

$$\min_{\mathbf{y} \in \mathcal{Y}} \mathbf{a}^\top \mathbf{y} + \max_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[Q(\mathbf{y}, \mathbf{w})] \quad (5.54a)$$

$$Q(\mathbf{y}, \mathbf{w}) = \min_{\mathbf{x}} \{ \mathbf{b}^\top \mathbf{x} : \mathbf{T}\mathbf{y} + \mathbf{W}\mathbf{x} \geq \mathbf{h} - \mathbf{H}\mathbf{w} \} \quad (5.54b)$$

where the set \mathcal{Y} represents the first-stage constraints including constraints (5.31)-(5.34), the matrix inequality expression in (5.54b) consists of the second-stage constraints (5.29)-(5.30) and (5.35)-(5.51), and the second-stage decision variables are collected in \mathbf{x} .

As discussed above, in a common two-stage model, there is an unrealistic assumption that the non-anticipativity is not considered. In other words, it is assumed that the second-stage dispatch decisions are optimized simultaneously with the disclosure of all uncertainty realizations in the beginning [205]. However, the uncertain wind power is revealed sequentially in practice and the dispatch decisions can only be made according to the uncertainty realizations up to current period, i.e., the dispatch decision at time t is dependent on the wind power realizations from time 1 to t which can be expressed as $\mathbf{w}_{[t]}$. Accordingly, we can formulate the two-stage multi-period problem enforcing non-anticipativity as follows:

$$\min_{\mathbf{y} \in \mathcal{Y}, \mathbf{x}(\cdot)} \mathbf{a}^\top \mathbf{y} + \max_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[\mathbf{b}^\top \mathbf{x}(\mathbf{w}_{[t]})] \quad (5.55a)$$

$$s.t. \mathbf{T}\mathbf{y} + \mathbf{W}\mathbf{x}(\mathbf{w}_{[t]}) \geq \mathbf{h} - \mathbf{H}\mathbf{w} \quad (5.55b)$$

where $\mathbf{x}(\mathbf{w}_{[t]})$ represents a function of $\mathbf{w}_{[t]}$. Generally, the distributionally robust multi-period problem is complex and intractable, and the solution method will be introduced in next subsection.

5.3.2 Solution Methodology

In this subsection, the ambiguity set for wind power is first designed to describe its possible probability distribution, then linear decision rule approach is introduced to approximate the multi-period problem, and the intractable distributionally robust multi-period problem is finally reformulated as a tractable problem.

5.3.2.1 Ambiguity Set for Wind Power

In DRO method, an ambiguity set is used to capture all possible probability distributions of random variables sharing common statistical characteristics such as moment information. In this work, a new ambiguity set based on moment information of wind power [206] is also designed. More specifically, by defining $\mathbf{w}_{[t]} = (\mathbf{w}_1, \dots, \mathbf{w}_t)$ and $\mathbf{w}_t = (w_{nt})$, the studied ambiguity set is given below:

$$\mathcal{D} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{N^w T}) \left| \begin{array}{l} \mathbb{P}(\mathbf{w} \in \mathcal{W}) = 1, \\ E_{\mathbb{P}}(\mathbf{w}) = \boldsymbol{\mu}, \\ E_{\mathbb{P}}((w_{nt} - \mu_{nt})^2) \leq \sigma_{nt}, \forall n, t, \\ E_{\mathbb{P}}[(\sum_{l=k}^t \mathbf{1}'(\mathbf{w}_l - \boldsymbol{\mu}_l))^2] \leq \gamma_{kt}, \\ \forall k \leq t, t \in [T] \end{array} \right. \right\} \quad (5.56)$$

where $\mathcal{P}_0(\cdot)$ is the set of all distributions, \mathcal{W} is support set defined as $\mathcal{W} = [\underline{\mathbf{w}}, \overline{\mathbf{w}}]$, $\boldsymbol{\mu}$ is the estimated mean vector of wind power, σ_{nt} and γ_{kt} are parameters related with variance which can be used to adjust the conservatism. The parameters in this set can be estimated from historical wind power data, i.e., the set can be constructed in a data-driven manner.

There are two main features in the ambiguity set introduced above. First, the partial cross-moment information is included which helps capture both the temporal correlation and spatial correlation of wind power [207]. Second, this set is a second-order conic representable set and the corresponding DRO problem can be transformed into a second-order conic program which can be solved by many off-the-shelf solvers. To obtain a tractable DRO problem with the set (5.56), the following lifted ambiguity set is proposed [206] by introducing auxiliary variables which keeps the optimal solution equivalent:

$$\overline{\mathcal{D}} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{N^w T} \times \mathbb{R}^{N^v T} \times \mathbb{R}^{T(T+1)/2}) \left| \begin{array}{l} \mathbb{P}((\mathbf{w}, \mathbf{u}, \mathbf{v}) \in \overline{\mathcal{W}}) = 1, \\ E_{\mathbb{P}}(\mathbf{w}) = \boldsymbol{\mu}, \\ E_{\mathbb{P}}(u_{nt}) \leq \sigma_{nt}, \forall n, t, \\ E_{\mathbb{P}}(v_{kt}) \leq \gamma_{kt}, \forall k \leq t, t \in [T] \end{array} \right. \right\} \quad (5.57)$$

where \mathbf{u} and \mathbf{v} are auxiliary variables, and $\overline{\mathcal{W}}$ is the lifted support set defined as below:

$$\overline{\mathcal{W}} = \left\{ (\mathbf{w}, \mathbf{u}, \mathbf{v}) \left| \begin{array}{l} \underline{\mathbf{w}} \leq \mathbf{w} \leq \overline{\mathbf{w}}, \\ (w_{nt} - \mu_{nt})^2 \leq u_{nt}, \forall n, t, \\ [\sum_{l=k}^t \mathbf{1}'(\mathbf{w}_l - \boldsymbol{\mu}_l)]^2 \leq v_{kt}, \\ \forall k \leq t, t \in [T] \end{array} \right. \right\}. \quad (5.58)$$

It is equivalent to deal with the DRO problem with the lifted ambiguity set since the original set \mathcal{D} is equivalent to the set of marginal distributions of \mathbf{w} under all $\mathbb{P} \in \overline{\mathcal{D}}$. In addition, we can further design a tighter lifted support set $\tilde{\mathcal{W}}$ by incorporating the upper bounds of \mathbf{u} and \mathbf{v} which can significantly improve the performance of the optimal solution, and the improvement is verified in case studies. The set $\tilde{\mathcal{W}}$ is expressed as follows:

$$\tilde{\mathcal{W}} = \{\overline{\mathcal{W}}, u_{nt} \leq \bar{u}_{nt}, v_{kt} \leq \bar{v}_{kt}, \forall n, t, k \leq t\} \quad (5.59)$$

where the upper bounds can be obtained as: $\bar{u}_{nt} = \max\{(w_{nt} - \mu_{nt})^2, (\bar{w}_{nt} - \mu_{nt})^2\}$ and $\bar{v}_{kt} = \max\{(\sum_{l=k}^t \mathbf{1}'(\underline{\mathbf{w}}_l - \boldsymbol{\mu}_l))^2, (\sum_{l=k}^t \mathbf{1}'(\overline{\mathbf{w}}_l - \boldsymbol{\mu}_l))^2\}$.

5.3.2.2 Linear Decision Rule

The multi-period problem (5.55) is computationally challenging since the recourse variable \mathbf{x} is a function of all past uncertainty realizations. In addition, the explicit expression of the recourse policy and worst-case expectation are generally intractable to acquire. One effective approach to solve this multi-period problem is linear decision rule (LDR) method

which is also known as affine decision rule [79] [208]. LDR method enforces the recourse variable to be linearly dependent on some random variables to overcome the intractability. Actually, the LDR method depending on all past uncertainty realizations prior to time t still makes the problem very computationally difficult, thus a simplified LDR method is adopted in this work which can achieve sufficiently good results [78]. In the simplified LDR method, the recourse variable is assumed to be a linear function of the uncertain parameters at current time period. Particularly, the LDR method for a single recourse variable can be expressed as follows:

$$x_t(\mathbf{w}_{[t]}, \mathbf{u}_{[t]}) = x_t^0 + \sum_n x_{nt}^w w_{nt} + \sum_n x_{nt}^u u_{nt} \quad (5.60)$$

where x_t^0 is a constant, x_{nt}^w and x_{nt}^u are related linear coefficients which will be considered as decision variables in the new problem. In addition, the auxiliary variable u_{nt} is also included in this LDR method since this enhanced LDR method can improve the results as shown in [206]. Considering the variables u_{nt} and v_{kt} are both related with the second-order moment information, v_{kt} is neglected here to reduce the number of decision variables.

Based on (5.60), we can write the LDR method for all recourse variables in a matrix form as follows:

$$\mathbf{x}(\mathbf{w}, \mathbf{u}) = \mathbf{x}^0 + \mathbf{X}^w \mathbf{w} + \mathbf{X}^u \mathbf{u} \quad (5.61)$$

where \mathbf{x}^0 denotes the constant vector, \mathbf{X}^w and \mathbf{X}^u are coefficients matrices. With LDR method, the non-anticipativity is automatically included and we can obtain a tractable problem for the complex multi-period problem.

