Bridging probabilistic forecasts and power system optimization considering uncertainty: A joint chance-constrained perspective

Yifu Ding

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

Acknowledgements

Looking back on the past four years, I feel so grateful to study and research at Oxford, where I met so many brilliant minds. At this moment even I am far from Oxford and my hometown, my heart is with my family, dearest friends, colleagues, and mentors. I won't be here without your help.

First and foremost, I would like to sincerely thank my supervisor, Prof. Malcolm McCulloch, for his continuous support and guidance during my Ph.D. study. Your leadership and great optimism helped me and the entire EPG group stay strong and active during this unprecedented pandemic.

I also would like to thank my advisor, Prof. Benjamin Hobbs, during my visiting research at Johns Hopkins University. Thank you for hosting me in your group and supporting me in the final stage of my Ph.D. I learned my very first lecture on the power market and game theory from you. Your kindness and enthusiasm will inspire me to continue researching.

Then, my gratitude goes to my colleagues for their help in my research. To Renke, Derek, and Prof. Dan Rogers, thank you for the insightful discussion on microgrid design. To colleagues and alumna of EPG, thanks for your friendship, support and academic writing suggestions. To colleagues at Turing Institute and JHU, thanks for the insightful discussion on machine learning and the US power market. My gratitude also goes to my friends inside and outside Wolfson college. This page cannot name all of you, but thanks for all these joyful moments in the common room, punting, tennis and squash courts. To my dearest friends Ping Lu and Ke Zhang, thank you for always being by my side, comforting me and giving me a hand whenever I experience difficulties.

In particular, I would like to thank my thesis examiners, Prof. Pierre Pinson and Prof. Kostas Margellos, for their comprehensive suggestions in each chapter of this thesis.

Finally, my deepest love goes to my parents, especially my mother. You teach me to be a kind, independent, and confident woman. Thank you so much for your companionship at every important moment and for always believing in me and my decisions.

Abstract

Power systems globally are significantly integrating renewable energy to achieve net zero emission targets. Nonetheless, these renewable energy sources are weather-dependent and uncertain, bringing challenges to ensure operational reliability. Additionally, extreme weather and natural disasters damage power infrastructure, leading to contingencies or large-scale blackouts occasionally. These pressing issues necessitate uncertainty-aware power system management and give rise to the main research question, *how to apply the joint chance-constrained (JCC) formulation in uncertainty-aware power system operations using probabilistic forecasting information?*

A literature review discusses different machine learning (ML) probabilistic forecasting models and uncertainty-aware formulations in power system operations. Selected ML models are then compared for weather-informed net-load probabilistic forecasting. Leveraging probabilistic forecast, this thesis investigates and applies two uncertainty-aware formulations in power system operations, the JCC and distributionally robust optimization (DRO). Case studies are demonstrated on solar-powered networked microgrids.

The first case study investigates solar-powered networked microgrids during utility contingencies. A distributionally robust joint chance-constrained (DRJCC) framework is developed to control this networked microgrid considering the worst-case solar forecast error distribution. The DRJCC problem is reformulated into a tractable individual chance-constrained formulation with unspecified individual violation rates. A novel evolutionary algorithm is designed to approximate the optimal individual violation rates, known as the NP-hard Bonferroni Approximation problem. The solution is proven to be reliable and less conservative in out-of-sample tests compared to other methods.

The second one focuses on networked microgrids in the power markets. It envisages that, due to stochastic solar power generations, the imperfect real-time reserve provisions from multiple microgrids simultaneously could result in unexpected load shedding. Leveraging historical forecast error samples, the reserve bidding strategy of each microgrid is formulated into a two-stage Wasserstein-metrics DRO model. A JCC regulates the under-delivered reserve capacities of all microgrids in a non-cooperative game. To approximate the optimal individual violation rates of microgrids with the changing correlation, the Bayesian optimization method is introduced. This DRJCC game framework is simulated with up to 14 players in a 30-bus network. The proposed Bayesian optimization method can identify the optimal reserve violation rates, which effectively regulate the joint violation rate of the under-delivered reserve whereas securing the profit of microgrids.

Contents

List of Figures

List of Tables

List of Abbreviations

AI . . . . artificial intelligence
ANN . . . . . . artificial neural network
ARMA . . . . . autoregressive moving average
ARIMA . . . . autoregressive integrated moving average
BNN . . . . . . Bayesian neural network
BTM . . . . . . Behind the meter
CAISO . . . . . California Independent System Operator
CDF . . . . . . cumulative distribution function
CNN . . . . . . convolutional neural network
CNN-SE . . . . convolutional neural network with squeeze-and-excitation modules
CRPS . . . . . . continuous ranked probability score
CVaR . . . . . . conditional value-at-risk
DC . . . . direct current
DBN . . . . . . deep belief network
DER . . . . . . distributed energy resources
DLMP . . . . . distribution locational marginal price
DNN . . . . . . deep neural network
DRCC . . . . . distributionally robust chance-constrained
DRJCC . . . . . distributionally robust joint chance-constrained
DRO . . . . . . distributionally robust optimization
DSO . . . . . . distribution system operator
ELM . . . . . . extreme learning machine
EENS . . . . . . expected energy not served
ED . . . . economic dispatch
ES . . . . ensemble learning (Chapter 2)

ES . . . . energy storage

EV . . . . electric vehicle

G . . . . Gaussian-based

GAN . . . . . . generative adversarial network

GBM . . . . . . gradient boost machine

GELM . . . . . generalized extreme learning machine

GPR . . . . . . Gaussian process regression

GRU . . . . . . gated recurrent unit

JCC . . . . . . joint chance-constrained

KDE . . . . . . kernel density estimation

KKT . . . . . . Karush-Kuhn-Tucker rule

HV . . . . high voltage

LMP . . . . . . locational marginal price

LSTM . . . . . long short-term memory

LOLP . . . . . loss of load probability

LP . . . . linear programming

LV . . . . low voltage

MAE . . . . . . mean absolute error

MAPE . . . . . mean absolute percentage error

MCM . . . . . . markov-chain mixture distribution

MCMC . . . . . markov chain Monte Carlo

MC . . . . Monte Carlo

MG . . . . microgrid

MDN . . . . . . mixture density network

MPC . . . . . . model predictive control

MPEC . . . . . mathematical programming with equilibrium constraints

MSE . . . . . . mean square error

MILP . . . . . . mixed integer linear programming

MLR . . . . . . multivariate linear regression

ML . . . . machine learning

NF . . . . normalized flow

NG . . . . parametric non-Gaussian
NN . . . . neural network
NP . . . . non-parametric
NP-hard . . . . non-deterministic polynomial-time hardness
OPF . . . . . . optimal power flow
PDF . . . . . . probability distribution function
PPF . . . . probabilistic power flow
PV . . . . Photovoltaics
QRF . . . . . . quantile random forest
QRA . . . . . . quantile regression average
ReLU . . . . . . Rectified Linear Unit
RF . . . . random forest
RL . . . . reinforcement learning
RO . . . . robust optimization
RT . . . . real time
RNN . . . . . . recurrent neural network
UFLS . . . . . . under-frequency load shedding
SE . . . . square exponential
SVM . . . . . . support vector machines
SOCP . . . . . . second-order cone programming
SoC . . . . state of charge
SDP . . . . semidefinite programming
SCC . . . . . . single chance-constrained
SSA . . . . sub-Saharan Africa
TCL . . . . . . thermostatically controlled loads
T&D . . . . . . transmission and distribution
TSO . . . . . . transmission system operator
UK . . . . United Kingdom
US . . . . United States
VAE . . . . . . variational auto-encoders
VPP . . . . . . virtual power plants
VoLL . . . . . . value of lost load
WS . . . . winkler score

1
Introduction

Contents

1.1 Motivations

1.1.1 Maintaining power system reliability in a net-zero transition

Limiting global warming to 1.5 degrees Celsius above the pre-industrial level requires reducing global emissions to net zero by 2050 [1]. More than 130 countries have committed to net-zero carbon emissions targets as of 2022 [2]. Despite the different status of net-zero carbon emissions targets across countries, decarbonizing the largest carbon-emitting industries, including power generation, heating, and transportation, is of paramount importance. For the power generation sector, key measurements for carbon reduction include phasing out fossil fuel plants such as coal and gas and introducing renewable energy sources such as wind and solar power. However, as shown in Figure 1.1, the majority of countries or regions that are committed to net zero have a share of electricity production from renewables that are less than 50 % as of 2022. This means that much effort is needed to integrate renewable energy in the next two decades globally.

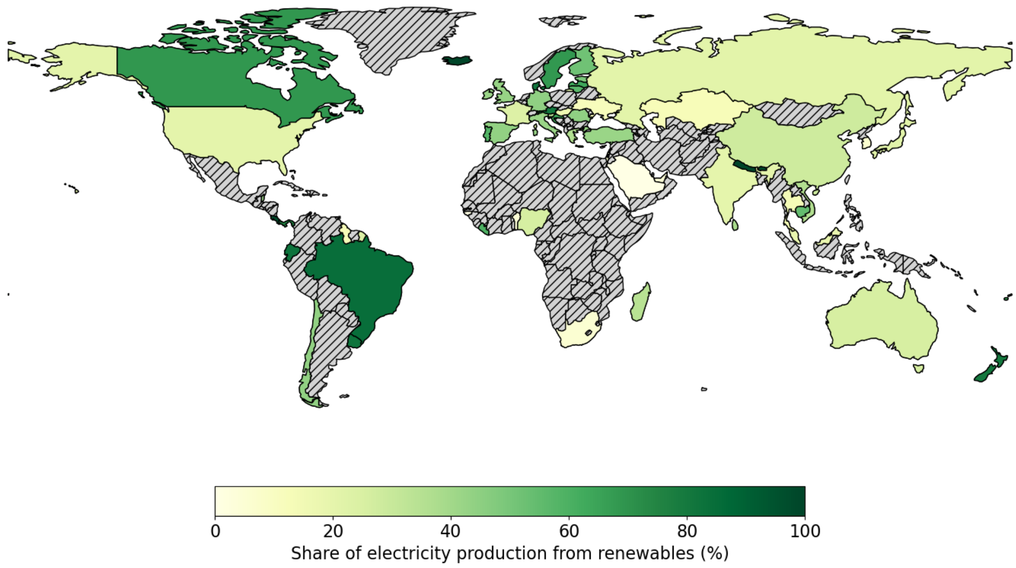

Figure 1.1: Share of electricity production from renewable in countries or regions that have committed to the net-zero emission target [2] (The gray shading area are regions has no action for net-zero target)

The integration of renewable energy sources raises reliability issues in power system operations. As renewable energy resources are inherently weather-dependent and uncertain, system operators need to predict their power outputs to maintain system balance. This will inevitably produce forecast errors, which must be compensated by backup thermal generations and system flexibility. On the other hand, due to climate change, frequent extreme weather events and natural disasters damage infrastructure or disrupt the normal operations of power systems. This could result in severe frequency perturbations and then disastrous blackouts [3]. Renewable energy resources are connected to the grid via controlled power electronics instead of synchronized turbines. During contingencies, these inverter-based resources, if not being programmed for synthetic inertia, cannot provide ancillary or grid security services such as frequency or inertia response like thermal power plants [4]. To enhance power system reliability and resilience on the road to net-zero systems, transformative changes are needed to assist the integration of renewable energy.

1.1.2 Shifting towards uncertainty-aware power system management

The operational reliability of power systems refers to meeting electricity demands and maintaining system balance in light of various changing conditions [5]. For a power system integrating a high amount of renewable energy, one of the approaches to improve operational reliability is accounting for the stochastic nature of renewable energy sources and shifting towards uncertainty-aware power system management. An example is the use of probabilistic energy forecasting. Unlike deterministic energy forecasting, it produces a variety of grid scenarios that capture the instantaneous uncertainty caused by external factors such as weather.

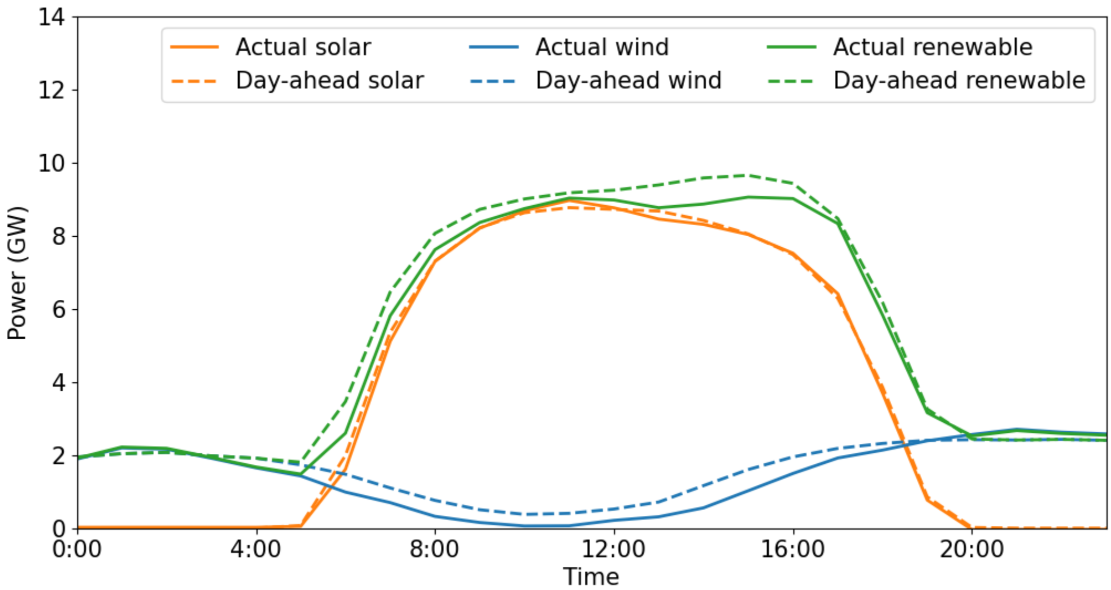

Figure 1.2: Day-ahead forecasted and actual wind, solar, and total renewable power in California, US on 27 June 2022 [6]

Figure 1.2 shows differences between day-ahead deterministic forecasts and actual values in California, US. These forecast differences must be compensated by the contracted reserve or demand flexibility. The oversubscription of the reserve could result in low-efficient and costly operations of stand-by thermal power plants. System operators in California and Australia are implementing weather-informed probabilistic net load forecasting to estimate reserve and ramping requirements [7, 8, 9]. Compared to reserve bounds set by historical error records, probabilistic net load forecasting produces a reserve margin that gives a higher probability coverage for imbalances with less cost [9].

Another example is adapting to uncertainty-aware in power system control and operations, even though most utilities still use deterministic scheduling today. With the uncertainty-aware formulation, system operators could decide on the operational reliability level and conduct operations such as power dispatch or schedule at different risk levels. These uncertainty-aware optimization results consider the worst-case scenario or low-frequency events, so operators can effectively prepare for extreme circumstances and arrange contingency plans such as load shedding [10]. Apart from system operators, small-scale participators such as aggregators in the electricity market can coordinate uncertain behind-the-meter (BTM) generations and flexibility [11], but they are exposed to uncertain and high fluctuations in real-time prices and underperforming risks. Using uncertainty- or risk-aware formulations, they can make informed decisions that maximize their profits and minimize their exposure to potential losses [12].

1.1.3 Enhancing system reliability and resilience with renewable networked microgrids

Upgrading the infrastructure or increasing the reserve margin can improve system reliability in the long run. This, however, entails considerable investments and a long implementation time. As power systems shift from the centralized to decentralized paradigm [13], networked microgrids are the short-term alternatives for reliable power supply. These microgrids consist of dispatchable units such as diesel generators, non-dispatchable units such as renewable energy sources, energy storage, and local loads.

In sub-Saharan Africa (SSA) and South Asia, where the main grid is under-developed and suffers from intermittent power supplies, microgrids are installed as the backup supply for households and small businesses as a sustainable and scalable solution [14, 15]. In more developed countries such as the US, networked microgrids can be transported to or self-organized at points where high-voltage lines are down due to wildfires or hurricanes [16]. These microgrids supply critical loads, including lighting and hospital facilities, for hours or days until the primary energy source recovers. They can also provide security services to the upstream grid, such as reserve or restoration services [3].

Controlling networked microgrids under renewable uncertainty is challenging. Especially when a microgrid is an isolated and small entity, it has a lower inertia and can experience larger frequency perturbations compared to a large grid. Microgrid control can be classified into three types, primary, secondary, and tertiary control. The primary control is the local control that concerns voltage regulator and frequency within the microgrid, whereas the secondary and tertiary control concerns economic dispatch and external coordination with the main grid, respectively [17]. Uncertainty-aware formulations can be applied in all three control levels to address power imbalance, voltage, and frequency violations due to different uncertainty resources.

In summary, this thesis will apply and investigate advanced uncertainty-aware optimization in power system operations, leveraging probabilistic forecasting to enhance operational reliability. To illustrate the effectiveness of these methods, this thesis will present case studies of controlling and coordinating networked microgrids in different grid operation scenarios. These case studies will highlight the cost-effectiveness, robustness, and scalability of uncertainty-aware control and coordination strategies.

1.2 Research scope

Uncertainty-aware optimization can be applied in various power system operations, and their time horizons range from several seconds to one decade. These include the component-level dynamic control in seconds to minutes, power system dispatch and operations in minutes to hours, to power system reinforcement and planning in days to years or even decades. Therefore, the definitions of uncertainty and reliability vary from one research context to another, and it is necessary to define the research scope before studying the main research question.

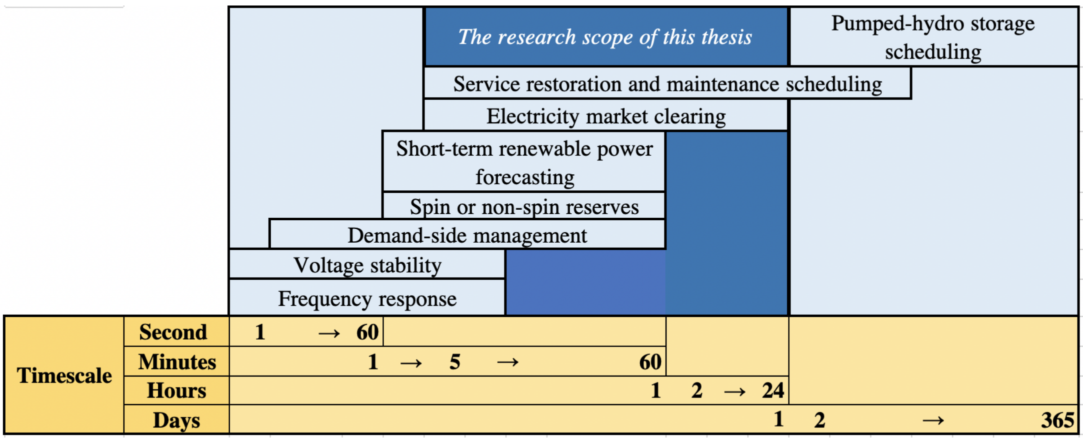

Figure 1.3: Research scope of this thesis in power system operations

Figure 1.3 shows an example of power system dynamic control and operations in minutes to days. This thesis mainly investigates uncertainties and risks resulting from intra-day or day-ahead energy forecasting. The optimization formulations in the discussion are for short-term power system operations within one day. Specifically, this involves demand-side response, intra-day and day-ahead energy forecasting, microgrids scheduling and dispatch, and day-ahead and real-time market clearing. The research scope does not include stability control, sizing and planning for microgrids, and mid or long-term energy forecasting. The reliability in this thesis refers to operational reliability, which mainly concerns network violation events in minutes and the operational performance of energy services in the day-ahead and real-time electricity market.

1.3 Thesis outline and contributions

1.3.1 Main research question and chapter abstracts

The principal research question of this thesis is *"How to apply the JCC formulation in uncertainty-aware power system operations using probabilistic forecasting information?"* In this thesis, Chapters 2 - 5 progressively address this problem by reviewing the literature, proposing new methods, and applying them in two case studies of networked microgrids. These chapters are summarized as below,

Chapter 2 introduces the main research question by conducting a literature review of ML probabilistic forecasting and uncertainty-aware optimization formulations. In particular, this thesis focuses on two formulations in power system applications: distributionally robust optimization (DRO) and JCC optimization. The non-convexity and NP-hardness of the JCC problem are discussed. Existing methods to approximate the solution to the JCC problem are compared and discussed.

Chapter 3 compares two types of ML probabilistic forecasting models for the aggregated net-load profiles, including additive Gaussian process and deep learning models. The net-load profiles from low-voltage substations with varying levels of PV penetration are forecasted. These probabilistic forecast results are also used in the case study in Chapter 5.

Chapter 4 first employs the distributionally robust joint chance-constrained (DRJCC) formulation in controlling a solar-powered DC networked microgrid. This networked microgrid aims to provide a reliable power supply during utility contingencies. DRO formulation secures the solution robustness even for the worst-case solar forecast error distribution, while JCC formulation regulates the joint rate of network violation events. To solve the DRJCC problem in a non-conservative manner, the NP optimal Bonferroni approximation is formulated, and an evolutionary algorithm is proposed to solve the problem. The proposed algorithm can secure the desired operational reliability level and significantly reduce solution conservativeness compared with the Bonferroni approximation.

Chapter 5 applies the DRJCC formulation in a game-theoretical market framework. In a two-settlement power market, the real-time reserve provision of microgrids is imperfect due to uncertain solar generations. If multiple microgrids breach the day-ahead contract simultaneously, it could result in a grid contingency and unexpected load shedding. This game models the system operator's regulation in a JCC formulation and models risk-averse bidding strategies of microgrids in a two-stage DRO. Instead of enforcing the conservative Bonferroni approximation, Bayesian optimization is proposed to approximate the optimal solution under the changing correlation among players. The scalability of this game framework is tested by a non-cooperative game with an increasing number of players. The proposed Bayesian approach can regulate the joint violation rate while ensuring each microgrid's profit.

Chapter 6 concludes the main results of this thesis. Three recommendations are outlined from the thesis results to shed light on future works.

1.3.2 Contributing publications

The list of research publications directly related to this thesis is presented as follows.

- Y. Ding, M. McCulloch, Additive Gaussian process prediction for electrical loads compared with deep learning models. In Proceedings of the Twelfth ACM International Conference on Future Energy Systems (pp. 499-506). https://doi.org/10.1145/3447555.3466592 (**Chapter 3**)

- Y. Ding, T. Morstyn, M. McCulloch, Distributionally robust joint chance-constrained optimization for networked microgrids considering contingencies and renewable uncertainty. IEEE Transactions on Smart Grid 13, no. 3 (2022): 2467-2478. https://doi.org/10.1109/TSG.2022.3150397 (**Chapter 4**)

- Y. Ding, M. McCulloch, Distributionally robust optimization for networked microgrids considering contingencies and renewable uncertainty." In 2021 60th IEEE Conference on Decision and Control (CDC), pp. 2330-2335. (**Chapter 4**)

- Y. Ding, A. Vijay, D. Neal, D. Rogers, and M. McCulloch, "Model predictive control for grid-ready microgrids in developing countries," 2020 IEEE PES/IAS PowerAfrica, 2020, pp. 1-5, doi: 10.1109/PowerAfrica49420.2020.9219857. (**Chapter 4**)

- Y. Ding, S. Wang, B. Hobbs, Coordinating renewable microgrids for reliable energy and reserve services: a distributionally robust chance-constrained game model, 2023 (submitted to the 14th ACM International Conference on Future Energy Systems, **Chapter 5**)

- Y. Ding, B. Hobbs, Joint chance-constrained game for coordinating microgrids in energy and reserve markets: a Bayesian Optimization approach (submitted to IEEE Transactions on Smart Grid, **Chapter 5**)

The list of research publications indirectly related to this thesis is presented as follows.

- Green Radio: Dynamic power saving configuration for mobile networks, The Alan Turing Institute, http://doi.org/10.5281/zenodo.3786852 (**Chapter 2**)

2
Literature Review

Contents

***Chapter Summary**: Probabilistic energy forecasting can anticipate future uncertainty and risk. It can be applied in various power system applications using uncertainty-aware formulations. The theory of ML probabilistic energy forecasting has been studied extensively. Compared to the statistical model, ML models have the advantage of dealing with non-linearity in data, and their structures can be flexibly adjusted for downstream operations. This motivates the research to investigate connections between prediction and optimization. This chapter identifies three critical stages between prediction and optimization. Assumptions, merits, and demerits of probabilistic forecasting and uncertainty-aware formulations in different types are discussed. Finally, this chapter summarizes research trends and gaps based on the existing literature.*

Existing literature has studied probabilistic energy forecasting in the four major representations - prediction intervals, quantiles, scenarios, and density [12]. Figure 2.1 outlines four types of probabilistic energy with UK national demand data. These types of probabilistic energy forecasting have been used in different grid applications at the transmission, distribution, and end-user levels, as listed in Table 2. For certain applications, such as predicting solar power, wind power, electrical loads, and electricity prices, probabilistic forecasting is a very mature technique, and methodologies have been extensively studied for all four types.

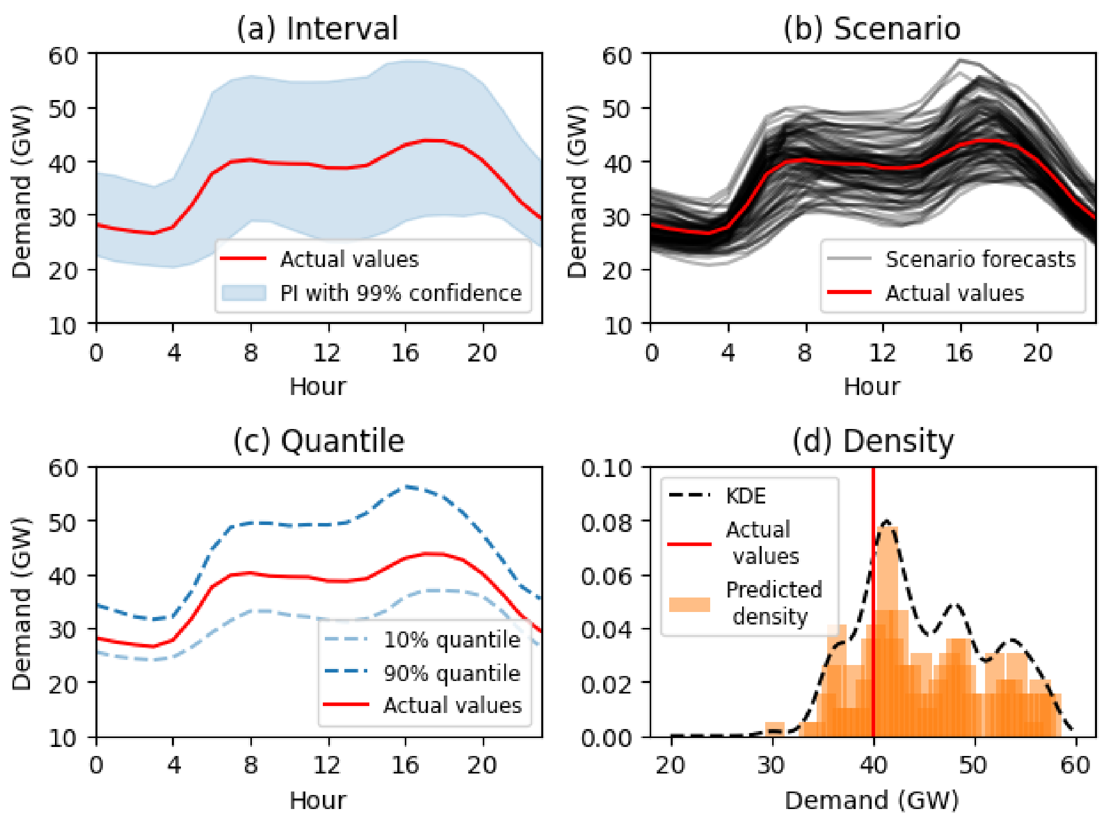

Figure 2.1: Four major representations of probabilistic forecasting (a) interval (b) scenarios (c) quantile and (d) density

Some research topics in probabilistic energy forecasting have recently emerged, including system inertia and grid states. Ref. [46] designs a density forecasting model for system inertia in the UK power system. Ref. [44] develops a quantile ML model to predict the node voltages in a distribution power network. Then the optimal power flow (OPF) is simulated by considering several EV charging scenarios to manage the voltage violations at a reduced cost.

Although the theory of probabilistic energy forecasting is well-established, deterministic forecasting is still dominant in practical power system operations. For example, most US system operators only use a particular quantile forecast to calculate the worst-case scenario as of 2020 [49]. One of the main reasons is that probabilistic forecasting takes practice to understand, and there is no standard for applying diverse probabilistic forecasting techniques in uncertainty-aware optimization and power system operations.

In the face of these challenges, this chapter identifies three main stages, from upstream energy forecasting to downstream uncertainty-aware power system optimization. These stages are ML prediction, integration paradigms, and uncertainty-aware formulations. Compared to the existing reviews solely focusing on either forecasting designs [50, 51] or optimization methods [52, 53], this review tries to identify connections between two and how different choices impact the decision-making.

2.1 ML probabilistic forecasting in different representations

Probabilistic forecasting can be conducted using either statistical or ML approaches. Statistical approaches range from multivariate linear regression (MLR) [54], autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) type models [55] to recently-developed generalized additive models (GAM) [56]. In

	Wind	Solar	Loads		EV	Heat	Inertia	Price	Grid States	
			HV	LV					Line rates	Voltage
(a) Interval	[18]	[19]	[20]	[21]		[22]		[23]		
(b) Scenario	[24]	[25]	[26]	[27]	[28]	[29]		[30]		
(c) Quantile	[31]	[32]	[33]	[34]	[35]	[36]		[37]	[38]	[39]
(d) Density	[40]	[41]	[42]	[43]	[44]	[45]	[46]	[47]	[48]	

Table 2.1: Existing works of four probabilistic forecasting representations in the short-term power system applications (Each position shows an example of relevant works, and the blank space represents the topics yet to be explored.)

general, statistical models are significantly faster to compute and easier to interpret. The magnitude of marginal effects (i.e., how fast the predicted variable changes with each feature) and statistical significance of regressors (i.e., the importance of each feature for the final prediction result) can be readily identified [57]. However, statistical models are generally limited in capturing non-linearities in the data.

ML can overcome this bottleneck by employing information-rich NN structures. In the past decades, ML structures have evolved from shallow learning to deep learning structures which can process millions of data points. Shallow learning structures include support vector machine (SVM), random forest (RF), and artificial neural network (ANN) [58]. Deep learning structures are built upon complex neural network (NN) topologies consisting of hundreds of individual neurons or units. These structures include convolutional neural network (CNN) [59], gated recurrent unit (GRU) [60], and long short-term memory (LSTM) network [61]. Recent works also introduce the transformer [62], attention [63, 64] and multitask learning mechanisms [65] in the NN structures to enhance the model performance and reduce the training efforts.

Another advantage of the ML probabilistic forecasting model is that its structure can be changed based on the need for decision-making. NNs usually use mean square error (MSE) as the loss function in deterministic predictions. Not only can ML predictors change loss functions, the number of hidden layers and neurons on each layer for different power system operations. Additionally, there are different kinds of model frameworks to leverage different predictors, including ensemble models [66, 67, 68, 69], load decomposition [70, 71, 72, 73, 20], and intelligent selection [42].

The ensemble framework incorporates a family of forecasters in either one type with different hyper-parameter settings or multiple types. Considering the pros and cons of various ML algorithms, an ensemble of models usually performs better than individuals by canceling out the bias from a single model. Refs. [66, 67, 68, 69] employ the ensemble framework and optimize the weight of each model based on a probabilistic evaluation score. Refs. [70, 71, 72, 73, 20] decompose the electrical load first using the wavelet technique or empirical mode technique, and forecast decomposed components separately using different ML models. To yield the best performance, ref. [42] create a closed-loop reinforcement learning (RL) framework to select the best forecaster competitively in the rolling horizon.

Table 2.2 shows a list of the literature surveyed, which summarizes ML forecasting methods in each of the four forecasting types. This table lists their ML or NN structures and underlying distributions (i.e., Gaussian distribution (G), parametric non-Gaussian distribution (NG), or non-parametric random distribution (NP)). This shows how different forecast design choices impact decision-making optimization. As shown in Table 2.2, quantile and scenario forecasting has no underlying distribution, and density forecasting can produce a Gaussian-based or mixture of Gaussian-based density predictions. This chapter uses the **bold font** to denote the vectors of variables and parameters such as $\mathbf{x}$, and the subscript refers to the individual elements of vector such as $\mathbf{x}_i$.

2.1.1 Prediction Interval

In energy forecasting, a prediction interval $\mathscr{I}$ covers future realizations of an uncertain variable with a certain probability of $100(1-\beta)\%$.

$$\mathscr{I} \in \{[l,u] | \mathbb{P}(l \leq y \leq u) = 100(1-\beta)\%\} \tag{2.1}$$

Prediction intervals could have distribution assumptions or no assumptions and are often centered around the mean value of an uncertain variable. A prediction interval could inform system operators of the likelihood of outcomes within a range. This result can be used in robust optimization (RO). This section introduces two ML methods to generate the prediction interval.