5.3.2.3 Problem Reformulation

To solve the proposed two-stage multi-period distributionally robust problem for CCHP microgrid system, we need to reformulate it into a tractable problem. First, we need to deal with the second-stage worst-case expectation $\max_{\mathbb{P} \in \overline{\mathcal{D}}} E_{\mathbb{P}}[\mathbf{b}^\top \mathbf{x}]$ in the two-stage framework to reduce the computational burden [209]. The worst-case expectation can be written in an integral form as below:

$$\max_{\mathbb{P} \in \overline{\mathcal{D}}} \int_{\overline{\mathcal{W}}} (\mathbf{b}^\top \mathbf{x}) dP(\mathbf{w}, \mathbf{u}, \mathbf{v}) \quad (5.62a)$$

$$s.t. \int_{\overline{\mathcal{W}}} dP(\mathbf{w}, \mathbf{u}, \mathbf{v}) = 1 : (\lambda) \quad (5.62b)$$

$$\int_{\overline{\mathcal{W}}} \mathbf{w} dP(\mathbf{w}, \mathbf{u}, \mathbf{v}) = \boldsymbol{\mu} : (\boldsymbol{\eta}) \quad (5.62c)$$

$$\int_{\overline{\mathcal{W}}} u_{nt} dP(\mathbf{w}, \mathbf{u}, \mathbf{v}) \leq \sigma_{nt}, \forall n, t : (\beta_{nt}) \quad (5.62d)$$

$$\int_{\overline{\mathcal{W}}} v_{kt} dP(\mathbf{w}, \mathbf{u}, \mathbf{v}) \leq \gamma_{kt}, \forall k \leq t : (\alpha_{kt}) \quad (5.62e)$$

where the symbols in the parenthesis are related dual variables that are used later. According to the strong duality, the equivalent dual problem of (5.62) is given as follows [206] [209]:

$$\min_{\lambda, \boldsymbol{\eta}, \boldsymbol{\beta} \geq 0, \boldsymbol{\alpha} \geq 0} \lambda + \boldsymbol{\eta}^\top \boldsymbol{\mu} + \boldsymbol{\beta}^\top \boldsymbol{\sigma} + \boldsymbol{\alpha}^\top \boldsymbol{\gamma} \quad (5.63a)$$

$$s.t. \lambda + \boldsymbol{\eta}^\top \mathbf{w} + \boldsymbol{\beta}^\top \mathbf{u} + \boldsymbol{\alpha}^\top \mathbf{v} \geq \mathbf{b}^\top \mathbf{x}, \forall (\mathbf{w}, \mathbf{u}, \mathbf{v}) \in \overline{\mathcal{W}} \quad (5.63b)$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are corresponding dual vectors composed of β_{nt} and α_{nt} , respectively.

By combining the first-stage problem with (5.63) and considering the LDR method in (5.61), we can get the following equivalent formulation of the two-stage multi-period problem (5.55):

$$\min_{\mathbf{y} \in \mathcal{Y}} \mathbf{a}^\top \mathbf{y} + \min_{\lambda, \boldsymbol{\eta}, \boldsymbol{\beta} \geq 0, \boldsymbol{\alpha} \geq 0} \lambda + \boldsymbol{\eta}^\top \boldsymbol{\mu} + \boldsymbol{\beta}^\top \boldsymbol{\sigma} + \boldsymbol{\alpha}^\top \boldsymbol{\gamma} \quad (5.64a)$$

$$s.t. \lambda + \boldsymbol{\eta}^\top \mathbf{w} + \boldsymbol{\beta}^\top \mathbf{u} + \boldsymbol{\alpha}^\top \mathbf{v} \geq$$

$$\mathbf{b}^\top (\mathbf{x}^0 + \mathbf{X}^w \mathbf{w} + \mathbf{X}^u \mathbf{u}), \forall (\mathbf{w}, \mathbf{u}, \mathbf{v}) \in \overline{\mathcal{W}} \quad (5.64b)$$

$$\mathbf{T} \mathbf{y} + \mathbf{W} (\mathbf{x}^0 + \mathbf{X}^w \mathbf{w} + \mathbf{X}^u \mathbf{u}) \geq \mathbf{h} - \mathbf{H} \mathbf{w}. \quad (5.64c)$$

The constraints (5.64b) and (5.64c) can be further recast equivalently as follows:

$$\begin{aligned} \lambda - \mathbf{b}^\top \mathbf{x}^0 \geq \max_{(\mathbf{w}, \mathbf{u}, \mathbf{v}) \in \overline{\mathcal{W}}} & [(\mathbf{b}^\top \mathbf{X}^w)' - \boldsymbol{\eta}]^\top \mathbf{w} \\ & + [(\mathbf{b}^\top \mathbf{X}^u)' - \boldsymbol{\beta}]^\top \mathbf{u} - \boldsymbol{\alpha}^\top \mathbf{v} \end{aligned} \quad (5.65a)$$

$$\mathbf{T} \mathbf{y} - \mathbf{h} + \mathbf{W} \mathbf{x}_0 \geq \max_{(\mathbf{w}, \mathbf{u}, \mathbf{v}) \in \overline{\mathcal{W}}} (-\mathbf{W} \mathbf{X}^w - \mathbf{H}) \mathbf{w} - \mathbf{W} \mathbf{X}^u \mathbf{u}. \quad (5.65b)$$

Note that the above two constraints are actually robust constraints with the support set $\overline{\mathcal{W}}$ and they have the same structure. To eliminate the max operator in the right-hand side, we can transform the maximization problem into a minimization problem based on dual theory, and the minimization problem is equivalent to the existence of a feasible solution where the min operator can be neglected. Take constraint (5.65a) as an example, and we can get the dual problem of the right maximization based on conic duality [205] [210] (see Appendix B) as follows:

$$\begin{aligned} \min_{\boldsymbol{\psi}} & \overline{\mathbf{w}}^\top \tilde{\boldsymbol{\delta}} - \underline{\mathbf{w}}^\top \tilde{\boldsymbol{\delta}} - \boldsymbol{\mu}^\top \tilde{\boldsymbol{\theta}} - \frac{1}{2} \mathbf{1}^\top \tilde{\boldsymbol{\theta}} + \frac{1}{2} \mathbf{1}^\top \hat{\boldsymbol{\theta}} \\ & - \sum_t \sum_{k \in [1:t]} \sum_{l=k}^t \mathbf{1}^\top \boldsymbol{\mu}_l \bar{\rho}_{kt} - \frac{1}{2} \mathbf{1}^\top \tilde{\boldsymbol{\rho}} + \frac{1}{2} \mathbf{1}^\top \hat{\boldsymbol{\rho}} \end{aligned} \quad (5.66a)$$

$$\tilde{\delta}_{nt} - \bar{\delta}_{nt} + \bar{\theta}_{nt} + \sum_{k=1}^t \sum_{l=t}^T \bar{\rho}_{kl} = \mathbf{e}_{nt}^\top (\boldsymbol{\eta} - (\mathbf{b}^\top \mathbf{X}^w)'), \forall t, n \quad (5.66b)$$

$$(\tilde{\theta}_{nt} + \hat{\theta}_{nt})/2 = \mathbf{e}_{nt}^\top [\boldsymbol{\beta} - (\mathbf{b}^\top \mathbf{X}^u)'], \forall t, n \quad (5.66c)$$

$$(\tilde{\rho}_{kt} + \hat{\rho}_{kt})/2 = \alpha_{kt}, \forall t, k \leq t \quad (5.66d)$$

$$\sqrt{(\bar{\theta}_{nt}^2 + \tilde{\theta}_{nt}^2)} \leq \hat{\theta}_{nt}, \forall t, n \quad (5.66e)$$

$$\sqrt{(\bar{\rho}_{kt}^2 + \tilde{\rho}_{kt}^2)} \leq \hat{\rho}_{kt}, \forall t, k \leq t \quad (5.66f)$$

where $\psi = \{\bar{\delta}, \tilde{\delta}, \bar{\theta}, \tilde{\theta}, \hat{\theta}, \bar{\rho}, \tilde{\rho}, \hat{\rho}\}$ are dual variables corresponding to the constraints in $\overline{\mathcal{W}}$, $\mathbf{1}$ is a vector with all 1 elements, and e_{nt} is a zero vector except that the $(2(t-1) + n)$ th element is 1. Thus, the constraint (5.65a) is recast as follows:

$$\begin{aligned} \lambda - \mathbf{b}^\top \mathbf{x}^0 &\geq \bar{\mathbf{w}}^\top \bar{\boldsymbol{\delta}} - \underline{\mathbf{w}}^\top \tilde{\boldsymbol{\delta}} - \boldsymbol{\mu}^\top \bar{\boldsymbol{\theta}} - \frac{1}{2} \mathbf{1}^\top \tilde{\boldsymbol{\theta}} + \frac{1}{2} \mathbf{1}^\top \hat{\boldsymbol{\theta}} \\ &\quad - \sum_t \sum_{k \in [1:t]} \sum_{l=k}^t \mathbf{1}^\top \boldsymbol{\mu}_l \bar{\rho}_{kt} - \frac{1}{2} \mathbf{1}^\top \tilde{\boldsymbol{\rho}} + \frac{1}{2} \mathbf{1}^\top \hat{\boldsymbol{\rho}} \end{aligned} \quad (5.67a)$$

$$(5.66b) - (5.66f). \quad (5.67b)$$

Since constraint (5.65b) has the same structure with (5.65a), similar approach can be applied to deal with (5.65b) by introducing new dual variables and replacing the right-hand side of the above dual constraints with the elements of the coefficients matrices of $(\mathbf{w}, \mathbf{u}, \mathbf{v})$ in (5.65b), and the detailed formulation is omitted here. Therefore, constraints (5.65a) and (5.65b) are transformed into a finite number of linear and second-order conic constraints, and the original two-stage multi-period problem is finally reformulated as a tractable mixed-integer second-order conic program (MISOCP) which is actually a single minimization problem and can be solved by some off-the-shelf solvers.

5.3.3 Case Studies

In this subsection, case studies are conducted to validate the performance of the proposed approach. Related parameters are first set and historical wind power data are collected to construct the ambiguity set. Then the simulation results and comparison with other methods are presented. All the experiments are implemented in Matlab environment solved by the MOSEK solver [211] on a personal computer (Intel Core i7-6700 CPU 3.4 GHz and 8GB RAM).

5.3.3.1 Data and Parameter Settings

In this work, a CCHP microgrid composed of three MTs, a gas furnace, two wind farms, one energy storage and thermal storage system, and electric and thermal loads which are all in a single bus system is considered. The optimization horizon is $T = 24$ hours with the scale of one hour. Based on the hourly wind power data of December, 2018 from [212], we can estimate the mean, upper and lower bounds used in the ambiguity set as shown in Fig. 5.10 which are properly scaled. It is assumed that two wind farms have the same power profile for simplicity. In addition, the other parameters in the ambiguity

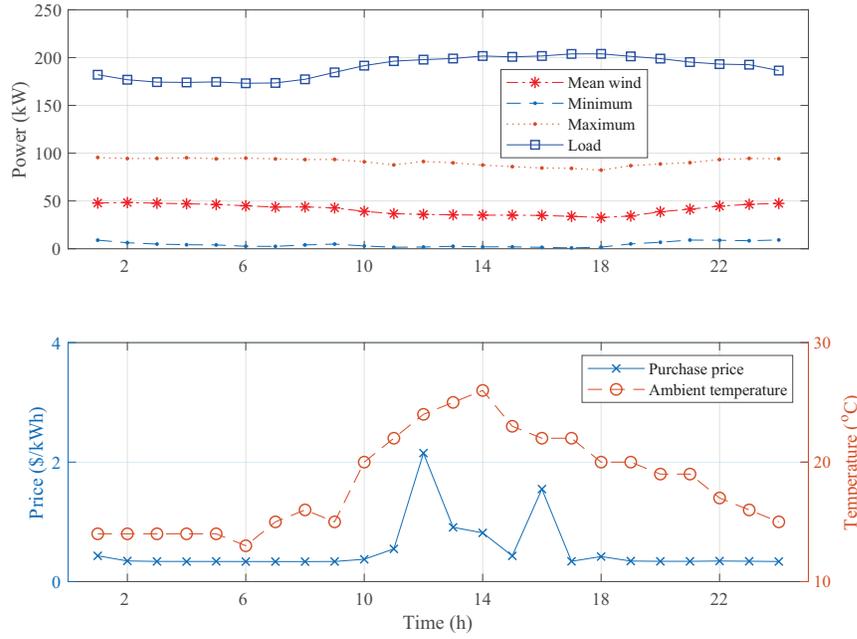


Figure 5.10: Data profiles.

Table 5.4: Parameters of MTs

Unit	\underline{P}_i	\bar{P}_i	DT_i	UT_i	R_i^{up}/R_i^{dn}	η_i^{MT}	$\eta_{i,loss}^{MT}$
MT1	5	20	1	1	10	0.295	0.115
MT2	5	50	2	2	25	0.285	0.15
MT3	50	100	2	3	50	0.3	0.1

set can also be estimated from the data with covariance matrix. The scaled electric load and purchase price are collected from AESO [213], the forecast ambient temperature of Edmonton on July 16 is used here which is related to the thermal load, and they are all shown in Fig. 5.10. In addition, the price of selling electricity to the main grid is assumed to be 0.8 of the purchase price.

The main parameters of MTs are given in Table 5.4 which are collected from relevant references [112] [205]. In addition, the start-up (shut-down) costs of three MTs are 3, 3, 1.5, respectively, and the no-load operation costs are 3, 6, 1, respectively. The rest parameter values about the gas furnace, storage systems and other constraints used in this work are listed in Table 5.5.

5.3.3.2 Simulation Results

With the data and parameter settings introduced above, we can solve the two-stage multi-period CCHP microgrid energy management problem. First, we can get the unit commitment (UC) decisions of three MTs, and the results are presented in Table 5.6. From

Table 5.5: Main parameter values

Parameters	Value	Parameters	Value
η_j^{e+}/η_j^{e-}	0.95	E_{m0}^q	90 kWh
η_m^{q+}/η_m^{q-}	0.9	$\underline{E}_m^q/\overline{E}_m^q$	20/180 kWh
η_g^{GF}	0.93	$\underline{h}_g^{GF}/\overline{h}_g^{GF}$	0/80 kW
$\theta_t^{in}/\overline{\theta}_t^{in}$	18/22 °C	$\overline{p}^{AC}/\overline{p}^{HC}$	200 kW
c^{air}	1.85 kWh/°C	Q^G	9.78 kWh/m ³
c^{gas}	0.5 \$/m ³	$\underline{r}_j^{e+}/\overline{r}_j^{e+}$	0/100 kW
c_j^{ess}/c_m^q	0.0035 \$/kWh	$\underline{r}_j^{e-}/\overline{r}_j^{e-}$	0/100 kW
COP^C/COP^H	0.83/0.8	$\underline{r}_m^{q+}/\overline{r}_m^{q+}$	0/100 kW
E_{j0}^e	90 kWh	$\underline{r}_m^{q-}/\overline{r}_m^{q-}$	0/100 kW
$\underline{E}_j^e/\overline{E}_j^e$	20/180 kWh	R^{tr}	1.3 °C/kW

Table 5.6: UC results of MTs

Hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
MT1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0
MT2	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0
MT3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

this table, it can be seen that MT3 is always on to supply power since it has the lowest generation cost, while MT1 and MT2 are started up at time period 11 when the load increases. The total cost for this case is \$115.1317, and the solution time is about 38 min.

To further verify the performance of LDR method, the second-stage recourse variables are investigated for a realized wind power data. More specifically, the lower bounds of wind power data are used, and the output of MTs and the ESS storage level, the electricity transaction with main grid and indoor temperature settings are illustrated in Fig. 5.11. Note that since there are not many so thermal loads in this system, the output of gas furnace is close to zero, and the TSS is almost not used, hence, their output are not shown here. From Fig. 5.11, we can see that three MTs approximately generate the maximum output as a result of the low wind power. In addition, the indoor temperature settings are also within the predefined comfortable range.

5.3.3.3 Comparison with Other Methods

In this subsection, the proposed CCHP microgrid energy management problem is further studied with the new support set $\tilde{\mathcal{W}}$ to improve the solution and the proposed approach is compared with other methods to validate its effectiveness. With the set $\tilde{\mathcal{W}}$ which includes upper bounds of auxiliary variables, we can reformulate the problem similarly and the optimal cost achieved is \$78.5489 which enhances the original objective. For comparison purpose, we study the problem with robust optimization (RO) approach and deterministic method. In RO method, the interval uncertainty set is used and the problem is solved with

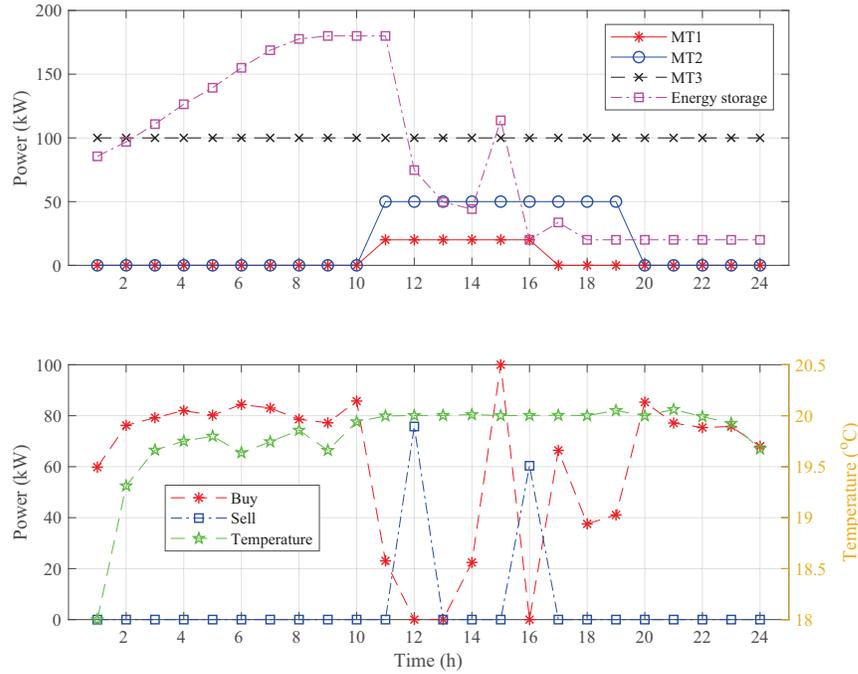


Figure 5.11: Output of recourse variables with LDR.

Table 5.7: Comparison results

Method	DRO	DRO2	RO	Deterministic
Total cost (\$)	115.1317	78.5489	628.0743	-74.2021
UC cost (\$)	108	147	240	126

column and constraint generation (C&CG) approach [61] [188]. For deterministic method, fixed wind power realizations (e.g., mean values) are used and there is no uncertainty. The comparison results are summarized in Table 5.7 including the total cost and first-stage UC cost. In this table, DRO is the proposed method and DRO2 is the DRO method with new support set $\tilde{\mathcal{W}}$. From the total cost, we can find that the proposed DRO method is less conservative than RO method which schedules more units to guard against uncertainty. In addition, the total cost (profits) of deterministic method is better than DRO method. However, there is no uncertainty information in deterministic method and the wind power realizations are assumed to be known in entire horizon. By contrast, the non-anticipativity of multi-period problem is enforced in the proposed DRO method which can deal with the uncertainty from renewable generation.

5.4 Summary

To ensure the non-anticipativity of uncertainty, multistage energy management is studied for grid-connected microgrids in this chapter, and RO and DRO uncertainty modeling

method are investigated, respectively.

First, a novel multistage robust energy management model for grid-connected microgrids is developed in this chapter, which considers the uncertainty of renewable generation and load demand. Compared with traditional two-stage models, the multistage model can deal with the non-anticipativity of uncertainty. To address the computationally difficult multistage problem, a robust version of dual dynamic programming named the RDDP method is proposed which combines the outer approximation and inner approximation in the forward and backward pass. A case study with real datasets is conducted and the experimental results verify the effectiveness of the method in achieving robust energy management. In addition, the comparison with other methods including SDDP, SAA, and deterministic method also demonstrates the advantage of the proposed multistage model and solution method with respect to the robust optimal solution.