I.1. Prediction interval with a coverage probability: NNs can be used to predict an interval using an appropriate loss function. Refs. [74, 75] construct ANNs to predict

	ML Categories	Ref.	ML or NN structures	Distributions
Interval	*I.1. Prediction interval with a coverage probability*	[74, 75]	ANN	NP
		[70]	SVM	
		[71]	GELM	
		[76]	ELM	
		[20]	LSTM	
	I.2.Patten recognition	[77]	Clustering	
Quantile	*Q.1.ML quantile forecasting*	[78]	SVM	NP
		[34, 79]	GBM	
		[80]	LSTM	
		[72]	RF	
		[81]	RNN	
		[67]	GRU	
		[82]	GPR	
		[33]	CNN, RNN, LSTM, GRU	
		[83]	ANN (Non-crossing)	
Scenario	*S.1. Generative models*	[84, 85]	GAN	NP
		[85, 86]	VAE	
		[85, 86]	NF	
	S.2. Markov chain	[87]	MCM	
		[88]	MCMC	
Density	*D.1. Gaussian process*	[89]	GPR	G
		[90, 91]	Sparse GPR	
		[92]	Additive GPR	
		[91]	Deep GPR	
		[93]	GPR with warped inputs	NG
	D.2. Density NN	[94]	ANN	G
		[95]	CNN, GRU	NG
	D.3. Bayesian NN	[43]	Bayesian LSTM	NG
		[65]	Bayesian LSTM	
		[96]	Bayesian MDN	

Table 2.2: A brief survey of ML probabilistic forecasting techniques in four major epresentations

the interval of electrical loads. Its loss function maximizes the coverage probability of future load scenarios. Ref. [76] uses an extreme learning machine (ELM) to integrate a linear programming model into the predictor. The model has an objective function to maximize the sharpness of the prediction interval and constraints are imposed for the good calibration of the interval.

I.2. Pattern recognition: Unsupervised learning algorithms such as clustering can be used for identifying an uncertainty interval. Ref. [77] clusters the mean and variance of electrical load profiles based on the calendar and weather features. Then a load interval is generated by combining load patterns based on different types of a day.

2.1.2 Quantile forecasting

Quantile regression is a statistical technique that estimates and conducts the inference about condition quantile functions [97]. Generally, there are no underlying distribution assumptions of quantile forecasting. Compared to a prediction interval, quantile forecasting can inform system operators of the threshold value of prediction outcomes with a chance, which is particularly important for risk evaluation [98]. However, it suffers from a practical issue known as quantile crossing.

Q.1. ML quantile forecasting ML models can use an asymmetrical pinball loss function to produce quantile forecasting. The predicted time series l_t is a realization of a stochastic process whose probability distribution can be divided into Q quantiles. The pinball loss function minimizes the sum of asymmetric absolute residuals between the prediction $\tilde{l}_t$ and ground truth l_t, which is given by,

$$\tilde{l}_t = \underset{\tilde{l}_t}{\operatorname{argmin}} \sum_{t=0}^{T} \rho(q, \varepsilon_t) \tag{2.2}$$

where the forecast error at time t denotes as $\varepsilon_t = l_t - \tilde{l}_t$, and the asymmetric absolute residuals known as the pinball loss $\rho(q, \varepsilon_t)$ is calculated as,

$$\rho(q, \varepsilon_t) = \begin{cases} q\varepsilon_t, & \text{for } \varepsilon_t \geq 0 \\ (q-1)\varepsilon_t, & \text{for } \varepsilon_t < 0 \end{cases} \tag{2.3}$$

This pinball loss function can be applied in shallow ML structures, including GBM [34], SVM [78], and RF [72], and deep ML structures including RNN [81], CNN [33], GRU [67], and LSTM [80]. Notably, a long-lasting issue when using quantile forecasting is *quantile crossing*. It means quantiles that are predicted curve-by-curve might cross at one point. The crossed quantiles cannot form a prediction interval to be used in the operations. Adding non-crossing constraints in the forecasting model [83] can help address this issue.

2.1.3 Scenario forecasting

Scenario forecasting generates a set of prediction scenarios $\mathbf{s} := \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n\}$, and each scenario $\mathbf{s}$ is considered as a realization of an uncertain distribution $\mathbb{P}$. These scenarios are generally considered to have an equal occurrence probability, although a weight of probability could be imposed. The most intuitive scenario generation technique is to run the ML predictor under different scenarios, which is commonly used for weather-informed energy forecasting [26]. This section introduces two ML prediction techniques that can directly produce a spread of future scenarios.

S.1. Deep generative models Deep generative model can train NNs to replicate the observed distributions, which include normalized flow (NF), variational auto-encoders (VAE), and generative adversarial network (GAN). These ML models have been used for generating scenarios from a given probability distribution. A review and comparison of these three models are in refs. [85] and [86]. In particular, the GAN model consists of a generator and a discriminator. The generator produces the volatility samples to 'mimic' the actual forecasting error distribution. A discriminator is used to identify the generated samples from the actual distribution iteratively until these samples cannot be identified and the system reaches an equilibrium. For example, ref. [84] uses a GAN to produce scenarios of load errors based on the past load forecast error distributions.

S.2. Markov-chain Markov chain (also known as Markov process) is a statistical technique to generate future scenarios, which has been applied in the short-term wind power [99], solar power [100] and electrical load forecasting [88, 87]. This model can predict the next state from past states and current observations by modeling the transitions between states. For instance, ref. [87] divides the past load states into N scenarios and models the relationship of current and past load states in a $N \times N$ transition matrix. A Markov-chain mixture (MCM) distribution model is used to predict the whole load distribution by modeling the transitions for each loading scenario in the distribution.

2.1.4 Density forecasting

Density forecasting is to predict the full probability distribution of an uncertain variable $\mathbb{P}$ at any future time, which has the full stochasticity within a prediction interval. In other words, the system operator could know the exact chance of a particular prediction outcome from density forecasting. Theoretically, it could be used for almost all kinds of uncertainty-aware optimization. However, non-Gaussian probability distributions might not be modeled precisely and need approximation when used for optimization. ML methods for density forecasting include nonparametric Bayesian models, Bayesian inference, and density NNs.

D.1. Gaussian process regression: Gaussian process regression (GPR) is one of the most classical nonparametric Bayesian models. This model estimates the regression function based on the prior distribution and kernels (i.e., covariance function) and then optimizes the mean function using posterior observations. It generates Gaussian-based density forecasting. Given the input data x, x', a Gaussian distribution is defined by the mean $\mu(x)$ and the kernel $k(x, x')$ functions, which is given by,

$$f(x) \sim \mathscr{G}\mathscr{P}(\mu(x), k(x, x')) \tag{2.4}$$

GPR shows great advantages in feature interpretability. Since a nonnegative sum of kernel functions is still a kernel function, each kernel can encode individual feature information for different weather variables such as temperature, wind speed, and solar irradiance [92, 91, 101] in an additive framework.

However, it also suffers from several drawbacks. First, the computational cost of the hyper-parameter optimization process increases cubically with the problem dimension. *Sparse GPR* has been proposed which could reduce the computational cost [90, 91]. Another drawback of GPR is the cumbersome kernel design process requiring domain knowledge. Recent work [91] adopts *Deep GPR*, which utilizes the NN structure to form a ‘self-tuning’ covariance function without human intervention. This method has successfully experimented with electrical load predictions for multiple sites. Finally, GPR can only produce a Gaussian-based density. For power system applications which has non-Gaussian distributions, such as residential electrical loads, the *warped GPR* has been introduced. It uses an exogenous function that can transform a normal distribution into a log-normal distribution (i.e., a non-negative distribution) so that the prediction distribution is closer to observations [93].

D.3. Density neural network: By using the maximum likelihood as the loss function, NNs can predict the mean and variance of a probability density [94, 65], named density NNs. However, density NN using the maximum likelihood loss function can only generate a Gaussian-based density prediction. Refs. [95, 96] developed a predictor which can create a mixture of Gaussian distributions (i.e., mixture density network (MDN)). For example, in ref. [95], a CNN-GRU structure was used to build an MDN for probabilistic electrical load predictions, It outputs the mean, variance, and weight of different Gaussian distributions in the mixture density.

D.4. Bayesian neural network: Bayesian neural network (BNN) was first proposed in ref. [102] to capture uncertainty, which conducts Bayesian inference on a NN by means of the neuron dropout. In this work, Bayesian dropout is applied in each hidden layer to maximize the log-likelihood of density forecasting compared to the posterior observations. Using this BNN technique, ref. [65] developed an LSTM network with Bayesian dropout, which is used for probabilistic predictions of clustered customers’ load profiles. Ref. [43] constructed the LSTM network to predict the mean and variance based on the Gaussian distribution. Bayesian inference is conducted over the hyperparameter of LSTM networks to enhance prediction accuracy.

2.2 Integrating forecasting into optimization

In light of the pros and cons of ML predictors in different types, research is motivated to investigate how different forecasting designs will impact downstream decision-making. This section will introduce two types of forecasting paradigms for power system decision-making. Examples are discussed and compared concerning several power system applications.

Prediction then optimization: The 'prediction-then-optimization' paradigm is the most common approach in practical decision-making and operational problems [103]. The forecasting results are input to power system optimization unidirectionally. This, however, has several drawbacks. First, the upstream predictor and downstream optimization have different objectives. In other words, the most accurate prediction might not be the most suitable input for decision-making. For example, the costs of under- and over-estimating electrical loads are different in the economic dispatch [104]. The MSE loss function has symmetrical penalties and thus generates a less cost-effective solution. Second, some forecasting information can be forfeited when input to optimization. For example, the crossed quantile forecasting cannot be used and have to be discarded before optimization.

	Applications	**Prediction-then-Optimization**	**Prediction-and-Optimization**
Point	Predict electricity prices	Predict electricity prices using the MSE loss function [105]	Integrating the regret cost of decision-making into the loss function [106]
Interval	Predict wind power interval	Predict two bounds using ELM to maximize prediction reliability [18]	ELM includes a bi-level optimization for more cost-effective forecasting [107]
Scenario	Generate net load forecast errors scenarios at various network nodes	GAN directly predicts electrical load forecast error scenarios [84]	GAN integrates the power network matrix [108]

Table 2.3: Comparison of two paradigms, prediction-then-optimization and prediction-and-optimization, for the same energy forecasting applications

Prediction and optimization: To address the drawbacks of the 'prediction-then-optimization' approach, the research combines prediction and optimization processes to yield a better overall result in decision-making. These prediction methods are named the 'optimization-in-the-loop' prediction [109], decision-oriented [106], or cost-oriented predictions [110]. ML predictors are redesigned in several ways to consider downstream decision-making processes, for example, using an asymmetrical loss function for different over- or under-estimations [111], introducing the decision regret costs into the loss function of ML predictors [109, 106], integrating operational constraints or parameters into a physical-informed ML predictor [107, 108]. Table 2.3 shows different ML techniques for the 'prediction-and-optimization' in three cases (i.e., point, interval, and scenario forecasting).

As an example of the point forecast, ref. [106] points out decision-making errors (i.e., induced cost as decisions are made on inaccurate forecasts) and forecast errors differ. The work derives the cost of decision-making errors and integrates it into the gradient descent algorithm of the loss function. Similarly, for predicting intervals, ref. [107] uses an extreme learning machine (ELM) to incorporate a decision-making optimization model and generate cost-aware forecasting. The final prediction interval achieves the minimum operational cost in decision-making compared to the other four ML predictors without integrating optimization. For scenario forecasting, ref. [108] integrates the power network constraints into the GAN model to predict the net load error scenarios. The generated error scenarios lead to a less conservative reserve schedule compared to the robust set as the benchmark.

2.3 Uncertainty-aware formulations in short-term power system operations

Each kind of uncertainty-aware optimization has uncertainty assumptions and only uses specific types of probabilistic forecasting as inputs. In this section, different uncertainty-aware optimizations are compared and discussed, and examples are given in terms of their grid applications.

Problem description: A power system optimization problem has an uncertain variable ζ. The sources of uncertainty could be stochastic renewable generation, component failures, or prices in the electricity market. Assuming that constraints hold for any ζ, which follows a distribution $\zeta \in \mathbb{P}$ or a set of distributions $\zeta \in \mathbb{P} \subset \mathscr{P}$. this power system optimization has a set of deterministic decision variables $\mathbf{x} \in \mathbb{R}^d$ given by,

$$\min_{x} \quad f(\mathbf{x}, \zeta) \tag{2.5a}$$

$$\text{subject to} \quad \mathbf{h}(\mathbf{x}, \zeta) = 0, \tag{2.5b}$$

$$\mathbf{g}(\mathbf{x}, \zeta) \leq 0 \tag{2.5c}$$

Based on this problem, six uncertainty-aware formulations are discussed and compared.

Robust optimization: Using the prediction interval or scenario forecasting, RO constructs an uncertainty set based on the worst-case scenario to minimize the cost function (i.e., the max-min problem) [21, 112]. Ref. [21] uses the fuzzy model to predict the intervals of solar power, wind power, and loads. These prediction intervals are used in a RO formulation for an island microgrid. Ref. [112] assumes that solar power forecast deviations are bounded in a box-shaped uncertainty set and controls a LV power network to minimize the operation costs considering the maximum solar power deviations.

Probabilistic power flow: Probabilistic power flow (PPF) was first proposed in ref. [113]. This method improves on the conventional MC sampling by estimating several critical points $\{\mathbf{s}_1, \mathbf{s}_2, \ldots.., \mathbf{s}_n\}$ of an uncertain distribution $\mathbb{P}$. Power flow is then conducted at each estimation point $\mathbf{s}$. Ref. [114] summarizes different point estimation methods for the OPF problem, and it can also be extended for market clearing problem with locational marginal prices (LMPs) [115]. Predicted intervals and quantiles can be used to determine critical points in the PPF formulation.

Single chance-constrained optimization: While RO based on the worst-case scenario often leads to an overly-conservative solution, the chance-constrained (CC) formulation can flexibly control the system reliability and decide the trade-off between the reliability and cost. Given a violation chance α, the single chance-constrained (SCC) formulation is given by,

$$\mathbb{P}(\mathbf{g}(\mathbf{x},\zeta) \leq 0) \geq 1-\alpha \tag{2.6}$$

This means constraint $\mathbf{g}(\mathbf{x},\zeta) \leq 0$ will not be violated with at least a chance of $1-\alpha$, where $\alpha \in (0,1)$. The SCC formulation models the risk or reliability metrics in power system operations. For example, the economic dispatch problem could have the loss of load probability (LOLP) or expected energy not served (EENS) in the constraints [116, 117], and the OPF problem could have line thermal capacity or voltage violations in the constraints [118, 119, 120].

Conditional value-at-risk (CVaR): Original from the portfolio management in the financial sector, value-at-risk (VaR) and CVaR are the two most common metrics for risk measurement [121]. VaR refers to the threshold value of the expected loss at the worst α% of cases, whereas CVaR refers to the average expected loss of all the worst α% of cases. The CVaR formulation of constraint (2.5b) is,

$$\mathrm{CVaR}_{\mathbb{P}}^{\alpha}[\mathbf{g}(\mathbf{x},\zeta)] \leq 0 \tag{2.7}$$

CVaR formulation can also be incorporated into the objective function, often in the two-stage optimization. At the probability level α, the VaR is θ (i.e., $\mathbb{P}[\mathbf{g}(\mathbf{x},\zeta) \leq \theta] \geq 1-\alpha$). The CVaR term is given by,

$$\mathrm{CVaR}_{\mathbb{P}}^{\alpha}[\mathbf{g}(\mathbf{x},\zeta)] = \min_{\theta}\{\theta + \frac{1}{\alpha}\mathbb{E}_{p}[\mathbf{g}(\mathbf{x},\zeta)-\theta]^{+}\} \tag{2.8}$$

Compared to SCC formulation, which only concerns the probability of violations, CVaR formulation also concerns the magnitude of violation events. This allows the optimization to capture high-impact and low-frequency events. An example is the unprecedented peak loads, or generation failures that lead to the worst-case power outage in a very small probability [122]. The SCC and CVaR formulations are equivalent when the uncertain distribution is a normal distribution. However, given a non-Gaussian uncertain distribution (i.e., unimodal), the CVaR formulation yields a more conservative result [123].