In addition, distributionally robust multi-period energy management for CCHP microgrids is also investigated in this chapter. Different from previous literature, a two-stage multi-period model is proposed which considers the non-anticipativity of dispatch process. To capture the uncertainty of wind power, a second-order conic representable ambiguity set is designed based on moment information (e.g., mean and covariance) which can also describe the temporal and spatial correlation of random variable. With LDR method, the complex multi-period problem is finally reformulated as a tractable mixed-integer second-order conic program (MISOCP) problem. In addition, a tight support set with upper bounds of auxiliary variables introduced in the lifted ambiguity set is investigated to further improve the solutions. The performance of the proposed approach is validated by case studies based on real-world data, particularly, the comparison experiments with RO and deterministic methods are conducted to show the effectiveness of the proposed approach.

6

Conclusions and Future Work

Tradition power systems are evolving towards the new smart grid with the development of advanced communication technologies. In addition, the utilization of renewable energy also attracts wide attention to alleviate the energy crisis problem. However, with the increasing penetration of renewable generation, many new challenges with regard to uncertainty have arisen in the reliable and secure operation of smart grid. Therefore, it is necessary and important to study the influence of uncertainty for better decision-making process in power system operation.

To reduce the impact of uncertain and intermittent renewable generation, uncertainty modeling and related optimization problems in smart grid are investigated in this thesis. The uncertainty modeling techniques studied in this work consist of interval forecast method, DRO method and RO method, and the involved problems include wind energy forecast, chance-constrained energy management for islanded microgrids, data-driven distributionally robust UC and ED problem, and multistage energy management for grid-connected microgrids. More specifically, the direct interval prediction for wind power based on RNN and multiobjective interval prediction for wind speed based on WNN are first studied, respectively. Based on the interval forecast, the moment based ambiguity set with an interval form is designed and chance-constrained energy management problem is studied for islanded microgrids. Since UC scheduling is not included in the chance-constrained energy management problem, data-driven distributionally robust UC and ED problem are then investigated with distance-based ambiguity sets. Finally, multistage energy management problem for grid-connected microgrids is studied based on RO and DRO method which considers the non-anticipativity of uncertainty.

6.1 Contributions of Thesis

The main contributions of this thesis can be summarized as follows:

- The RNN model exhibiting the dynamic system behaviors is investigated for wind power interval forecast, and this is the first time to apply RNN model to do interval forecast for wind power. A new evaluation index for the PI width is proposed to enhance the RNN model training; this new index, unlike the previous measures, further considers and uses the known information in the training process. DA is introduced for the first time to solve the PI problem and a new weight update method which combines linear decrease and random walk is designed to improve the algorithm search ability. In addition, delay embedding approach rather than the typical correlation analysis is employed to reconstruct the time series data and determine the input of the prediction model. It is suitable to process wind power data with chaotic characteristics.
- The WNN model is proposed for wind speed interval prediction is a multiobjective framework. Although WNN model has been studied before for point forecast tasks, it is the first time to conduct interval prediction based on WNN model in this study, i.e., this is a new interval forecast method for wind speed. A novel multiobjective evolutionary algorithm PICEA-g is investigated to train the NN model which considers two objectives. Considering the multiobjective essence of PI construction, the proposed multiobjective problem formulation is a more direct problem formulation compared with the indirect single-objective transformation, and this is more reasonable and practical. Case studies are implemented to validate the proposed prediction method based on real-world datasets. More specifically, the proposed model is compared with various single-objective and multiobjective interval prediction models based on the quality of solutions and Pareto front, and comparison results show the efficiency of the proposed approach, that is, the WNN-based multiobjective interval prediction model can achieve better forecast results.
- A CC energy management problem is formulated for an islanded microgrid considering DG units, ESS, renewable generation and various load demand. Unlike the previous chance constraints in the literature, the common power balance is presented in a probabilistic version for islanded microgrids in this work. The objective of the proposed model is to minimize the total system cost including the generation cost and emission cost of DG units and degradation cost of ESS where the consideration of emission cost is necessary in practice with increasing attention of

environment problem. A novel ambiguity set is proposed to describe the uncertain probability distribution of renewable generation. In particular, the box-type ambiguity set is used to capture the uncertain moment information (e.g., mean and variance) of renewable generation. This ambiguity set has not been studied previously for microgrid energy management problem with uncertainty. Based on the ambiguity set, the DRO method is utilized to solve the microgrid energy management problem by transforming the CC problem into a tractable second-order conic programming (SOCP) problem which can be solved by off-the-shelf solvers efficiently. A case study with real datasets to verify the effectiveness of the proposed method is presented. The comparison with the DRO method with known moment information is also carried out to show the robustness of the approach. In addition, sample average approximation (SAA) and stochastic optimization with normal distribution method, which are two common methods to deal with chance constraint, are also applied to solve the problem for comparison purpose.

- A data-driven distributionally robust chance-constrained two-stage UC model is proposed in this work which determines the commitment decision and base-case dispatch plan in the first stage and minimizes the re-dispatch cost due to possible power imbalance in the second stage. Specifically, a chance constraint is used to restrain the power imbalance in the first stage, and the re-dispatch cost resulted from load curtailment or wind power spillage is considered in the second stage. This is a new model by combining the new DRO technique and two-stage chance-constrained model compared with those in previous literature. A new problem reformulation method is proposed with the studied distance-based ambiguity set. Particularly, based on the proposed ambiguity set, the original complicated UC problem is reformulated into a tractable two-stage optimization problem which can be solved in a decomposition framework, i.e., the second-stage objective function is transformed into a convex combination of conditional value-at-risk (CVaR) and worst-case cost. According to the available historical wind power data size, the constructed ambiguity set can be adjusted, thus the conservativeness of the solution can also be altered accordingly. Moreover, there is a marginal diminishing effect for data size in controlling the conservativeness. In addition, the new reformulation method helps explicitly reveal the value of additional data in reducing the conservatism of the problem, and we can flexibly acquire the corresponding stochastic problem and robust problem.
- A data-driven two-stage distributionally robust model is proposed for energy and reserve dispatch problem with wind power, and a new ambiguity set based on L_∞ norm is designed to describe the uncertainty of wind power probability

distribution. The second-stage worst-case expectation is reformulated into a convex combination of CVaR and an expected cost so that the original problem can be solved as a stochastic linear program problem, the effectiveness of the proposed method, especially the value of data, is validated by experiments based on IEEE 6-bus test system and 118-bus test system.

- A new multistage robust energy management model for gridconnected microgrid is developed, which considers the nonanticipative uncertainty from a practical perspective compared with previous two-stage models, and the uncertainty considered comes from renewable generation and load demand. To deal with the computational difficulty of multistage robust model, a novel decomposition method similar to SDDP method, i.e. the robust dual dynamic programming method (RDDP), is proposed to solve the problem. This robust version of dual dynamic programming makes the complex multistage problem computationally tractable to find the worst-case optimal solutions. It decomposes the multistage problem into small-stage problems and tries to approximate the unknown cost-to-go function with a lower and upper bound. A case study with real datasets is carried out to verify the effectiveness of the proposed model and method. In particular, the simulation results are analysed and the comparison with other methods including the common SDDP is discussed.
- A two-stage multi-period energy management model is proposed for CCHP microgrids which considers the non-anticipativity of dispatch decisions, and the DRO technique is adopted as the uncertainty modeling method. In order to capture the uncertain distribution of wind power, a new second-order conic representable ambiguity set is designed, and this moment-based set can describe the temporal and spatial correlation of random variable. In addition, linear decision rule is investigated to help transform the multi-period problem into a tractable problem. In addition, together with the lifted ambiguity set, a tight support set with upper bounds is developed to further improve the solutions.

6.2 Directions for Future Work

In this thesis, several problems in smart grid with renewable generation are studied including interval prediction based on NN model, microgrid energy management, UC and ED based on DRO or RO method. Although good results are achieved in this work, there are still some open issues based on this research, for example, how to further improve the forecast performance, how to design more practical ambiguity sets and models, and so on. Particularly, the following topics are proposed for future work:

- The RNN and WNN model used in the interval forecast adopt a typical three layer structure, more complicated NN model such as deep NN model [214] may be investigated to improve the forecast performance.
- More advanced evolutionary algorithms may be introduced to help train the NN model and enhance the performance. In addition, it is also possible to study long-term interval forecast model for wind power and wind speed.
- For DRO method, some new ambiguity sets may be investigated to capture the uncertainty of renewable generation, e.g., the ambiguity set combining the moment and distance information may be designed.
- Two-stage UC and ED models are very common in the literature. Therefore, a potential research topic is the multistage UC and ED problem considering distributional uncertainty, and the corresponding solution method is also worth studying.
- For microgrid energy management, the economic cost is mostly is considered as the objective. It is possible to develop a multiobjective energy management model [215] in the future which considers multiple objectives such as the cost, pollutant emissions and reliability index.
- Although renewable generation is a very common source of uncertainty, there are other uncertain factors in smart grid including load demand and electricity price. Hence, multiple uncertainties may be considered in related problems for future research.
- Demand response is a critical and effective measure to enhance the interaction between demand side resources and renewable generation in power system [216]. The integration of demand response in current microgrid energy management problem which may reduce the cost by shifting some loads is worth studying in the future.