Joint chance-constrain optimization: In power systems, there are interlinked constrained or events that are required to be satisfied jointly. Compared to SCC formulation, JCC formulation concerns the whole-system reliability, which considers the joint violation rate of interlinked constraints or events. This thesis considers two main power system applications. This first one is different network violation events resulting from the same uncertainty variable [124, 125]. Another application is the stochastic game incorporating different agents, and each has an uncertain pay-off function in the electricity markets [126]. Assuming that the optimization problem has a set of individual constraints $\mathbf{g} := \{\mathbf{g}_0, \mathbf{g}_1, ..., \mathbf{g}_i\}$ and the subscript i is the index of individual constraints. The JCC of these constraints can be written as,

$$\mathbb{P}(\bigcup_{i=0}^{N_c} \mathbf{g}_i(\mathbf{x}, \zeta) \leq 0) \geq 1 - \alpha_j \tag{2.9}$$

where N_c is the number of individual constraints, and their joint violation rate is constrained below a chance of $1 - \alpha_j$. The JCC problem is generally considered an NP-hard problem in operation and control research, but its solution can be approximated using different methods. Existing methods to solve the JCC problem include scenario-based or approximation-based methods. The scenario-based method solves possible scenarios from the historical samples [127, 128] considering the required confidence level. Given a violation level $\alpha_j \in (0, 1)$ and a confidence level $\beta \in (0, 1)$, one can select the number of scenarios according to ref. [129],

$$\sum_{k=0}^{n-1} \binom{N}{k} \alpha_j^k (1 - \alpha_j)^{n-k} \leq \beta \tag{2.10}$$

Therefore, the number of scenarios N that is needed for a problem with n_x decision variables is given by,

$$N \geq \frac{1}{\alpha_j}\frac{e}{e-1}(n_x - 1 + \ln\frac{1}{\beta}) \tag{2.11}$$

It can be observed that the scenario-based approach requires a large volume of samples for a high-confidence solution. Its scalability heavily relies on statistical techniques, such as sample average estimation, to ease the computation burden [130].

An approximation-based method decomposes an intractable JCC problem into a series of tractable SCCs. It approximates individual violation rates based on *Boole's inequality*, meaning that the sum of individual violation rates is no less than the joint violation rate. The simplest but most conservative approach is *Bonferroni Approximation*, which divides the joint violation rate into many equal individual violation rates [131]. Improvements can be made to reduce conservativeness. Refs. [132, 133] use a tractable CVaR approximation to solve the JCC problem, and obtained around 10% conservativeness reduction compared with *Bonferroni Approximation* in several numerical experiments [132].

Another method is to pre-solve the problem first with a conservative guess and then tighten bounds progressively by identifying the joint violation probability of all possible events. For example, ref. [124] uses the ML classification method to identify all the intersections of active constraints and tighten the union bounds of all constraints in the iterative framework. However, the decomposed SCCs still have the same violation rate. Ref. [125] develops an iterative framework and allocates the violation rate proportional to the violation probability from the pre-solved power flow results. It reported this framework can achieve negligible conservativeness. However, such a framework requires complete information on the simulated system, which is less suited for some applications, such as market bidding. Notably, the problem of optimally allocating individual violation rates of SCCs is termed as the *optimal Bonferroni Approximation* problem. Ref. [134] theoretically proves that this problem is an NP-hard problem.

Distributionally robust optimization: A common challenge in forecasting or uncertainty quantification is that the probability distribution function is not exactly known or modeled. DRO provides a robust alternative other than assuming a single, fixed uncertainty distribution. It constructs a set based on historical samples or density forecasting - termed ambiguity set - including all possible uncertainty distributions. The essence of this optimization is to find the solution that satisfies all constraints for the worst-case distribution in the ambiguity set. For example, if $\mathscr{P}$ is the ambiguity set constructed for the uncertainty variable ζ, the distributionally robust CC and CVaR formulations of constraint (2.5b) are represented as below respectively,

$$\inf_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(\mathbf{g}(\mathbf{x},\zeta) \leq 0) \geq 1-\alpha \tag{2.12}$$

$$\sup_{\mathbb{P}\in\mathscr{P}} \mathrm{CVaR}_{\mathbb{P}}^{\alpha}[\mathbf{g}(\mathbf{x},\zeta)] \leq 0 \tag{2.13}$$

The DRO term can also be incorporated into the objective function. This is usually formulated as a two-stage worst-case expected cost function [135] (i.e., the max-min problem). It can then be reformulated as a CVaR or MILP model to solve [136]. Different methods to construct the ambiguity set determine the trade-off between solution conservativeness and robustness. There are two methods to construct ambiguity sets, moment-based and moment-free methods. The moment-based method utilizes the moment and structure information of distributions. Examples include the symmetric [137, 138], unimodal [139, 140] and multimodal [135] ambiguity sets. However, the moment-based method only leverages part of the information from the probability density. For example, ref. [138] constructs a GAN scenario forecasting model to predict wind power outputs at different nodes in a network and then conducts day-ahead energy and reserve dispatch using DRO. The scenario forecasting, however, only updates the first two moments (i.e., mean and variance) of the ambiguity set.

The moment-free method constructs the ambiguity set directly from the uncertainty samples, which can leverage the probabilistic forecasting completely, compared to the moment-based method. The ambiguity set can be constructed as a ball space. The reference distribution is at the center, and the statistical distance from all uncertainty distributions to the fixed center. These statistical distance metrics include Kullback-Leibler divergence [141] and Wasserstein distance [142, 143, 136].

Formulations	RO	PPF	SCC	CVaR	JCC	DRO
Reliability adjustment	$\times$	✓	✓	✓	✓	$\times$
'Worst-case' scenario	✓	$\times$	$\times$	✓	$\times$	✓

Table 2.4: Characteristics of six uncertainty-aware optimization formulations

Six formulations are compared by two characteristics, as presented in Table 2.4. They are considerations of the reliability levels and the worst-case scenario.

2.4 Research trends and gap analysis

Several research trends and gaps are identified based on the discussion and comparisons in this chapter.

The use of 'prediction-and-optimization' method: As introduced in section 2.2, recent works focus on integrating optimization into prediction in different power system applications, including energy arbitrage and power dispatch. From empirical results, these prediction designs are proven more task-specific and cost-effective in short-term power system operations. However, no theoretical and math proof exists of the convergence of modified ML models. Research could focus on this area and apply the proposed 'prediction-and-optimization' method in a more general case.

Hybrid formulations to leverage merits of different optimization formulations: For some power system applications, the ability to adjust reliability and the consideration of the worst-case case is desirable. The hybrid formulations such as DRCC [139, 140, 137, 135] and DRJCC [144, 126] are well-suited. These works combine the SCC or JCC formulations with the DRO formulation. In this case, the new formulation has merits of two formulations.

Tractable reformulations of DRO using density forecasting: Density forecasting could be integrated into power system operations using DRO formulation. However, tractable reformulations of DRO depend on the ambiguity set constructed. Some reformulations are very computation-expensive if considering high-order moments (e.g., skewness) or unique structures (i.e., multi-modal). Therefore, the choice of ambiguity sets and their construction methods should be carefully evaluated based on empirical density forecasting.

New approximation solutions to NP-hard JCC problem: As aforementioned, the JCC problem is NP-hard, but its solutions could be approximated. Existing methods have been proposed to give solutions with different levels of conservativeness. In particular, the allocation of optimal individual violation rates from a joint violation rate is considered an NP-hard problem (i.e., *optimal Bonferroni approximation)* and no method is offered in the existing literature. Furthermore, limited power system applications use JCC formulations, although JCC formulations secure whole-system reliability.

3

ML Probabilistic Forecasting for Weather-informed Net Load Profiles

Contents

Chapter Summary*: With the proliferation of behind-the-meter (BTM) generations such as solar PV panels, net load forecasting becomes increasingly important in power system managements. Day-ahead probabilistic net load forecasting can be used to size dynamic operating reserve and anticipate network congestion conditions. In essence, the aggregated net load profile of sub-stations is the imbalance between the total load consumptions and BTM generation outputs. However, the characteristics of these BTM distribution assets and loads (e.g., solar generation penetration levels) are invisible to the distribution system operator (DSO) due to privacy issues. Also, various local weather conditions impact net load profiles. All these increase difficulties in predictions.*

This section compares deep-learning and Gaussian process models for the aggregated probabilistic net load forecasting using historical load data and local weather information. In particular, their performances under various weather conditions are evaluated and compared using a set of probabilistic evaluation scores.

3.1 Introduction

The distribution network in developed countries faces paradigm shifts. First, an increasing number of BTM generations and energy storage (ES) have been installed in the distribution network. Second, power system management changes from the centralized and unidirectional to the decentralized and bidirectional fashion [13]. These changes help a bottom-up integration of renewable power, but it also leads to several issues in the network, such as reserve power flow, network congestions [145], and high ramping of thermal generators [146]. Therefore, day-ahead net load forecasting is crucial for network management considering these BTM renewable generations. One of the most famous net load profiles is the California duck curve due to the increasing solar PV penetration in the California area [147]. As a consequence, in several places where BTM solar generations have been widely installed, such as Australia [7], and California [9], power system operators began to employ the data-driven probabilistic net load forecasting to estimate dynamic reserve and ramping product requirements. In Germany, DSO uses the probabilistic net load forecasting at the substation level (referring to the vertical grid load in the operations) to anticipate possible network congestions [10].

The net load profile is the imbalance between the total load and renewable generation. Therefore, when conducting probabilistic forecasting, uncertainties of both generations and demands, as well as their correlations should be considered simultaneously. Existing literature has proposed several methods for net load probabilistic forecasting, including quantile regression [69], deep learning Bayesian model [43, 96], GPR [93] and statistical models [56]. These models use two kinds of net loads in predictions, which are the *decomposed* or *aggregated* net loads.

The *decomposed* net load profiles include both renewable generation outputs and demand profiles. These components are predicted separately and added back to obtain net load probabilistic forecasting. For example, ref. [148] predicts three components of net load profiles - residential loads, solar power, and residues separately. The final probabilistic net load forecasting is obtained by combining predictions of these three parts as well as modeling their correlations.

The *aggregated* net load approach predicts net loads directly, which assumes that BTM generations are invisible. This is a more practical case in reality. First, small-scale BTM generations, such as solar PV generations, are costly to measure at each household [149], and household load profiles in the high granularity are hard to obtain due to privacy issues. Second, the correlation between renewable power and load varies across regions due to microclimate [150], and tedious to model location-by-location accurately. Ref. [43] developed a Bayesian deep learning NN to predict net load profiles. This work predicted net load profiles considering different levels of PV visuality. Results show that available BTM generation profiles can improve the overall prediction accuracy. Ref. [56] uses net-load profiles of UK sub-regions and conducts the weather-informed probabilistic forecasting for each region using a GAM statistical model. A class of weather variables is created and employed as prediction features. This work concluded that the detailed weather information of sub-regions, such as variances, and the maximum and minimum values of local weather variables, could improve prediction accuracy.

Weather features play a critical role in load predictions when using the aggregated net load approach. Since different ML predictors process individual features in different ways. the design of the ML predictor should consider how to process the weather information. For example, GPR offers a higher degree of flexibility and interpretability by encoding the characteristics of each feature into kernels [151], such as different weather variables. However, GPR needs a knowledge-based and cumbersome kernel design, and high computation cost for the high-dimension input data [152]. Deep learning NNs can overcome this dimension problem. However, NNs operate 'black-box' predictors and do not differentiate different characteristics of prediction features [94].

This work develops an additive GPR and deep-learning model (i.e., CNN-LSTM) for weather-informed probabilistic net load forecasting. A set of probabilistic scores evaluates their performances.

3.2 Data processing and feature extraction

3.2.1 An introduction of datasets

This research uses PV Solar Panel Energy Generation trial data from UK Power Network. This trial covers 20 substations with ten domestic premises. It collected solar generations and loads at the household level, substation netload, local voltage, and frequency measurements over 480 days in 10- or 30-minute resolution from 27 July 2013 to 19 November 2014. A detailed description of the dataset and trial is provided in ref. [153].

This work selects four substations with historical weather records from onsite weather stations. Their solar PV penetration levels are measured by how many kW of solar PV panels per customer have been installed, denoted as I_{pv}. Figure 3.1 shows the mean value and 1-σ deviation of half-hour net load profiles at four substations (named YMCA, MDE, FR, and EC), and their solar PV penetration levels I_{pv} are labeled on the titles. It can be observed that the net load uncertainty depends on the variability of solar power. For example, FR and EC substations have higher solar PV penetrations due to variable solar power outputs under different weather conditions throughout the year.

3.2.2 Feature engineering

A range of prediction features $F = \{F_1, F_2, ..., F_n\}$ are created in two categories, weather features F_w and time features F_t. Figure 3.2 gives a summary of all features. Weather

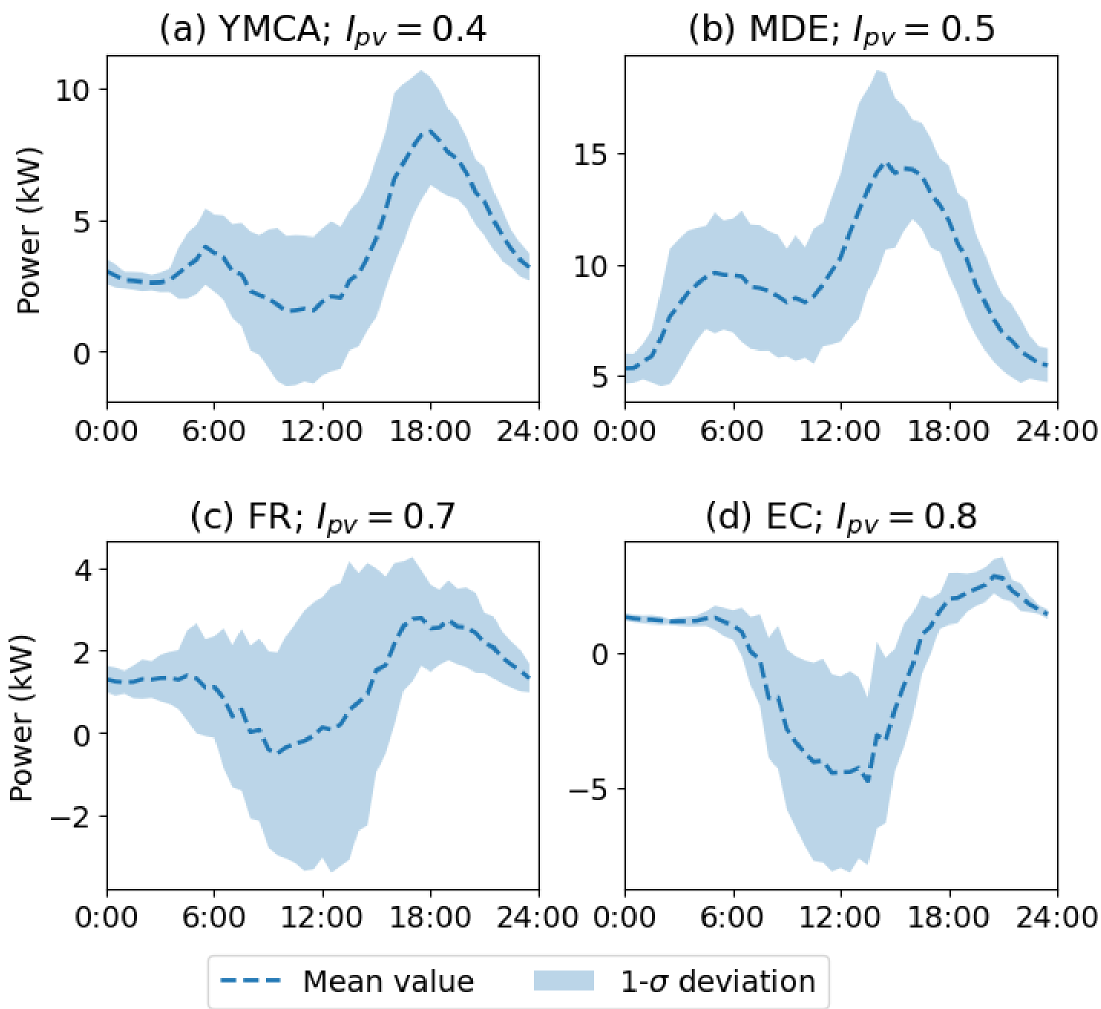

Figure 3.1: Aggregated net load profiles at four substations with different levels of solar PV generation penetration (The blue dash line represents the mean value; The shaded area represents the 1-σ deviation)

features (orange) include six meteorological variables from onsite weather stations. Time features (green) include the hour of the day, the day of the month, the month of the year, the order of time intervals, and three selected lag variables with the highest auto-correlation values (i.e., 48-hour, 50-hour, and 52-hour ahead). The correlations of all features are calculated, and results are shown in Figure 3.2. Their value ranges from -1 to 1, and the positive value means a positive correlation.

The net load value is strongly relevant to several weather variables, including the outdoor temperature, outdoor humidity, and solar irradiance. It is also strongly related to

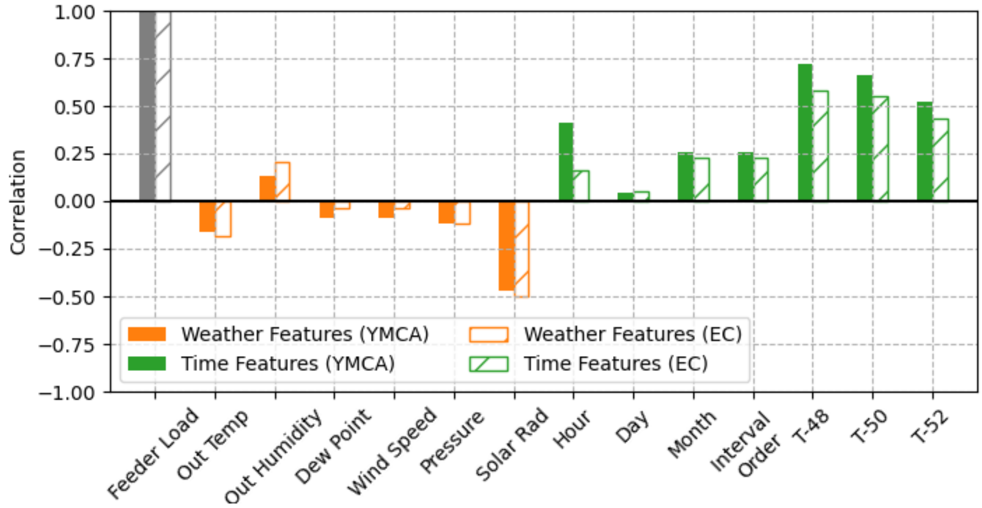

Figure 3.2: Summary of feature correlations

the past load profiles, but their correlation rapidly decreases with the lag time increases. This work chooses the ten most significant prediction features. They are outdoor temperature, humidity, solar irradiance, hour of the day, the month of the year, and the order of time intervals, 48-hour, 50-hour, and 52-hour ahead lag values. The weather data is the historical measurements (i.e., actual values) from local weather stations. However, the perfect weather forecast is unattainable in reality. Therefore, different weather forecast scenarios (i.e., perfect weather forecast and naive weather forecast) are simulated in predictions, detailed in Section 3.4.1.

Before the selected features are fed into prediction models, net load profiles and prediction features are scaled between 0 to 1. After the prediction, the result will be transformed back to the original scale. Feature scaling will help reduce the training time by preventing NNs from converging slowly.

3.3 Principles of model designs

This research conducts day-head net load probabilistic forecasting, in other words, predicting 1×48 future time steps of net loads. The section illustrates two kinds of

ML model designs, additive GPR and deep-learning probabilistic forecasting models. For additive GPR, the training data is all the selected features in the past 7 days (i.e, 7×48 time steps). The deep learning NN model is constructed as a sequence-to-sequence prediction model and trained by a dataset for half a year. Then NNs use a sequence of all features in the past seven days as inputs to predict the net load the next day.

3.3.1 Additive Gaussian process

GPR is a non-parametric process that can inherently produce a probabilistic result. An underlying assumption of GPR is that any continuous function can be modeled as the combination of covariates and their covariance, also known as the positive-semidefinite kernel functions. The kernels of a GPR model can be added or multiplied and still be valid kernels. The additive GP model is used to predict the load profile considering multiple features, and this research does not consider minor high-order interactions between covariates [151].

GPR describes the distribution of a non-linear function $f(x)$. Given the input data x, x', the distribution is defined by the mean $\mu(x)$ and the covariance $k(x, x')$.

$$f(x) \sim \mathscr{GP}(\mu(x), k(x, x')) \tag{3.1}$$

The kernel function $k(x, x')$ with hyper-parameters θ defines the covariance between any two realizations from the distribution $f(x)$ and $f(x')$,

$$k(x, x'|\theta) = cov(f(x), f(x')) \tag{3.2}$$

If the input data includes all prediction features $X_1 = (x_1, x_2, ..., x_D)$, their function values $f(X) = (f(x_1), f(x_2), ..., f(x_D))^\top$ should follow the joint multivariate Gaussian distribution. In the GPR, the mean function $\mu(x)$ is usually set to be zero, and thus the prior distribution is given by,

$$f(X) \sim N(0, K_{x,x}(\theta)) \tag{3.3}$$

where $K_{x,x}(\theta)$ is a $D \times D$ matrix and each element represents the kernel between $f(x_i)$ and $f(x_j)$ as,

$$[K_{x,x}(\theta)]_{i,j} = k(x_i, x_j|\theta) \tag{3.4}$$

Similarly, given the training data X_1 and the test data X_2, their functions, $f(X_1)$ and $f(X_2)$, follow the joint multivariate Gaussian distribution,

$$\begin{bmatrix} f(X_1) \\ f(X_2) \end{bmatrix} \sim N(0, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}) \tag{3.5}$$

where $\Sigma_{11} = K_{X1,X1}$, $\Sigma_{12} = K_{X1,X2}$, $\Sigma_{22} = K_{X2,X2}$ and $\Sigma_{21} = K_{X2,X1}$. The posterior predictive distribution based on the prior observations $X1$ is given by,

$$p(f(X_2)|f(X_1), X_1, X_2, \theta) \sim N(\tilde{f}(X_2), cov(f(X_2))) \tag{3.6}$$

where its mean function and covariance are,

$$\tilde{f}(X_2) = (\Sigma_{11}^{-1}\Sigma_{12})^\top f(X_1) \tag{3.7}$$

$$cov(f(X_2)) = \Sigma_{22} - (\Sigma_{11}^{-1}\Sigma_{12})^\top \Sigma_{12} \tag{3.8}$$

For each kernel, the hyper-parameter values θ, such as variance σ^2 (scaling factor for the deviation from the main function) and length scale l (control the smoothness) need to be decided. Based on the approximate Bayesian inference technique [152], the point estimation of the optimal hyper-parameters is conducted by maximizing the log marginal likelihood.

$$\begin{aligned} \log(p(f(X)|X, \theta) = &-\frac{1}{2} f(X)^\top K(X,X)^{-1} f(X) \\ &-\frac{1}{2}\log|K(X,X)| - \frac{n}{2}\log 2\pi \end{aligned} \tag{3.9}$$

Since a nonnegative sum of kernel functions is still a kernel function, instead of using one universal kernel for all features, this research employs a sum of N kernels $\sum_{n=0}^{N} K_{x,x}^{n}(\theta_n)$ to encode the relationships between different features. One could even multiply the kernel with sparse matrices to select the active dimension (i.e. features) for kernels. Thus, similar to (3.3), its joint Gaussian distribution is written as,

$$f(X) \sim N(0, \sum_{n=0}^{N} K_{x,x}^{n}(\theta_n)) \tag{3.10}$$

The predictive mean function and covariance (3.7) and (3.8) also change accordingly. However, the critical question is how to design and select the optimal kernel for each feature. The following paragraphs demonstrate the main procedure for designing and selecting the optimal kernel, and the proposed GPR model is built using GPflow.

This research uses a variety of kernels, including linear, square exponential, and periodic kernels. A detailed description of these kernels can be consulted in ref. [152]. The letter K denotes the kernel, and the subscript denotes the type of kernels. The feature in the bracket is the active dimension (i.e., feature). For example, a summation of kernels for time and weather features is given by,

$$K(F) = K(F_t) + K(F_w) \tag{3.11}$$

Time features include the hour of the day, the month of the year, the order of time intervals, and lag variables. First, to capture non-linear trends of these time series, the model applies the square exponential (SE) kernel K_{se} on the hour of the day, the month of the year, and the order of time intervals, respectively. To capture daily and weekly periods of load profiles, a multiplication of two periodic kernels, K_{per1} and K_{per2} are used. Their periods (i.e., kernel length scale) are set to be 48 and $7 \times 48 = 336$, respectively. For the lag variables, a linear kernel K_{lag} encodes the relationship between the past and present values. The kernel expression for time features is summarized as,

$$\begin{aligned} K(F_t) = K_{se}(F_{hour}, F_{month}, F_{order}) + K_{linear}(F_{lag}) \\ + K_{per1}(F_{hour}) \cdot K_{per2}(F_{month}) \end{aligned} \tag{3.12}$$

There are three selected weather features: outdoor temperature, humidity, and solar irradiance. As the exact relationships between these features are uncertain, kernels with different levels of smoothness (i.e., how fast the regression function can vary) are considered, and the optimal combination is selected by repetitive training and cross-validation.

$$K(F_w) = K_{x1}(F_{temp}) + K_{x2}(F_{humid}) + K_{x3}(F_{solar}) \tag{3.13}$$

Two types of kernels, linear and SE kernels are used to represent a steady and non-linear change, respectively. These options are considered for each weather feature. Therefore there are 2^{N^F} combinations where N^F denotes the number of weather features.

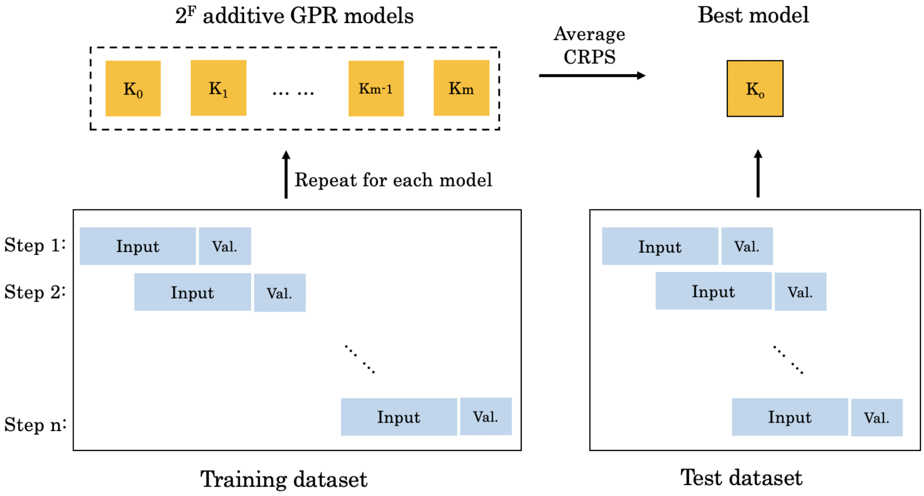

Figure 3.3: Procedure for optimal kernel selection (Val.: the validation dataset)

To select the best model among 2^{N^F} combinations, training and validation are conducted to evaluate the performance of each combination. As shown in Figure 3.3, a total length of $32 \times 48 = 1536$ time intervals are selected from the dataset. It is divided into the training and test parts with a ratio of 0.75: 0.25. The training part is segmented into the sliding window. In each window, the first $7 \times 48 = 336$ intervals are inputs to predict the rest of 1×48 intervals for result validation. This process is repeated for each combination. The score of each kernel combination is calculated by the average continuous rank probability score (CRPS). Given the predictive net load distribution $F(t)$ and actual observation y, the general form of CRPS is defined as,

$$\mathrm{CRPS}(F,y) = \int_{-\infty}^{\infty} [F(t) - H(t-y)]^2 dx \tag{3.14}$$

Where $H(t-y)$ denotes the Heaviside function, which takes the value 0 when $t < y$ and the value one otherwise. CPRS calculates the Euclidean distance between the CDFs of predictions and observations, similar to MSE for deterministic predictions. Figure 3.4 gives a schematic representation of the CRPS function. Since GPR outputs a standard Gaussian distribution, CPRS for a Gaussian distribution $\mathscr{N}(\mu, \sigma^2)$ with the mean μ and variance σ^2 is calculated as follow [154],

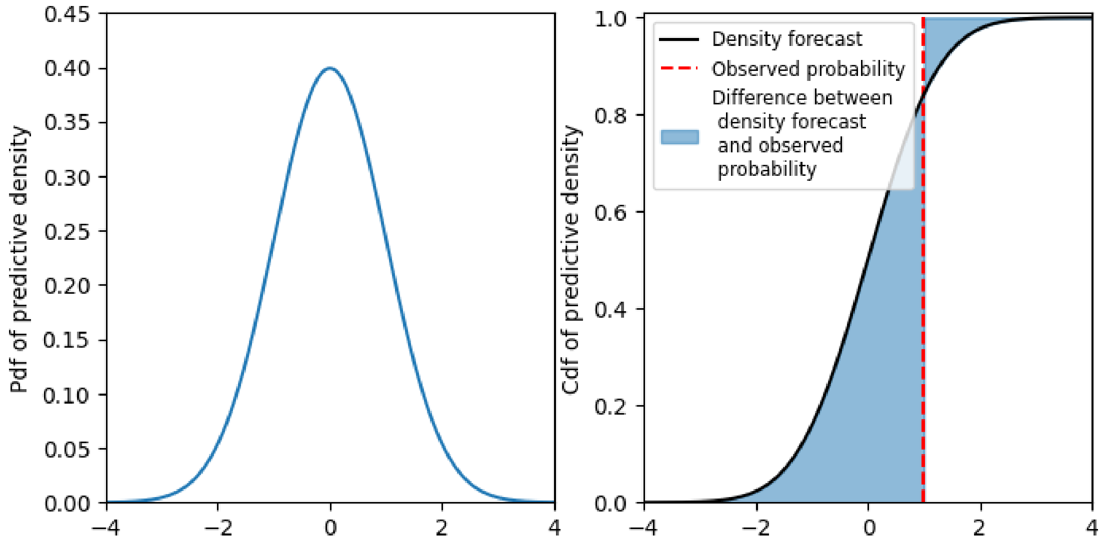

Figure 3.4: Schematic diagram of CRPS function

$$\mathrm{CRPS}_{normal}(\mathscr{N}(\mu, \sigma^2), y) = \sigma\{\frac{y-\mu}{\sigma}[2\phi(\frac{y-\mu}{\sigma}) - 1] + 2\psi(\frac{y-\mu}{\sigma}) - \frac{1}{\sqrt{\pi}}\} \tag{3.15}$$

Where $\phi(\frac{y-\mu}{\sigma})$ and $\psi(\frac{y-\mu}{\sigma})$ denote the PDF and CDF of Gaussian distributions, respectively. The model with the lowest CRPS is selected as the best model. After choosing the optimal kernels, this model is used to predict the next-day aggregated net-load profile.

3.3.2 Deep learning probabilistic prediction models

For NN predictors, MSE is often used as the loss function in deterministic prediction. The probabilistic prediction uses other loss functions, such as the pinball loss or log-likelihood functions. This section demonstrates two kinds of probabilistic prediction models based on deep NNs, namely, quantile and deep ensemble NNs.

First, the whole dataset includes $180 \times 48 = 8760$ time intervals, and they are divided into three parts, the training, validation, and test sets, in a ratio of 0.7: 0.2: 0.1. Therefore, three parts have 6132, 1752 and 876 time intervals respectively. The training and validation parts are used to train ML models based on the loss function. After training, the model is performed on the test data. The test data is further segmented into the sliding window for day-ahead forecasting, and each window has a total length of $(7+1) \times 48 = 384$ intervals. The first $7 \times 48 = 336$ intervals are the input sequence to predict the following 48 future observations (i.e., the label sequence).

Table 3.1: The architecture and parameter selection of CNN-LSTM model

Layers	**Parameters**	**Values**
Lambda	Input sequence length	240
	Batch size	(**32**, 64, 128)
1D Convolutional	Filters	(32, **64**, 128)
	Kernel size	(3, **4**, 5, 6)
	Activation function	ReLU
LSTM	Cell number	(40, 60, 80, **100**, 120)
Dense	-	-
	Epoch	(10, **20**, 30)
	Learning rate	(0.01, **0.001**, 0.0001)

A hybrid ML structure is used, namely, the CNN-LSTM model. Specifically, CNN extracts effective features from given grids or matrices. LSTM belongs to the RNN architecture and is capable to learn long-term dependencies by avoiding the vanishing gradient problem of standard RNNs [155]. This hybrid structure can takes advantage of the respective merits of CNN and LSTM. It has been applied in the short-term predictions with spatial-temporal features, for example, the load profile prediction [59], and solar irradiance prediction [156] for multiple locations.

In this work, this ML structure is used to extract important features and reduce the training time. The proposed model consists of four hidden layers, lambda, 1-D convolutional, LSTM, and dense layer. Table 3.1 presents the model architecture and detailed parameters. The lambda layer takes a feature sequence as inputs and truncates them into a new dataset. Its sample length is the same as the kernel size in the next convolutional layer. The dataset is then convolved by 1D-CNN at each time step, to extract and map important features. LSTM network takes the processed sequential data and makes predictions. The dense layer is fully connected to its previous layer and concludes the result. The proposed deep learning models are built in Python Tensorflow [157]. The parameters of each layer and the training process are selected using the grid search. The tested values for each parameter are listed in Table 3.1, and the selected ones after grid search are highlighted in **bold**.