Bibliography

- [1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid the new and improved power grid: A survey," *IEEE Commun. Surv. Tut.*, vol. 14, no. 4, pp. 944–980, Fourth 2012.
- [2] H. Liang, A. K. Tamang, W. Zhuang, and X. S. Shen, "Stochastic information management in smart grid," *IEEE Commun. Surv. Tut.*, vol. 16, no. 3, pp. 1746–1770, Third 2014.
- [3] N. W. Miller, D. Guru, and K. Clark, "Wind generation," *IEEE Ind. Appl. Mag.*, vol. 15, no. 2, pp. 54–61, March 2009.
- [4] Canadian Wind Power Association (CANWEA). [Online]. Available: <https://canwea.ca/wind-facts/why-wind-works/>, accessed August 30, 2019.
- [5] L. Xie, Y. Gu, X. Zhu, and M. G. Genton, "Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 511–520, Jan 2014.
- [6] S. Fan, L. Chen, and W. Lee, "Short-term load forecasting using comprehensive combination based on multimeteorological information," *IEEE Trans. Industry Appl.*, vol. 45, no. 4, pp. 1460–1466, July 2009.
- [7] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 306–318, Feb 2009.
- [8] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, 2014.
- [9] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Trans. Neural Netw.*, vol. 22, no. 3, pp. 337–346, March 2011.
- [10] S. S. Haykin *et al.*, *Neural networks and learning machines third edition*. Pearson Upper Saddle River, New York: Prentice Hall, USA, 2009.
- [11] F. Katiraei, R. Iravani, N. Hatziargyriou, and A. Dimeas, "Microgrids management," *IEEE Power Energy Mag.*, vol. 6, no. 3, pp. 54–65, May 2008.

- [12] W. Shi, N. Li, C. Chu, and R. Gadh, "Real-time energy management in microgrids," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 228–238, Jan 2017.
- [13] S. F. Rafique and Z. Jianhua, "Energy management system, generation and demand predictors: a review," *IET Gener. Transm. Distrib.*, vol. 12, no. 3, pp. 519–530, 2018.
- [14] H. Scarf, "A min-max solution of an inventory problem in studies in the mathematical theory of inventory and production," *The social cost of foreign exchange reserves*, pp. 201–9, 1958.
- [15] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Oper. Res.*, vol. 58, no. 4-part-1, pp. 902–917, 2010.
- [16] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, 2010.
- [17] Q. P. Zheng, J. Wang, and A. L. Liu, "Stochastic optimization for unit commitment: a review," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1913–1924, July 2015.
- [18] N. P. Padhy, "Unit commitment—a bibliographical survey," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1196–1205, May 2004.
- [19] J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumaili, and V. Miranda, "Wind power forecasting uncertainty and unit commitment," *Appl. Energy*, vol. 88, no. 11, pp. 4014–4023, 2011.
- [20] X. Xia and A. Elaiw, "Optimal dynamic economic dispatch of generation: A review," *Electr. Power Syst. Res.*, vol. 80, no. 8, pp. 975–986, 2010.
- [21] Y. Ding, M. Xie, Q. Wu, and J. stergaard, "Development of energy and reserve pre-dispatch and re-dispatch models for real-time price risk and reliability assessment," *IET Gener. Transm. Distrib.*, vol. 8, no. 7, pp. 1338–1345, 2014.
- [22] A. Street, A. Moreira, and J. M. Arroyo, "Energy and reserve scheduling under a joint generation and transmission security criterion: An adjustable robust optimization approach," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 3–14, Jan 2014.
- [23] H. B. Azad, S. Mekhilef, and V. G. Ganapathy, "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 546–553, April 2014.
- [24] E. B. Ssekulima, M. B. Anwar, A. Al Hinai, and M. S. El Moursi, "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review," *IET Renew. Power Gener.*, vol. 10, no. 7, pp. 885–989, 2016.

- [25] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "Arima-based time series model of stochastic wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–676, May 2010.
- [26] H.-C. Wu and C.-N. Lu, "A data mining approach for spatial modeling in small area load forecast," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 516–521, May 2002.
- [27] E. E. Elattar, "Prediction of wind power based on evolutionary optimised local general regression neural network," *IET Gener. Transm. Distrib.*, vol. 8, no. 5, pp. 916–923, 2014.
- [28] S. Li, D. C. Wunsch, E. A. O'Hair, and M. G. Giesselmann, "Using neural networks to estimate wind turbine power generation," *IEEE Trans. Energy Convers.*, vol. 16, no. 3, pp. 276–282, Sep. 2001.
- [29] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Trans. Ind. Appl.*, vol. 48, no. 3, pp. 1064–1069, May 2012.
- [30] N. Amjady, F. Keynia, and H. Zareipour, "Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization," *IEEE Trans. Sustain. Energy*, vol. 2, no. 3, pp. 265–276, July 2011.
- [31] A. Khosravi and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 849–856, Oct 2013.
- [32] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.
- [33] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004.
- [34] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845–1856, Nov 2010.
- [35] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Optimal prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1166–1174, May 2014.
- [36] D. Sáez, F. Ávila, D. Olivares, C. Cañizares, and L. Marín, "Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 548–556, March 2015.

- [37] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Trans. Neural Netw.*, vol. 22, no. 9, pp. 1341–1356, Sep. 2011.
- [38] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction interval construction and optimization for adaptive neurofuzzy inference systems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 983–988, Oct 2011.
- [39] R. Ak, V. Vitelli, and E. Zio, "An interval-valued neural network approach for uncertainty quantification in short-term wind speed prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2787–2800, Nov 2015.
- [40] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 303–315, Feb 2014.
- [41] S. Qin, F. Liu, J. Wang, and Y. Song, "Interval forecasts of a novelty hybrid model for wind speeds," *Energy Rep.*, vol. 1, pp. 8–16, 2015.
- [42] A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "A new fuzzy-based combined prediction interval for wind power forecasting," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 18–26, Jan 2016.
- [43] M. Khodayar and M. Teshnehlab, "Robust deep neural network for wind speed prediction," in *Proc. 4th Iranian Joint Congr. Fuzzy Intell. Syst. (CFIS)*, Sept 2015, pp. 1–5.
- [44] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770–2779, Dec 2017.
- [45] S. Guo, Y. Li, and S. Xiao, "Wind speed forecasting of genetic neural model based on rough set theory," in *Proc. 5th Int. Conf. Critical Infrastructure (CRIS)*, Sept 2010, pp. 1–6.
- [46] C. M. Huang, Y. C. Huang, K. Y. Huang, S. J. Chen, and S. P. Yang, "Deterministic and probabilistic wind power forecasting using a hybrid method," in *Proc. IEEE ICIT*, March 2017, pp. 400–405.
- [47] R. L. Welch, S. M. Ruffing, and G. K. Venayagamoorthy, "Comparison of feedforward and feedback neural network architectures for short term wind speed prediction," in *Proc. IEEE IJCNN*, 2009, pp. 3335–3340.
- [48] S. Anbazhagan and N. Kumarappan, "Day-ahead deregulated electricity market price forecasting using recurrent neural network," *IEEE Syst. J.*, vol. 7, no. 4, pp. 866–872, Dec 2013.

- [49] O. Özgönenel and D. W. Thomas, "Short-term wind speed estimation based on weather data," *Turk. J. Electr. Eng. & Comp. Sci.*, vol. 20, no. 3, pp. 335–346, 2012.
- [50] C. Wang, J. Wu, J. Wang, and Z. Hu, "Short-term wind speed forecasting using the data processing approach and the support vector machine model optimized by the improved cuckoo search parameter estimation algorithm," *Math. Probl. Eng.*, vol. 2016, 2016.
- [51] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renew. Sustain. Energy Rev.*, vol. 34, pp. 243–254, 2014.
- [52] S.-w. Fei and Y. He, "Wind speed prediction using the hybrid model of wavelet decomposition and artificial bee colony algorithm-based relevance vector machine," *Int. J. Elec. Power*, vol. 73, pp. 625–631, 2015.
- [53] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Trans. Energy Conver.*, vol. 21, no. 1, pp. 273–284, March 2006.
- [54] J. Wang, W. Zhang, Y. Li, J. Wang, and Z. Dang, "Forecasting wind speed using empirical mode decomposition and elman neural network," *Appl. Soft Comput.*, vol. 23, pp. 452–459, 2014.
- [55] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, April 2002.
- [56] C. Zhang, H. Wei, L. Xie, Y. Shen, and K. Zhang, "Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework," *Neurocomputing*, vol. 205, pp. 53–63, 2016.
- [57] N. A. Shrivastava, K. Lohia, and B. K. Panigrahi, "A multiobjective framework for wind speed prediction interval forecasts," *Renew. Energy*, vol. 87, pp. 903–910, 2016.
- [58] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Rev.*, vol. 53, no. 3, pp. 464–501, 2011.
- [59] Y. Zhang, N. Gatsis, and G. B. Giannakis, "Robust energy management for microgrids with high-penetration renewables," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 944–953, Oct 2013.
- [60] Y. Xiang, J. Liu, and Y. Liu, "Robust energy management of microgrid with uncertain renewable generation and load," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 1034–1043, March 2016.

- [61] Y. Guo and C. Zhao, "Islanding-aware robust energy management for microgrids," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1301–1309, March 2018.
- [62] W. Su, J. Wang, and J. Roh, "Stochastic energy scheduling in microgrids with intermittent renewable energy resources," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1876–1883, July 2014.
- [63] D. T. Nguyen and L. B. Le, "Optimal bidding strategy for microgrids considering renewable energy and building thermal dynamics," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1608–1620, July 2014.
- [64] Y. Cai, G. Huang, Z. Yang, Q. Lin, and Q. Tan, "Community-scale renewable energy systems planning under uncertainty an interval chance-constrained programming approach," *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 721–735, 2009.
- [65] J. Liu, H. Chen, W. Zhang, B. Yurkovich, and G. Rizzoni, "Energy management problems under uncertainties for grid-connected microgrids: A chance constrained programming approach," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2585–2596, Nov 2017.
- [66] Z. Wu, W. Gu, R. Wang, X. Yuan, and W. Liu, "Economic optimal schedule of chp microgrid system using chance constrained programming and particle swarm optimization," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2011, pp. 1–11.
- [67] Y. Huang, L. Wang, W. Guo, Q. Kang, and Q. Wu, "Chance constrained optimization in a home energy management system," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 252–260, Jan 2018.
- [68] H. Yu, C. Y. Chung, K. P. Wong, and J. H. Zhang, "A chance constrained transmission network expansion planning method with consideration of load and wind farm uncertainties," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1568–1576, Aug 2009.
- [69] B. V. Solanki, A. Raghurajan, K. Bhattacharya, and C. A. Canizares, "Including smart loads for optimal demand response in integrated energy management systems for isolated microgrids," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1739–1748, July 2017.
- [70] D. E. Olivares, J. D. Lara, C. A. Cañizares, and M. Kazerani, "Stochastic-predictive energy management system for isolated microgrids," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2681–2693, Nov 2015.
- [71] M. I. Alizadeh, M. P. Moghaddam, and N. Amjady, "Multistage multiresolution robust unit commitment with nondeterministic flexible ramp considering load and wind variabilities," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 872–883, April 2018.

- [72] S. Takriti, J. R. Birge, and E. Long, "A stochastic model for the unit commitment problem," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1497–1508, Aug 1996.
- [73] J. Wang, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with volatile wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1319–1327, Aug 2008.
- [74] L. Wu, M. Shahidehpour, and T. Li, "Stochastic security-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 800–811, May 2007.
- [75] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 800–810, May 2012.
- [76] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, Feb 2013.
- [77] C. Shao, X. Wang, M. Shahidehpour, X. Wang, and B. Wang, "Security-constrained unit commitment with flexible uncertainty set for variable wind power," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1237–1246, July 2017.
- [78] A. Lorca and X. A. Sun, "Multistage robust unit commitment with dynamic uncertainty sets and energy storage," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 1678–1688, May 2017.
- [79] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 39–49, Jan 2017.
- [80] C. Zhao and R. Jiang, "Distributionally robust contingency-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 94–102, Jan 2018.
- [81] W. Wei, F. Liu, and S. Mei, "Distributionally robust co-optimization of energy and reserve dispatch," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 289–300, Jan 2016.
- [82] Y. Zhang, S. Shen, and J. L. Mathieu, "Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1378–1388, March 2017.
- [83] F. Alismail, P. Xiong, and C. Singh, "Optimal wind farm allocation in multi-area power systems using distributionally robust optimization approach," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 536–544, Jan 2018.

- [84] X. Chen, W. Wu, B. Zhang, and C. Lin, "Data-driven dg capacity assessment method for active distribution networks," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3946–3957, Sept 2017.
- [85] H. Yang, S. Zhang, D. Qiu, J. Zhao, M. Lai, Z. Y. Dong, and Z. H. Dong, "Distributionally robust optimal bidding of controllable load aggregators in the electricity market," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1089–1091, Jan 2018.
- [86] F. Qiu and J. Wang, "Distributionally robust congestion management with dynamic line ratings," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2198–2199, July 2015.
- [87] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Math. Program.*, vol. 171, no. 1-2, pp. 115–166, 2018.
- [88] Y. Chen, Q. Guo, H. Sun, Z. Li, W. Wu, and Z. Li, "A distributionally robust optimization model for unit commitment based on kullback–leibler divergence," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5147–5160, Sept 2018.
- [89] C. Duan, L. Jiang, W. Fang, and J. Liu, "Data-driven affinely adjustable distributionally robust unit commitment," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1385–1398, March 2018.
- [90] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2587–2596, July 2016.
- [91] A. Bagheri, C. Zhao, and Y. Guo, "Data-driven chance-constrained stochastic unit commitment under wind power uncertainty," in *Proc. IEEE Power Energy Soc. Gen. Meeting, Chicago, USA*, July 2017, pp. 1–5.
- [92] T. Ding, Q. Yang, X. Liu, C. Huang, Y. Yang, M. Wang, and F. Blaabjerg, "Duality-free decomposition based data-driven stochastic security-constrained unit commitment," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 82–93, Jan 2019.
- [93] R. Zhu, H. Wei, and X. Bai, "Wasserstein metric based distributionally robust approximate framework for unit commitment," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2991–3001, July 2019.
- [94] W. Hou, R. Zhu, H. Wei, and H. TranHoang, "Data-driven affinely adjustable distributionally robust framework for unit commitment based on wasserstein metric," *IET Gener. Transm. Distrib.*, vol. 13, no. 6, pp. 890–895, 2018.
- [95] L. Yao, X. Wang, C. Duan, J. Guo, X. Wu, and Y. Zhang, "Data-driven distributionally robust reserve and energy scheduling over wasserstein balls," *IET Gener. Trans. Distrib.*, vol. 12, no. 1, pp. 178–189, 2018.

- [96] H. Gangammanavar, S. Sen, and V. M. Zavala, "Stochastic optimization of sub-hourly economic dispatch with wind energy," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 949–959, March 2016.
- [97] L. Yang, M. He, V. Vittal, and J. Zhang, "Stochastic optimization-based economic dispatch and interruptible load management with increased wind penetration," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 730–739, March 2016.
- [98] A. Baringo, L. Baringo, and J. M. Arroyo, "Day-ahead self-scheduling of a virtual power plant in energy and reserve electricity markets under uncertainty," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1881–1894, May 2019.
- [99] X. Lu, K. W. Chan, S. Xia, B. Zhou, and X. Luo, "Security-constrained multiperiod economic dispatch with renewable energy utilizing distributionally robust optimization," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 768–779, April 2019.
- [100] Z. Wang, Q. Bian, H. Xin, and D. Gan, "A distributionally robust co-ordinated reserve scheduling model considering cvar-based wind power reserve requirements," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 625–636, April 2016.
- [101] C. Yue, W. Wei, L. Feng, and S. Mei, "Distributionally robust hydro-thermal-wind economic dispatch," *Appl. Energy*, vol. 173, pp. 511–519, 2016.
- [102] X. Xu, Z. Yan, M. Shahidehpour, Z. Li, M. Yan, and X. Kong, "Data-driven risk-averse two-stage optimal stochastic scheduling of energy and reserve with correlated wind power," *IEEE Trans. Sustain. Energy*, pp. 1–1, 2019.
- [103] S. Talari, M. Yazdaninejad, and M. Haghifam, "Stochastic-based scheduling of the microgrid operation including wind turbines, photovoltaic cells, energy storages and responsive loads," *IET Gener. Transm. Distrib.*, vol. 9, no. 12, pp. 1498–1509, 2015.
- [104] M. Vahedipour-Dahraie, H. Rashidizadeh-Kermani, H. R. Najafi, A. Anvari-Moghaddam, and J. M. Guerrero, "Stochastic security and risk-constrained scheduling for an autonomous microgrid with demand response and renewable energy resources," *IET Renew. Power Gener.*, vol. 11, no. 14, pp. 1812–1821, 2017.
- [105] J. Liu, M. Hu, X. Dou, W. Xu, J. Yang, and X. Qiu, "Micro grid energy management based on two-stage robust optimization," in *2016 19th Int. Conf. Electr. Mach. Syst. (ICEMS)*, 2016, pp. 1–5.
- [106] M. V. Pereira and L. M. Pinto, "Multi-stage stochastic optimization applied to energy planning," *Math. Program.*, vol. 52, no. 1-3, pp. 359–375, 1991.

- [107] A. Papavasiliou, Y. Mou, L. Cambier, and D. Scieur, "Application of stochastic dual dynamic programming to the real-time dispatch of storage under renewable supply uncertainty," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 547–558, April 2018.
- [108] A. Bhattacharya, J. P. Kharoufeh, and B. Zeng, "Managing energy storage in microgrids: A multistage stochastic programming approach," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 483–496, Jan 2018.
- [109] P. Fatouros, I. Konstantelos, D. Papadaskalopoulos, and G. Strbac, "Stochastic dual dynamic programming for operation of der aggregators under multi-dimensional uncertainty," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 459–469, Jan 2019.
- [110] W. Gu, Z. Wang, Z. Wu, Z. Luo, Y. Tang, and J. Wang, "An online optimal dispatch schedule for cchp microgrids based on model predictive control," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2332–2342, Sept 2017.
- [111] M. Liu, Y. Shi, and F. Fang, "Combined cooling, heating and power systems: A survey," *Renew. Sustain. Energy Rev.*, vol. 35, pp. 1–22, 2014.
- [112] J. Wang, H. Zhong, Q. Xia, C. Kang, and E. Du, "Optimal joint-dispatch of energy and reserve for cchp-based microgrids," *IET Gener., Transm. Distrib.*, vol. 11, no. 3, pp. 785–794, 2017.
- [113] F. Fang, Q. H. Wang, and Y. Shi, "A novel optimal operational strategy for the cchp system based on two operating modes," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 1032–1041, May 2012.
- [114] Y. Jiang, J. Xu, Y. Sun, C. Wei, J. Wang, S. Liao, D. Ke, X. Li, J. Yang, and X. Peng, "Coordinated operation of gas-electricity integrated distribution system with multi-cchp and distributed renewable energy sources," *Appl. Energy*, vol. 211, pp. 237–248, 2018.
- [115] G. Li, R. Zhang, T. Jiang, H. Chen, L. Bai, H. Cui, and X. Li, "Optimal dispatch strategy for integrated energy systems with cchp and wind power," *Appl. Energy*, vol. 192, pp. 408–419, 2017.
- [116] M. Alipour, B. Mohammadi-Ivatloo, and K. Zare, "Stochastic scheduling of renewable and chp-based microgrids," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1049–1058, Oct 2015.
- [117] Z. Bao, Q. Zhou, Z. Yang, Q. Yang, L. Xu, and T. Wu, "A multi time-scale and multi energy-type coordinated microgrid scheduling solutionpart i: Model and methodology," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2257–2266, Sep. 2015.