Quantile deep NN

First, a quantile deep NN is developed based on the CNN-LSTM structure. Its loss function is the pinball loss function. Let the time series data of load profile l_t be a realization of a stochastic process L_t. The PDF of L_t is predicted using Q quantiles. The specific loss function of NN for each quantile $q \in Q$ is given by,

$$\tilde{l}_t = \underset{\tilde{l}_t}{\operatorname{argmin}} \sum_{t=0}^{T} \rho(q, \varepsilon_t) \tag{3.16}$$

where the calculated forecast error denotes as $\varepsilon_t = l_t - \tilde{l}_t$ at time t, and the pinball loss $\rho(q, \varepsilon_t)$ is calculated as,

$$\rho(q, \varepsilon_t) = \begin{cases} q\varepsilon_t, & \text{for } \varepsilon_t \geq 0 \\ (q-1)\varepsilon_t, & \text{for } \varepsilon_t < 0 \end{cases} \tag{3.17}$$

ML model is trained for each quantile $Q =: \{q_0, q_1, ..., q_n\}$ with different loss functions (3.16) independently and then predicts the net load profile. Eventually, all quantiles are compiled as an outcome.

Ensemble deep NN

Another kind of probabilistic prediction model is density NN, which has been introduced in Chapter 2. Considering the stochastic nature of NNs, this research adopts an ensemble density NN structure proposed first in [94] named *deep ensemble model*. It generates a standard Gaussian probabilistic prediction and adopts an ensemble learning framework to enhance accuracy.

The same CNN-LSTM structure is used, but its output layer and loss function are modified. The output layer has two outcomes, the predictive mean $\mu(x)$ and variance $\sigma^2(x)$, given the input feature vector x. The loss function is a negative log-likelihood function given by,

$$l = -log_p(y|x) = \frac{log\sigma^2(x)}{2} + \frac{(y-\mu(x))^2}{2\sigma^2(x)} + const. \tag{3.18}$$

A constant number is added to the loss function to keep the numerical value of the loss positive. For each NN, the initial weights of neurons are random, which leads to uncertain results.

$$\mu(x) = \frac{\sum_{n=0}^{M} \mu_n(x)}{M} \tag{3.19}$$

$$\sigma^2(x) = \frac{\sum_{n=0}^{M} \mu_n^2(x) + \sigma_n^2(x)}{M} - \mu^2(x) \tag{3.20}$$

Five density NNs are trained independently and are consisted of an ensemble framework. The final mean (4.6) and variance (3.25) of the predicted distribution are calculated from the results of all five density NNs.

3.4 Numerical experiments

3.4.1 Experiment settings

This work designs a series of experiments to compare additive GPR and deep learning models. The experiments include three cases using different sets of prediction features or based on different weather forecasting assumptions. They are summarized as below.

- Case 1: Probabilistic prediction using time features only (i.e., F_t).
- Case 2: Probabilistic prediction using time and weather features (i.e., $F_w + F_t$). The optimal kernel selection is conducted for three weather variables. The input weather forecasts are perfect over the prediction horizon (i.e., perfect weather forecast).
- Case 3: Probabilistic prediction using time and weather features (i.e., $F_w + F_t$). The optimal kernel selection is conducted for three weather variables. The input weather forecasts are historical weather in the previous day (i.e., naive weather forecast).

Each experiment is tested on two substations (YMCA and EC). All experiments are conducted on a MacBook Pro equipped with 8 GM memory and an Inter Core i5 processor at 3.1GHz.

3.4.2 Scoring rules

Compared to deterministic forecasting, probabilistic prediction requires not only prediction accuracy, but also result reliability (i.e., the degree to which the frequency of actual outcome matches the forecasted probability) and sharpness (i.e., the concentration of the predictive distributions). These properties need to be evaluated using a set of dedicated scoring rules. In this research, prediction results are evaluated from three perspectives - prediction accuracy, result reliability, and sharpness. Three scores including MAE, CRPS and Winkler scores, are selected to examine these properties respectively. These scores are divided by the full range of net load R_{nl} to obtain the normalized percentage values so that the prediction result for different substations with different power capacities can be compared.

$$R_{nl} = \max(y_t) - \min(y_t) \tag{3.21}$$

MAE examines prediction accuracy. It calculates the deviation between the point prediction and the actual value. Given the prediction $\tilde{y}_t$ and the actual value y_t, the normalized MAE (nMAE) is given by,

$$\text{MAPE} = \frac{1}{T}\sum_{t=0}^{T}\frac{|\tilde{y}_t - y_t|}{R_{nl}} \tag{3.22}$$

CPRS examines prediction reliability as in (3.14). The normalized CPRS (nCPRS) over the prediction horizon T is given by,

$$\text{nCRPS} = \frac{1}{T}\sum_{t=0}^{T}\frac{\text{CRPS}(\tilde{F}_t, y_t)}{R_{nl}} \tag{3.23}$$

Notably, CPRS is not calculated for predictions from the quantile deep NN model, as its underlying distribution is unknown.

The winkler score examines both prediction reliability and the sharpness of interval. The score calculates the width of prediction intervals and penalizes predictions out of the interval. Given a prediction interval between $100\alpha\%$ and $100(1-\alpha)\%$ at time t, and $u_{\alpha,t}$ and $l_{\alpha,t}$ are the upper and lower bounds of an interval, the winkler score $W_{\alpha,t}$ is calculated as,

$$W_{\alpha,t} = \begin{cases} (u_{\alpha,t} - l_{\alpha,t}) + \frac{2}{\alpha}(l_{\alpha,t} - y_t) & \text{for } y_t < l_{\alpha,t} \\ u_{\alpha,t} - l_{\alpha,t} & \text{for } l_{\alpha,t} \le y_t \le u_{\alpha,t} \\ (u_{\alpha,t} - l_{\alpha,t}) + \frac{2}{\alpha}(y_t - u_{\alpha,t}) & \text{for } y_t > u_{\alpha,t} \end{cases} \tag{3.24}$$

This research chooses a prediction interval between 10% to 90% (i.e. $\alpha = 10\%$). The normalized winkler score is thus given by,

$$nW_{\alpha\%} = \frac{1}{T}\sum_{t=0}^{T}\frac{W_{\alpha\%,t}}{R_{nl}} \tag{3.25}$$

3.5 Results and discussions

3.5.1 Predictions using time features

First, probabilistic predictions are performed using time features only (i.e., F_t). Figures 3.5 and 3.6 show half-hour day-ahead load prediction results during a week in two substations, YMCA and EC, respectively.

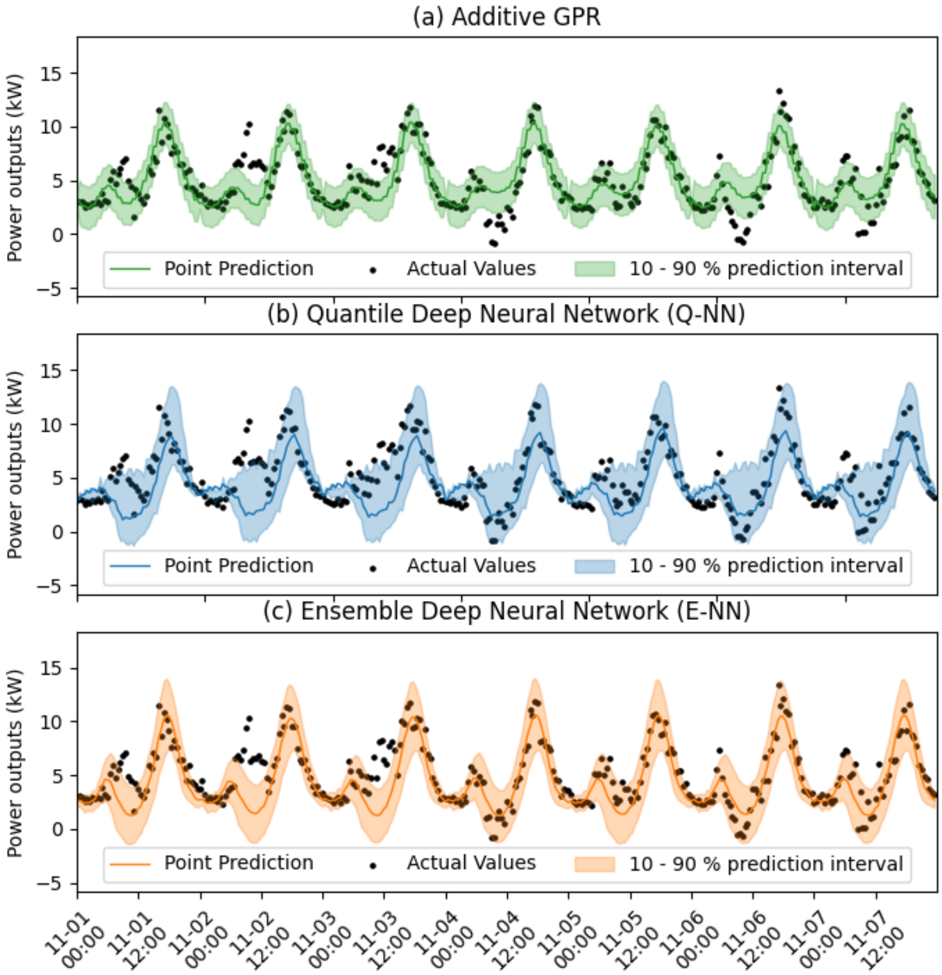

Figure 3.5: Day-ahead prediction results at YMCA using time features (The solid line shows the point prediction; The shade shows the 10-90% confidence interval)

All three models can capture the periodicity of the aggregated net load profiles. Their predictions for the YMCA substation (Figure 3.5) show two peaks in the early morning and late afternoon respectively, and predictions for the EC substation (Figure 3.6) show a deep valley in the load profile around the noon due to the exported solar power. For the underlying distribution, Q-NN gives an asymmetric prediction interval, while the other two predictors generate a standard Gaussian distribution. The kernel design of GPR can capture the dynamics within different time scales. However, additive GPR cannot capture weather-dependent changes in net load profiles in Case 1. For example, in Figure 3.6 (a), the low valleys of net load profiles due to the exported solar power cannot

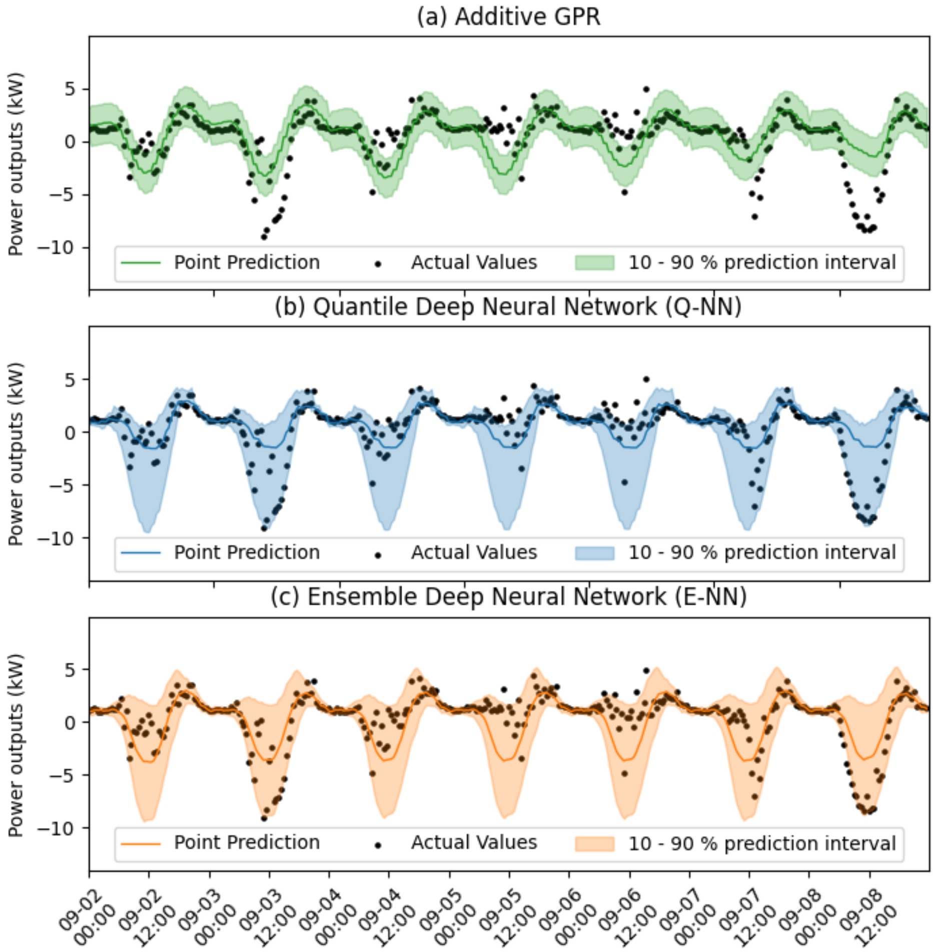

Figure 3.6: Day-ahead prediction results at EC using time features (The solid line shows the point prediction; The shade shows the 10-90% confidence interval)

be predicted, leading to unreliable results on particular days. Deep learning models in Figures 3.6 (b) and (c) can predict these low valleys in net load profiles, but their prediction intervals are much wider and conservative.

3.5.2 Predictions using time and weather features

Then predictions are performed using both time and weather features (i.e., $F_w + F_t$) in Cases 2 and 3, with different weather forecast assumptions. In terms of input data, deep learning models take all the features directly for training, while the optimal kernel selected is conducted for three weather features, outdoor temperature, outdoor humidity, and solar irradiance, in additive GPR. The best combination of kernels for the YMCA substation is SE, SE, and linear kernels. The best combination for EC substations is SE, SE, and SE kernels. The detailed CPRS for each combination in the training process is presented in Appendix B.

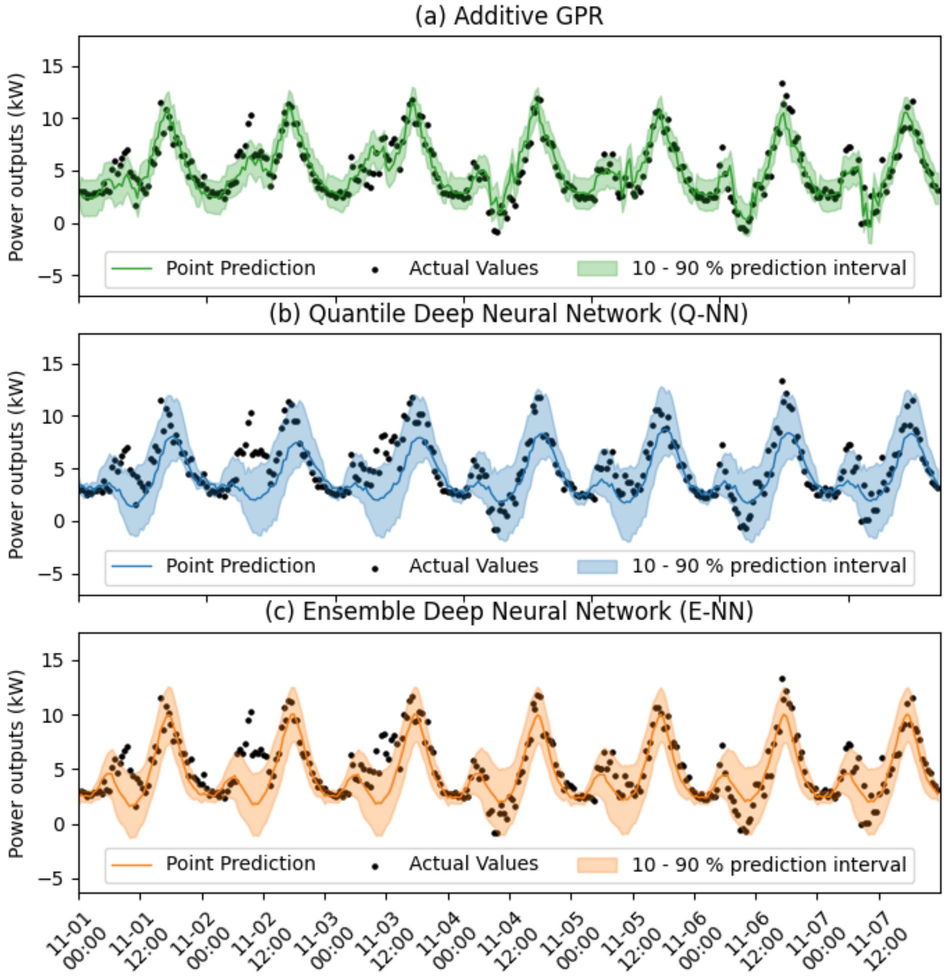

Figure 3.7: Day-ahead prediction results at YMCA using the full set of features (The solid line shows the point prediction; The shade shows the 10-90% confidence interval)