- [118] C. Zhang, Y. Xu, Z. Li, and Z. Y. Dong, "Robustly coordinated operation of a multi-energy microgrid with flexible electric and thermal loads," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2765–2775, May 2019.
- [119] J. Aghaei, V. G. Agelidis, M. Charwand, F. Raeisi, A. Ahmadi, A. E. Nezhad, and A. Heidari, "Optimal robust unit commitment of chp plants in electricity markets using information gap decision theory," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2296–2304, Sept 2017.
- [120] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using pso-tuned support vector machines," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 322–331, April 2015.
- [121] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction intervals for short-term wind farm power generation forecasts," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 602–610, July 2013.
- [122] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.
- [123] M. Ardalani-Farsa and S. Zolfaghari, "Chaotic time series prediction with residual analysis method using hybrid elman–narx neural networks," *Neurocomputing*, vol. 73, no. 13, pp. 2540–2553, 2010.
- [124] S. Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, 2016.
- [125] X.-S. Yang, *Nature-inspired metaheuristic algorithms*. Luniver press, 2010.
- [126] R. C. Eberhart and Y. Shi, "Tracking and optimizing dynamic systems with particle swarms," in *Proc. IEEE CEC*, vol. 1, 2001, pp. 94–100.
- [127] Y. Wang, D. Wu, C. Guo, Q. Wu, W. Qian, and J. Yang, "Short-term wind speed prediction using support vector regression," in *Proc. IEEE Power and Energy Soc. Gen. Meeting*, 2010, pp. 1–6.
- [128] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical systems and turbulence, Warwick 1980*. Springer, 1981, pp. 366–381.
- [129] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The tisean package," *Chaos*, vol. 9, no. 2, pp. 413–435, Jun. 1999.
- [130] Independent Electricity System Operator (IESO), Power Data. [Online]. Available: <http://www.ieso.ca/en/power-data/data-directory>, accessed Apr 20, 2017.

- [131] Canadian Wind Power Association (CANWEA). [Online]. Available: <http://canwea.ca/wind-energy/ontario/>, accessed May 18, 2017.
- [132] Nonlinear Time Series Analysis (TISEAN). Available: <https://www.mpipks-dresden.mpg.de/~tisean/>.
- [133] Y. Hariya, T. Kurihara, T. Shindo, and K. Jin'no, "Lévy flight pso," in *Proc. IEEE CEC*, 2015, pp. 2678–2684.
- [134] D. Yang, P. Jirutitijaroen, and W. M. Walsh, "Hourly solar irradiance time series forecasting using cloud cover index," *Sol. Energy*, vol. 86, no. 12, pp. 3531–3543, 2012.
- [135] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using gaussian processes and numerical weather prediction," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 656–665, 2014.
- [136] R. Koenker and K. Hallock, "Quantile regression: An introduction," *J. Econ. Perspect.*, vol. 15, no. 4, pp. 43–56, 2001.
- [137] A. Blanco, M. Delgado, and M. C. Pegalajar, "A real-coded genetic algorithm for training recurrent neural networks," *Neural networks*, vol. 14, no. 1, pp. 93–105, 2001.
- [138] F. Valencia, J. Collado, D. Sez, and L. G. Marn, "Robust energy management system for a microgrid based on a fuzzy prediction interval model," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1486–1494, May 2016.
- [139] J. Wang, T. Niu, H. Lu, Z. Guo, W. Yang, and P. Du, "An analysis-forecast system for uncertainty modeling of wind speed: A case study of large-scale wind farms," *Appl. Energy*, vol. 211, pp. 492–512, 2018.
- [140] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *J. R. Statist. Soc. B*, vol. 69, no. 2, pp. 243–268, 2007.
- [141] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Netw.*, vol. 3, no. 6, pp. 889–898, 1992.
- [142] N. Pindoriya, S. Singh, and S. Singh, "An adaptive wavelet neural network-based energy price forecasting in electricity markets," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1423–1432, Aug 2008.
- [143] H. Chitsaz, N. Amjady, and H. Zareipour, "Wind power forecast using wavelet neural network trained by improved clonal selection algorithm," *Energy Convers. Manag.*, vol. 89, pp. 588–598, 2015.
- [144] Y. Chen, B. Yang, and J. Dong, "Time-series prediction using a local linear wavelet neural network," *Neurocomputing*, vol. 69, no. 4, pp. 449–465, 2006.

- [145] H. Shao, X. Deng, and F. Cui, "Short-term wind speed forecasting using the wavelet decomposition and adaboost technique in wind farm of east china," *IET Gener. Transm. Distrib.*, vol. 10, no. 11, pp. 2585–2592, 2016.
- [146] S. A. Billings and H.-L. Wei, "A new class of wavelet networks for nonlinear system identification," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 862–874, July 2005.
- [147] A. Mellit, M. Benghaneim, and S. A. Kalogirou, "An adaptive wavelet-network model for forecasting daily total solar-radiation," *Appl. Energy*, vol. 83, no. 7, pp. 705–722, 2006.
- [148] R. Wang, R. C. Purshouse, and P. J. Fleming, "Preference-inspired coevolutionary algorithms for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 17, no. 4, pp. 474–494, Aug 2013.
- [149] R. Purshouse, C. Jalbă, and P. Fleming, "Preference-driven co-evolutionary algorithms show promise for many-objective optimisation," in *Proc. Evolutionary multi-criterion optimization*. Springer, 2011, pp. 136–150.
- [150] Z. Shi, R. Wang, and T. Zhang, "Picea-g using an enhanced fitness assignment method," in *Proc. 2014 IEEE Symp. Comput. Intell. Multi-Criteria Decision-Making (MCDM)*. Orlando, USA, 2014, pp. 72–77.
- [151] R. Wang, R. C. Purshouse, and P. J. Fleming, "Preference-inspired co-evolutionary algorithm using adaptively generated goal vectors," in *Proc. IEEE Congr. Evol. Comput.* Cancun, Mexico, 2013, pp. 916–923.
- [152] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach," *IEEE Trans. Evol. Comput.*, vol. 3, no. 4, pp. 257–271, Nov 1999.
- [153] Canada Weather Stats. Available at: <https://www.weatherstats.ca/>.
- [154] C. M. Fonseca, L. Paquete, and M. López-Ibáñez, "An improved dimension-sweep algorithm for the hypervolume indicator," in *Proc. Congr. Evol. Comput.* Palo Alto, USA, 2006, pp. 1157–1163.
- [155] Z. Shi, H. Liang, and V. Dinavahi, "Direct interval forecast of uncertain wind power based on recurrent neural networks," *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1177–1187, July 2018.
- [156] B. Zhao, Y. Shi, X. Dong, W. Luan, and J. Bornemann, "Short-term operation scheduling in renewable-powered microgrids: A duality-based approach," *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 209–217, Jan 2014.

- [157] A. Abdollahi, M. P. Moghaddam, M. Rashidinejad, and M. K. Sheikh-El-Eslami, "Investigation of economic and environmental-driven demand response measures incorporating uc," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 12–25, March 2012.
- [158] Z. Li, Q. Guo, H. Sun, and J. Wang, "Sufficient conditions for exact relaxation of complementarity constraints for storage-concerned economic dispatch," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1653–1654, March 2016.
- [159] C. Shao, X. Wang, X. Wang, C. Du, C. Dang, and S. Liu, "Cooperative dispatch of wind generation and electric vehicles with battery storage capacity constraints in scuc," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2219–2226, Sept 2014.
- [160] L. Guo, W. Liu, B. Jiao, B. Hong, and C. Wang, "Multi-objective stochastic optimal planning method for stand-alone microgrid system," *IET Gener., Transm. & Distrib.*, vol. 8, no. 7, pp. 1263–1273, 2014.
- [161] A. Das and Z. Ni, "A computationally efficient optimization approach for battery systems in islanded microgrid," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6489–6499, Nov 2018.
- [162] Y. Guan and J. Wang, "Uncertainty sets for robust unit commitment," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1439–1440, May 2014.
- [163] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, "A distributionally robust perspective on uncertainty quantification and chance constrained programming," *Math. Program.*, vol. 151, no. 1, pp. 35–62, 2015.
- [164] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Oper. Res.*, vol. 62, no. 6, pp. 1358–1376, 2014.
- [165] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program.*, pp. 1–32, 2013.
- [166] K.-w. Ding, M.-h. Wang, and N.-j. Huang, "Distributionally robust chance constrained problem under interval distribution information," *Optim. Lett.*, pp. 1–14, 2017.
- [167] K. Natarajan, M. Sim, and J. Uichanco, "Tractable robust expected utility and risk models for portfolio optimization," *Math. Financ.*, vol. 20, no. 4, pp. 695–731, 2010.
- [168] S. Sarykalin, G. Serraino, and S. Uryasev, "Value-at-risk vs. conditional value-at-risk in risk management and optimization," in *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*. Informs, 2008, pp. 270–294.
- [169] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, pp. 21–42, 2000.

- [170] A. Shapiro and A. Kleywegt, "Minimax analysis of stochastic problems," *Optim. Methods Softw.*, vol. 17, no. 3, pp. 523–542, 2002.
- [171] L. Vandenberghe, S. Boyd, and K. Comanor, "Generalized chebyshev bounds via semidefinite programming," *SIAM Rev.*, vol. 49, no. 1, pp. 52–64, 2007.
- [172] T. Summers, J. Warrington, M. Morari, and J. Lygeros, "Stochastic optimal power flow based on conditional value at risk and distributional robustness," *Int. J. Elect. Power Energy Syst.*, vol. 72, pp. 116–125, 2015.
- [173] B. Li, R. Jiang, and J. L. Mathieu, "Distributionally robust risk-constrained optimal power flow using moment and unimodality information," in *Proc. 55th IEEE Conf. Decision Control*, 2016, pp. 2425–2430.
- [174] S. Dharmadhikari and K. Joag-Dev, *Unimodality, convexity, and applications*. Elsevier, 1988.
- [175] B. Stellato, "Data-driven chance constrained optimization," Master's thesis, ETH Zürich, 2014.
- [176] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in matlab," in *Proc. IEEE Int. Symp. CACSD*, 2004, pp. 284–289.
- [177] S. Ahmed and A. Shapiro, "Solving chance-constrained stochastic programs via sampling and integer programming," in *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*. INFORMS, 2008, pp. 261–269.
- [178] B. Pagnoncelli, S. Ahmed, and A. Shapiro, "Sample average approximation method for chance constrained programming: theory and applications," *J. Optim. Theory Appl.*, vol. 142, no. 2, pp. 399–416, 2009.
- [179] Y. An and B. Zeng, "Exploring the modeling capacity of two-stage robust optimization: Variants of robust unit commitment model," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 109–122, Jan 2015.
- [180] C. Lee, C. Liu, S. Mehrotra, and M. Shahidehpour, "Modeling transmission line constraints in two-stage robust unit commitment problem," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1221–1231, May 2014.
- [181] Y. Zhang, X. Han, M. Yang, B. Xu, Y. Zhao, and H. Zhai, "Adaptive robust unit commitment considering distributional uncertainty," *Int. J. Elec. Power*, vol. 104, pp. 635–644, 2019.
- [182] Y. Chen, W. Wei, F. Liu, and S. Mei, "Distributionally robust hydro-thermal-wind economic dispatch," *Appl. Energy*, vol. 173, pp. 511–519, 2016.

- [183] M. Carrión and J. M. Arroyo, "A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1371–1378, Aug 2006.
- [184] A. Soroudi, *Power System Optimization Modeling in GAMS*. Switzerland: Springer, 2017.
- [185] R. Jiang and Y. Guan, "Risk-averse two-stage stochastic program with distributional ambiguity," *Oper. Res.*, vol. 66, no. 5, pp. 1390–1405, 2018.
- [186] Y. Cao, W. Wei, J. Wang, S. Mei, M. Shafie-khah, and J. P. S. Catalao, "Capacity planning of energy hub in multi-carrier energy networks: A data-driven robust stochastic programming approach," *IEEE Trans. Sustain. Energy*, pp. 1–1, 2018.
- [187] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *J. Banking Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.
- [188] B. Zeng and L. Zhao, "Solving two-stage robust optimization problems using a column-and-constraint generation method," *Oper. Res. Lett.*, vol. 41, no. 5, pp. 457–461, 2013.
- [189] IEEE 6-bus test system data. Available: <http://motor.ece.iit.edu/data/JEAS6bus.doc>.
- [190] Independent Electricity System Operator (IESO). Available: <http://www.ieso.ca/en/Power-Data/Data-Directory>, accessed Nov. 5, 2018.
- [191] IEEE 118-bus system data. Available: <http://motor.ece.iit.edu/data/JEAS118bus.doc>.
- [192] Z. Li, W. Wu, J. Wang, B. Zhang, and T. Zheng, "Transmission-constrained unit commitment considering combined electricity and district heating networks," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 480–492, April 2016.
- [193] A. Lorca and X. A. Sun, "Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1702–1713, July 2015.
- [194] C. Ju, P. Wang, L. Goel, and Y. Xu, "A two-layer energy management system for microgrids with hybrid energy storage considering degradation costs," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6047–6057, Nov 2018.
- [195] Z. Shi, H. Liang, S. Huang, and V. Dinavahi, "Distributionally robust chance-constrained energy management for islanded microgrids," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2234–2244, March 2019.

- [196] W. Hu, P. Wang, and H. B. Gooi, "Toward optimal energy management of microgrids via robust two-stage optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1161–1174, March 2018.
- [197] V. L. De Matos, A. B. Philpott, and E. C. Finardi, "Improving the performance of stochastic dual dynamic programming," *J. Comput. Appl. Math.*, vol. 290, pp. 196–208, 2015.
- [198] A. Georghiou, A. Tsoukalas, and W. Wiesemann, "Robust dual dynamic programming," *Oper. Res.*, vol. 67, no. 3, pp. 813–830, 2019.
- [199] A. Philpott, V. de Matos, and E. Finardi, "On solving multistage stochastic programs with coherent risk measures," *Oper. Res.*, vol. 61, no. 4, pp. 957–970, 2013.
- [200] A. B. Philpott and Z. Guan, "On the convergence of stochastic dual dynamic programming and related methods," *Oper. Res. Lett.*, vol. 36, no. 4, pp. 450–455, 2008.
- [201] IESO, "Ieso power data," Available at: <http://www.ieso.ca/en/power-data/data-directory>, accessed June 4, 2018.
- [202] NYISO, "Nyiso pricing data," Available at: http://www.nyiso.com/public/markets_operations/market_data/pricing_data/index.jsp, accessed June 4, 2018.
- [203] A. Shapiro, "Analysis of stochastic dual dynamic programming method," *Eur. J. Oper. Res.*, vol. 209, no. 1, pp. 63–72, 2011.
- [204] A. Soroudi, A. Rabiee, and A. Keane, "Information gap decision theory approach to deal with wind power uncertainty in unit commitment," *Electr Power Syst. Res.*, vol. 145, pp. 137–148, 2017.
- [205] S. Babaei, C. Zhao, and L. Fan, "A data-driven model of virtual power plants in day-ahead unit commitment," *IEEE Trans. Power Syst.*, pp. 1–1, 2018.
- [206] D. Bertsimas, M. Sim, and M. Zhang, "Adaptive distributionally robust optimization," *Manage. Sci.*, vol. 65, no. 2, pp. 604–618, 2018.
- [207] L. He, Z. Hu, and M. Zhang, "Robust repositioning for vehicle sharing," *Manuf. Serv. Oper. Manage.*, 2019.
- [208] C. Shang and F. You, "Distributionally robust optimization for planning and scheduling under uncertainty," *Comput. Chem. Eng.*, vol. 110, pp. 53–68, 2018.
- [209] L. Wu, C. He, X. Zhang, and T. Liu, "Distributionally robust scheduling of integrated gas-electricity systems with demand response," *IEEE Trans. Power Syst.*, pp. 1–1, 2019.

- [210] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, no. 1-3, pp. 193–228, 1998.
- [211] MOSEK. [Online]. Available: <https://www.mosek.com/>
- [212] Hourly aggregated wind power data. Available: <http://www.ercot.com/gridinfo/generation/>, accessed July 15, 2019.
- [213] AESO. Available: <https://www.aeso.ca/market/market-and-system-reporting/>, accessed July 15, 2019.
- [214] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 670–681, April 2019.
- [215] L. Wang, Q. Li, R. Ding, M. Sun, and G. Wang, "Integrated scheduling of energy supply and demand in microgrids under uncertainty: A robust multi-objective optimization approach," *Energy*, vol. 130, pp. 1–14, 2017.
- [216] J. Wang, H. Zhong, Z. Ma, Q. Xia, and C. Kang, "Review and prospect of integrated demand response in the multi-energy system," *Appl. Energy*, vol. 202, pp. 772–782, 2017.



Dual Problem and Change of Variables

The problem reformulation based on duality theory and change of variables in Chapter 3 is presented in this appendix. From (3.26)-(3.28), we can get the dual problem by introducing corresponding dual variables θ_0 , θ_1 and θ_2 as follows:

$$\inf_{\theta_0, \theta_1, \theta_2} \theta_0 + \theta_1 h(x) \mu_t + \theta_2 [h^2(x) \sigma_t^2 + (h(x) \mu_t)^2] \quad (\text{A.1a})$$

$$s.t. \theta_0 + \theta_1 r + \theta_2 r^2 \geq 0 \quad (\text{A.1b})$$

$$\theta_0 + \beta - h^0(x) + (\theta_1 - 1)r + \theta_2 r^2 \geq 0. \quad (\text{A.1c})$$

It can be verified that strong duality holds as $h^2(x) \sigma_t^2$ is positive [165] and the dual problem has feasible solutions when $\theta_2 > 0$. The constraints (A.1b) and (A.1c) can be transformed into the following equivalent constraints by considering the minimum value of the left hand side:

$$\inf_{\theta_0, \theta_1, \theta_2} \theta_0 + \theta_1 h(x) \mu_t + \theta_2 [h^2(x) \sigma_t^2 + (h(x) \mu_t)^2] \quad (\text{A.2a})$$

$$s.t. \theta_0 - \frac{\theta_1^2}{4\theta_2} \geq 0, \theta_0 + \beta - h^0(x) - \frac{(\theta_1 - 1)^2}{4\theta_2} \geq 0. \quad (\text{A.2b})$$

Then using the following variables change with auxiliary variables v , τ and $z > 0$, we can obtain the problem (3.29).

$$\theta_0 = v + \frac{(\tau - h(x) \mu_t)^2}{4z}, \theta_1 = \frac{\tau - h(x) \mu_t}{2z}, \theta_2 = \frac{1}{4z}.$$

B

Dual of SOCP Problem

The duality theory for SOCP problem is outlined in this appendix. Consider the following SOCP problem:

$$\min f^\top x \quad (\text{B.1})$$

$$s.t. \|A_i x + b_i\| \leq c_i^\top x + d_i, i = 1, \dots, N, \quad (\text{B.2})$$

where $x \in \mathbb{R}^n$ is the optimization variable, and $f \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{(n_i-1)n}$, $b_i \in \mathbb{R}^{n_i-1}$, $c_i \in \mathbb{R}^n$ and $d_i \in \mathbb{R}$ are corresponding problem parameters. The norm used in the constraints is usually the Euclidean norm, i.e., $\|y\| = \sqrt{y^\top y}$. Then the dual of this SOCP problem is as follows:

$$\max - \sum_{i=1}^N (b_i^\top z_i + d_i w_i) \quad (\text{B.3})$$

$$s.t. \sum_{i=1}^N (A_i^\top z_i + c_i w_i) = f, \quad (\text{B.4})$$

$$\|z_i\| \leq w_i, i = 1, \dots, N, \quad (\text{B.5})$$

where $z_i \in \mathbb{R}^{n_i-1}$ and $w \in \mathbb{R}^N$ are dual variables.