Case 2 is performed first with the perfect weather forecast. Figures 3.7 and 3.8 show results under the perfect weather during a week in YMCA and EC substations respectively. Results show additive GPR model can accurately capture the weather-dependent changes in net load profiles. For example, predictions in Figures 3.7 and 3.8 (a) capture a great amount of the exported power at the noon due to high solar irradiance. Two deep learning models show slight improvements, such as narrower

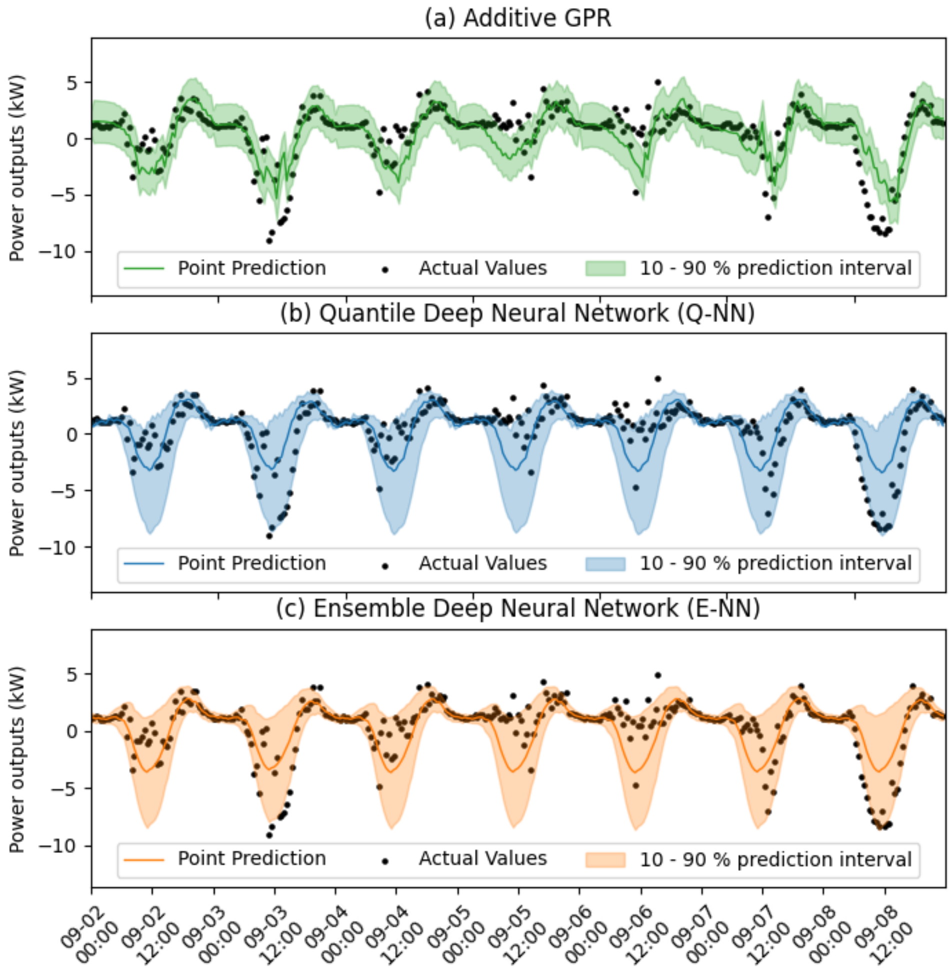

Figure 3.8: Day-ahead prediction results at EC using the full set of features (The solid line shows the point prediction; The shade shows the 10-90% confidence interval)

prediction intervals. However, they are generally less sensitive in capturing weather-dependent changes compared to additive GPR.

Case 3 is then performed on additive GPR with the naive weather forecast, which assumes the weather the next day is the same as the previous day. Figure 3.9 shows probabilistic forecast results at the YMCA substation under two different weather forecast assumptions, the perfect and naive weather forecasts. It can be observed that the weather forecast has a significant impact on netload predictions from additive GPR, especially around noon when the solar power output is the highest.

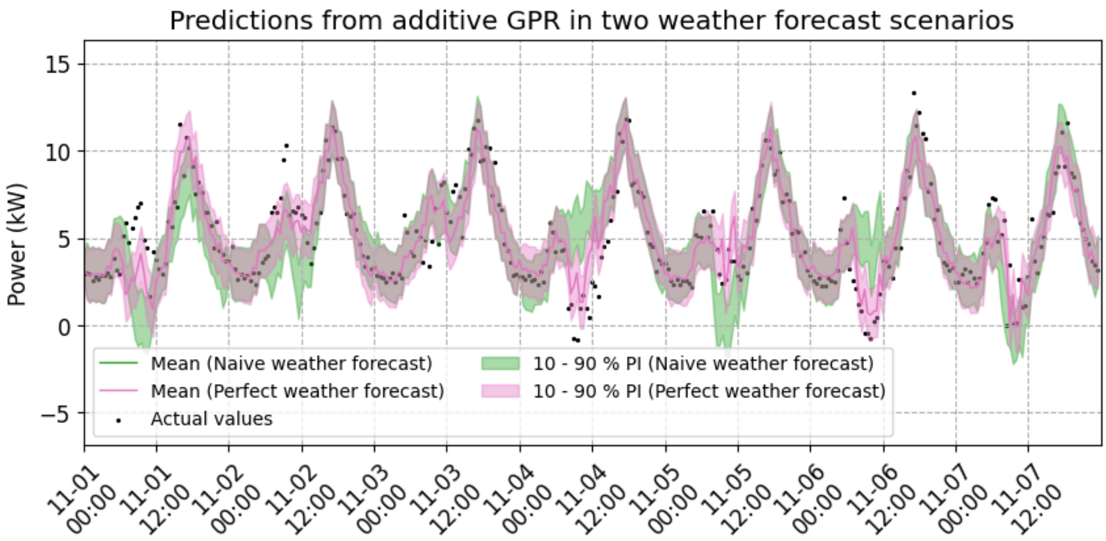

Figure 3.9: Predictions under two different weather forecast assumptions, the perfect and naive weather forecasts

3.5.3 Comparisons of prediction performances

Probabilistic evaluation scores (i.e., nMAE, nCPRS, and nWS) are calculated for all experiments, which are listed in detail in Appendix B. Figures 3.10 and 3.11 shows the simulation results in Cases 2 and 3 benchmarked with 'no weather' case respectively. The bar charts show the average prediction performances of two substations, and results for the YMCA and EC substations are plotted in black and red lines, respectively.

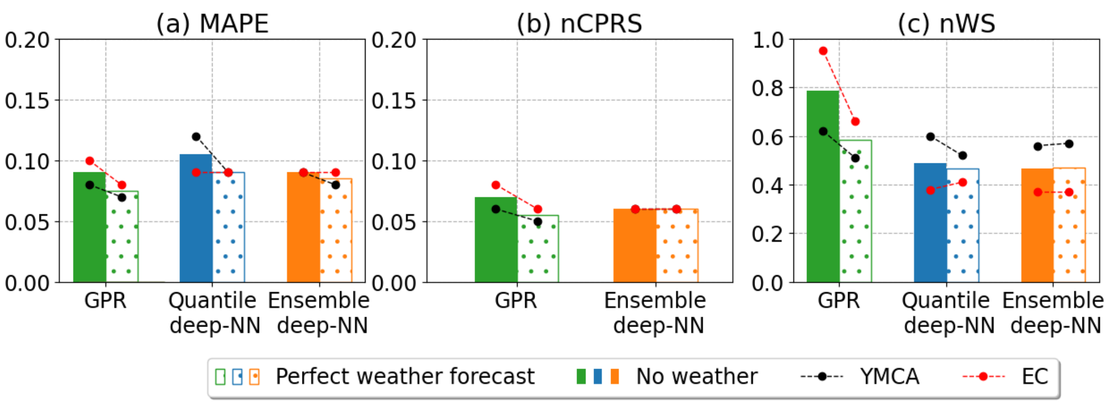

Figure 3.10: Summary of evaluation scores in Case 2 under the perfect weather forecast assumptions (The bars show the average prediction performance, and results for YMCA and EC substations are plotted in black and red lines respectively)

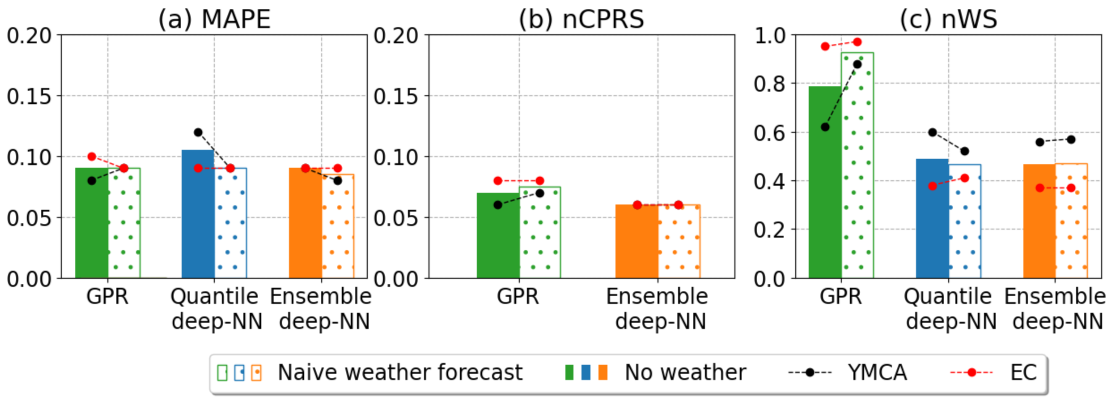

Figure 3.11: Summary of evaluation scores in Case 3 under the naive weather forecast assumptions (The bars show the average prediction performance, and results for YMCA and EC substations are plotted in black and red lines respectively)

In summary, additive GPR is sensitive to weather-dependent changes in net load profiles. After the optimal kernel selection for three weather variables, additive GPR can achieve a very accurate prediction under the perfect weather information (Figure 3.10). However, its prediction accuracy will significantly reduce if the weather forecast accuracy reduces (Figure 3.11). Under naive weather forecast assumptions, their prediction performances are worse than predictions from deep learning models. On the contrary, two deep learning models as 'black-box' predictors are influenced less by weather features.

3.6 Concluding remarks

This chapter designs and compares additive GPR and deep learning models for short-term probabilistic net load prediction using time and weather features. Their performances are simulated under different weather forecast scenarios and evaluated using probabilistic scores.

For the additive GP model, the optimal kernel selection is conducted for each weather variable, while the other two probabilistic deep learning models process all features without differences. The proposed additive GP model with the best kernel combination can capture weather-dependent changes in net load profiles. It yields significantly better performance when including the perfect weather forecast. However, the performance accuracy reduces significantly under the naive weather forecast. On the contrary, the other two deep learning models are less sensitive to weather-related changes.

In terms of the computation costs, the kernel selection procedure is a factorial test for features, and the number of the kernel combinations increases with both the number of feature and kernel options exponentially (i.e., the problem scale is of $\mathscr{O}(2^N)$ if each feature has two kernel options). In this case, deep learning models show more advantages to incorporating the high-dimensional feature data.

Future work could extend to designing an ensemble framework incorporating both additive GPR and deep learning NNs, which aims to secure accurate prediction performances under various weather forecast scenarios. Another extension is using different types of net load probabilistic forecasting to simulate dynamic reserve requirements in the power network. These probabilistic forecasting will lead to different reserve provision costs when considering different system constraints, such as the ramp-up or ramp-down constraints of generators.

4

Distributionally Robust Joint Chance-constrained Optimization for Networked Microgrids Considering Renewable Uncertainty

Contents

Chapter Summary: *In light of a reliable and resilient power system under extreme weather and natural disasters, networked microgrids integrating local renewable resources have been adopted extensively to supply demands when the main utility experiences blackouts. However, the stochastic nature of renewables and unpredictable contingencies are difficult to address with the deterministic energy management framework. The chapter proposes a comprehensive DRJCC framework that incorporates*

microgrid island, power flow, distributed batteries, and voltage control constraints. The interruptions of utility supply are addressed by considering two operational modes of microgrids, grid-connected and island modes. All the CCs are solved jointly, and each one is assigned to an optimized violation rate. To highlight, the JCC problem with the optimized violation rates has been recognized to be NP-hard and challenging to be solved. This chapter proposes a novel evolutionary algorithm that successfully tackles the problem and reduces the solution conservativeness (i.e., operation cost) by around 50% compared with the baseline Bonferroni Approximation. Considering the imperfect solar power forecast, this chapter constructs three data-driven ambiguity sets to model uncertain forecast error distributions. The solution is thus robust for any distribution in sets with the shared moment and shape assumptions. The proposed method is validated by robustness tests based on those sets and firmly secures the solution robustness.

4.1 Nomenclature

A.	*Set and index*
$t \in \mathscr{T}$	Index, set of timesteps
$b \in \mathscr{B}$	Index, set of buses
$s \in \mathscr{S}$	Index, set of distributed assets (i.e., storages, PV panels, loads)
B.	*Parameters*
N_B	Number of network buses
N_L	Number of lines
N_S	Number of distributed batteries
N_D	Number of loads
m_u	Cost factor of grid power [\$/W^2]
m_s	Cost factor of solar generation curtailment [\$/W^2]
m_l	Cost factor of load curtailment [\$/W^2]
m_r	Cost factor of droop control provision [\$/W^2]
m_d	Battery degradation costs [\$/W]

N_{PV}	Number of solar PV panels
N_{it}	Maximum number of evolution iterations
N_c	Number of SCCs
N_s	Number of forecast error samples
N_p	Number of individuals in one generation
η_{dis}, η_{ch}	Battery discharging / charging efficiency
$\overline{v}, \underline{v}$	Maximum / minimum bus voltage [V]
$\overline{SoC}, \underline{SoC}$	Maximum / minimum state of charge (SoC)
$\overline{P^c}, \overline{P^d}$	Maximum discharging / charging power [W]
$\overline{E^s}$	Battery energy capacity [Wh]
P^{pv}	Forecast solar power at each bus [W]
$P^{\tilde{pv}}$	Solar power forecast errors
P^l	Demands at each bus [W]
P^{cl}	Critical loads at each bus [W]
$\overline{\varepsilon}, \underline{\varepsilon}$	Upper / lower bounds of violation rates of CCs
ε_j	Violation rate of JCCs
$\mathbf{V}_o, \mathbf{I}_o$	Voltage / current operation point vector [W / A]
r_{thr}	Threshold ratio for the termination condition
$\mathbf{Y}$	Network admittance [S]
$\mathbf{C}$	Sparse matrices to map the distributed assets
$\mathbf{P}$	Individuals (population) in one generation

C.	*Variables*
P^d, P^c	Battery discharging / charging power [W]
ε_i	Violation rate of SCCs
μ_{pv}	Mean vector of solar forecast errors
Σ_{pv}	Covariance matrix of solar forecast errors
d^s	Droop provision coefficients
v	Voltage magnitude of buses [V]
$P^{pv'}$	Consumed solar power at each bus [W]
$P^{l'}$	Supplied loads at each bus [W]
P^{inj}	Net injected power at each bus [W]
P^g	Imported grid power [W]
R^{up}	Upward droop provision from batteries [W]
E^s	SoC of distributed batteries
r	Ratio to measure the solution convergence

4.2 Introduction

The reliability of power systems under the impact of increasing renewable penetration and extreme weather conditions is a rising concern. In the developed world, the decreasing number of dispatchable fossil-fuel power plants and reduced system inertia render the power system more vulnerable to natural disasters [158]. Recent examples include the rolling blackouts across California due to wildfires [159], and disastrous power outages in Texas due to the extremely cold weather [160]. In the developing world, increasing electricity demand and aging infrastructure result in frequent power outages. In SSA countries, the outage time of public utilities is commonly around 10% of the total operational time, and even reaches 50% in some instances [14].

To tackle utility failures and power cuts, microgrids aggregate local renewable energy resources and loads in a small network, and operate flexibly with or without the grid connection, thanks to modern inverter-based design [161]. Microgrids can thus supply loads when the grid experiences scheduled under-frequency load shedding (UFLS) [162] or unpredictable power cuts. This island capability is incorporated into the power scheduling of microgrids [163, 164]. Smart load shedding is also employed to mitigate power imbalances in microgrids. For example, refs. [165, 166] model users' utility functions using different appliances so that flexible loads can be shifted efficiently for peak shaving in distributed power networks.

On the other hand, intermittent renewable power sources such as solar power pose challenges to short-term power system operations. The imperfect forecasting brings uncertainty, which could result in network constraint violations and high power losses [112]. Stochastic and robust optimization have been proposed to address uncertainty. While robust optimization based on the worst-case scenario leads to an overly-conservative and cost-prohibitive solution, the chance-constrained (CC) formulation, as one of the predominant stochastic approaches, can directly control the system reliability to a predefined level and decide the optimal cost. The CC-OPF formulation was first proposed in [116], incorporating a series of single network CCs pertaining to voltage and power limits.

The most intuitive way to solve the CC problem is the scenario-based approach. As the exact solution to a CC problem is unattainable, this approach solves a great number of problem scenarios randomly drawn from the uncertainty distribution. To secure the estimation confidence level $1-\beta$, the number of random samples should be at least $\frac{2}{\varepsilon}(\ln\frac{1}{\beta}+n)$, given the violation rate ε and the dimension of decision variables n [167].

A more effective alternative is DRO. This approach constructs a set based on uncertainty data samples - termed the *ambiguity set* - including all possible uncertainty distributions. This formulation thus ensures constraints are satisfied for any distribution in the ambiguity set built upon distribution moments and shape information. The problem can be solved by being recast into tractable formulations, including LP, SDP and SOCP, depending on the degree of approximation [168]. However, defining an ambiguity set to characterize uncertain distributions is non-trivial, as one needs to decide the trade-off between solution robustness and conservativeness while considering the mathematical tractability. Compared with early works using the first two moments (i.e., mean and variance) such as [169], recent works utilize the high-order moments (e.g., skewness [135]), structural properties (e.g., unimodal [123] and symmetric [137]) to set tighter bounds. Another kind of method is the moment-free method. Ref. [170] constructs a ball space where possible distributions are centred at the reference distribution based on the training samples, and the ball radius is defined by Wasserstein-based distance metrics. However, such an approach is highly data-intensive and its performance is substantially influenced by the volume of data available [171].

All aforementioned literature [169, 171, 172, 135, 139, 137, 170] adopt the single CC formulation, in which each constraint is considered as an independent event with the pre-defined violation rate. However, in most power system applications, the JCC formulation is desired, which means that all constraints should be satisfied simultaneously and use one whole-system reliability metric. For example, a distribution feeder is considered to be reliable *if and only if* all the constraints, such as bus voltage limits, and power balance are met simultaneously.

However, solving the JCC problem is notoriously difficult since its reformulation generally results in intractable problems [173]. Only a few papers attempted to solve JCC problems using either the scenario-based or approximation-based methods. The scenario-based method, following the aforementioned principles, solves possible scenarios from historical samples to JCC problem. Ref. [127] first proposed the JCC-OPF formulation for the transmission network with high wind power penetration. A droop-type function, termed the distribution vector, was introduced to control generators concerning wind power forecast errors [174, 127]. These problems were solved in a great number of wind power forecast scenarios. Similar works include [175, 128, 176] for power networks integrating flexible loads or thermal storages. In general, the scenario-based approach offers a fairly accurate solution given a large volume of samples, but its scalability heavily relies on statistical techniques, such as sample average estimation, to ease the computation burden [130].

The approximation-based method is decomposing an intractable JCC problem into a series of tractable SCCs, then approximating individual violation rates. The simplest approximation method, termed the *Bonferroni Approximation*, assumes that all individual violation rates are the same and equal to the joint violation rate divided by the number of individual constraints, proposed first in [131]. This approximation has an extremely conservative assumption that neglects the intersections of constraints and treats all constraints equally. In this case, the solution conservativeness increases with the number of individual CCs [177]. To reduce the solution conservativeness, ref. [124] identifies all intersections of constraints using machine learning classification and obtained around 5% result improvement, compared to the *Bonferroni Approximation* as the baseline. Refs. [132, 133] approximate the JCC to CVaR constraints and introduce a scaling factor to control the tightness of the approximation. The improvement benchmarked against the *Bonferroni Approximation* is around 8-12%. However, none of the previous papers try to allocate the optimized violation rates for each constraint. Ref. [124] concludes these optimal violation rates are challenging to find. Furthermore, ref. [134] theoretically proves that a JCC problem with the optimized individual violation rates, termed the *optimized Bonferroni Approximation*, is a strongly NP-hard problem.

This chapter makes the following contributions which together address the challenges above:

1) A novel evolutionary algorithm is proposed to solve the JCC problem with optimized individual violation rates, which is an NP-hard problem and challenging to solve. Our method shows around a 50% reduction in the solution conservativeness (i.e., operation cost) benchmarked against the *Bonferroni Approximation*. This performance is the best to date compared to other approximation-based methods for the JCC problem. Moreover, these optimized violation rates are interpretable, accurately reflect the sensitivities of corresponding constraints to the operational cost.

2) The proposed JCC formulation is tested on three data-driven ambiguity sets, namely, symmetrical, unimodal and symmetrical & unimodal sets. These ambiguity sets are created and constructed using the empirical solar power forecast errors from a machine learning model to capture accurate statistical characteristics of uncertainty distributions in each time interval.

3) This DR-JCC energy management framework for the networked microgrid incorporates CCs pertaining to the power flow, bus voltage, energy storage power, and energy limits. The power flow is simulated under three uncertain distribution assumptions (i.e., ambiguity sets) to observe the no-violation cases. Only the proposed method can schedule the system to closely meet the reliability requirements, while the SCC and the benchmark case give either unreliable or overly-conservative results.

The rest of this chapter is organized as follows. Section 4.3 presents the centralized OPF formulation for a networked DC microgrid. Section 4.4 demonstrates the essential steps to reformulate the model into a DR-JCC framework with the optimized individual violation rates, and then this research proposes a novel evolutionary algorithm to solve this intractable problem. Section 4.5 presents a statistical analysis of empirical solar forecast errors and the rationale behind three data-driven ambiguity sets. Section 4.6 presents a case study to evaluate the model performance and test the solution robustness.

4.3 Centralized OPF for networked microgrids

Figure 4.1 shows an example of a networked DC microgrid in a rural area. The microgrid has a main busbar connected to multiple households and the main grid via an inverter. The grid often experiences unpredictable power cuts. Each household at the end point has a bidirectional multi-port DC-DC converter connected to local PV panels, distributed ES, and appliances. The centralized OPF is optimized in the receding horizon with 15-min time intervals and a one-day window. The formulation considers both the grid-connected and island mode simultaneously, allowing off-grid operation at any time step. The subsequent sections present the centralized OPF formulation with predetermined forecast errors.

1) *Preliminaries:* As in Figure 4.1, the framework considers a networked microgrid with N_B buses and N_L lines. The network buses are indexed by $b \in \mathscr{B}$, and the network admittance matrix is denoted as $\mathbf{Y} \in \mathbb{R}^{N_B \times N_B}$. Distributed assets are located at different buses, including energy storage, solar power generation, and flexible and inflexible loads, indexed by $s \in \mathscr{S}$. Bold letters $\mathbf{P}_t := \{P_{1,t}, P_{2,t}, ..., P_{n,t}\}$ represent decision variable vectors across distributed assets at time $t \in \mathscr{T}$. The positions of those distributed assets in the network are mapped by sparse matrices $\mathbf{C}^{pv} \in \mathbb{R}^{N_B \times N_{PV}}$, $\mathbf{C}^{s} \in \mathbb{R}^{N_B \times N_S}$, $\mathbf{C}^{l} \in \mathbb{R}^{N_B \times N_D}$. The

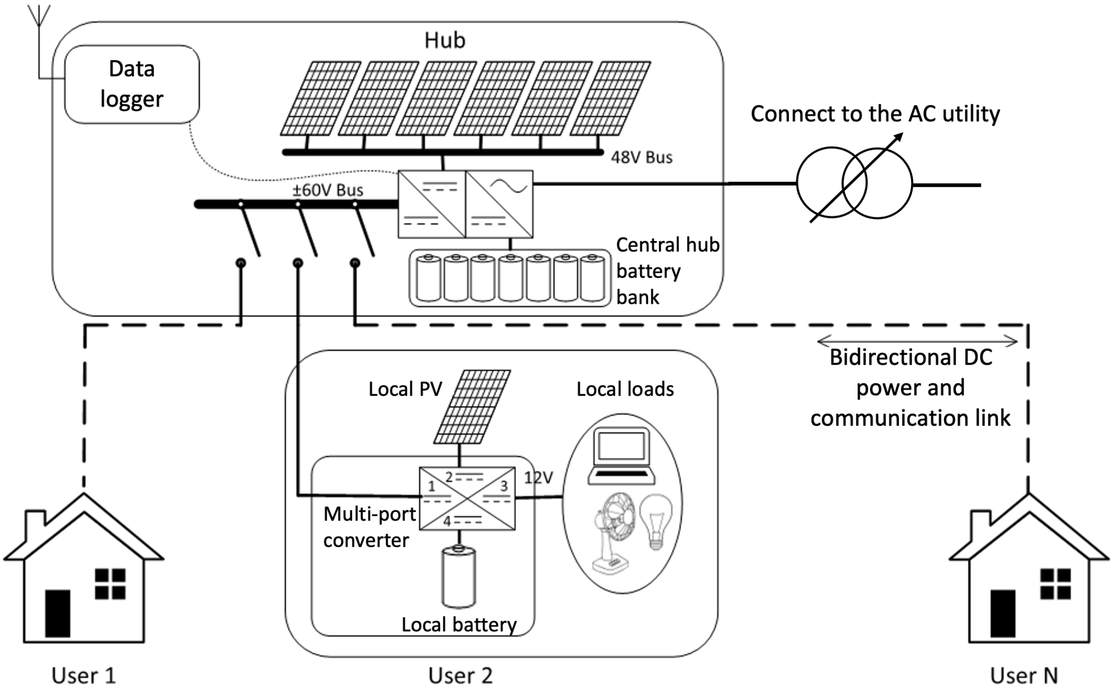

Figure 4.1: Networked DC microgrid in the rural area

multi-period centralized DC OPF has the time interval Δt. For all constraints, $\forall t \in \mathscr{T}$ and $\forall b \in \mathscr{B}$ hold unless otherwise specified.

2) *Objective function:* The objective function is formulated,

$$\begin{aligned}\mathscr{J} = \Delta t \sum_{t=0}^{T} \{ & m_u (P_t^g)^2 + m_r \sum_{s=0}^{N_S} (R_{s,t}^{up})^2 \\ & + m_s \sum_{s=0}^{N_{PV}} (P_{s,t}^{pv} - P_{s,t}^{pv'})^2 + m_l \sum_{s=0}^{N_D} (P_{s,t}^{l} - P_{s,t}^{l'})^2 \\ & + m_d \sum_{s=0}^{N_S} (P_{s,t}^{d} + P_{s,t}^{c}) \}\end{aligned} \tag{4.1a}$$

The objective function includes five terms, namely, utility tariff, droop control provision cost, solar power curtailment penalty, load shedding penalty, and battery degradation cost. Except for battery degradation costs, all costs are modeled as a quadratic function. Cost factors m_s and m_l are associated with the solar power self-consumption and users' utility, as detailed in Section V.

2) *SoC battery droop control*: Distributed ES in the network has two functions: shifting solar energy in time and providing P-V droop control to regulate bus voltages

within the range. For the voltage regulation function, distributed ES reserves energy to address positive solar power forecast errors and provide upward voltage regulation. When solar power is overestimated, it leads to a power imbalance and low bus voltages. Then each ES will release the reserved energy according to its droop coefficient and bus voltage deviation, until a new power balance is achieved. When the solar power is under-estimated, the surplus solar power will be curtailed considering the curtailment penalty in (4.1a).

$$\sum_{s=0}^{N_S} d^s_{s,t} = 1 \quad \forall d^s_{s,t} \in [0,1] \tag{4.1b}$$

$$R^{up}_{s,t} \geq d^s_{s,t} \sum_{s=0}^{N_{PV}} \tilde{P}^{pv}_{s,t} \tag{4.1c}$$

$$P^d_{s,t} + R^{up}_{s,t} \leq \overline{P^d} \tag{4.1d}$$

$$P^c_{s,t} - R^{up}_{s,t} \geq 0 \tag{4.1e}$$

$$\underline{SoC} \leq E^s_{s,t} - R^{up}_{s,t} \Delta t \tag{4.1f}$$

$$E^s_{s,t} \leq \overline{SoC} \tag{4.1g}$$

$$E^s_{s,t+1} = E^s_{s,t} - \frac{P^d_{s,t}}{\eta_{dis}} \Delta t + P^c_{s,t} \eta_{ch} \Delta t \tag{4.1h}$$

At any time step t, all distributed ES in the network should deliver the P-V droop provision to address exactly the total amount of solar power forecast errors (4.1b), while each distributed ES is coordinated to deliver a fraction known as the droop coefficient $d^s_{s,t}$ in (4.1c). Droop provision is constrained by the battery power output limits (4.1d) - (4.1e), energy constraint (4.1f) - (4.1g) and energy balance considering the round-trip efficiency (4.1h).

3) *Power flow and balance:* The users' loads are categorized into flexible and inflexible

loads, and flexible loads can be curtailed during a blackout.

$$P^{l'}_{s,t} = P^{l}_{s,t} \tag{4.1i}$$

$$0 \leq P^{pv'}_{s,t} \leq P^{pv}_{s,t} \tag{4.1j}$$

$$\langle P^{g}_{t}, P^{pv'}_{s,t}, P^{d}_{s,t}, P^{c}_{s,t}, P^{l'}_{s,t} \rangle \geq 0 \tag{4.1k}$$

$$P^{g}_{t} + \sum_{s=0}^{N_{PV}} P^{pv'}_{s,t} + \sum_{s=0}^{N_S} (P^{d}_{s,t} - P^{c}_{s,t}) = \sum_{s=0}^{N_D} P^{l'}_{s,t} \tag{4.1l}$$

At the grid-connected mode, the system cannot curtail any load (4.1i), but the solar power curtailment is allowed (4.1j). All stacked decision variables in (4.1k) should be greater than zero. The power balance is guaranteed by (4.1l).

Considering distributed ES with the droop provision, local solar power generations with forecast errors and loads, the injected power at each bus in the network is formulated as,

$$\mathbf{P}^{inj}_{t} = \mathbf{C}^{g}\mathbf{P}^{g}_{t} + \mathbf{C}^{s}(\mathbf{P}^{d}_{t} - \mathbf{P}^{c}_{t} + \mathbf{R}^{up}_{t}) + \mathbf{C}^{pv}(\mathbf{P}^{pv'}_{t} - \tilde{\mathbf{P}}^{pv}_{t}) - \mathbf{C}^{l}\mathbf{P}^{l'}_{t} \tag{4.1m}$$

The voltage at each bus depends on the power injected at that bus and the power flow between all neighboring buses,

$$\mathbf{P}^{inj}_{t} = \mathbf{diag}(\mathbf{v})\mathbf{I} = \mathbf{diag}(\mathbf{v})\mathbf{Y}\mathbf{v} \tag{4.1n}$$

$$\mathbf{P}^{inj}_{t} = \mathbf{diag}(\mathbf{V}_{o})\mathbf{Y}\mathbf{v} + \mathbf{diag}(\mathbf{I}_{o})(\mathbf{v} - \mathbf{V}_{o}) \tag{4.1o}$$

$$\underline{v} \leq \mathbf{v} \leq \overline{v} \tag{4.1p}$$

To deal with the non-convex constraint (4.1n), ref. [178] uses *Taylor series expansion* for linearization and validates its high fidelity. Based on this approach, constraint (4.1n) is linearized around the operating point $(\mathbf{V}_{o}, \mathbf{I}_{o})$ as (4.1o). Bus voltages are regulated within a certain range to ensure power quality (4.1p).

4) *Island mode:* Designed to operate during a blackout, the microgrid can island at any time step. The grid-connected and islanding schedules of the microgrid are

solved simultaneously as two scenarios of one problem. The optimization problem in the islanding scenario is identical to the grid-connected mode, except for the constraints about load curtailment and utility supply.

$$P_t^g = 0 \qquad t \in [t_o, t_o + H] \tag{4.1q}$$

$$P_{s,t}^{cl} \leq P_{s,t}^{l'} \leq P_{s,t}^{l} \qquad t \in [t_o, t_o + H] \tag{4.1r}$$

When the utility power is available, the battery SoC of these two scenarios should be the same. When the blackout happens (4.1q) during $[t_o, t_o + H]$, the microgrid is only required to supply inflexible loads (4.1r) during islanding, and the battery control reference follows the optimization result of the islanding scenario.

4.4 DR-JCC framework with the optimized individual violation rates

The centralized OPF formulation in Section 4.3 has an underlying assumption of predetermined solar power forecast errors, $P_{s,t}^{\tilde{pv}}$, while these errors follow an uncertain distribution in practice. The DR-JCC formulation is thus introduced to integrate uncertainty and secure solution robustness for uncertainty distributions. Introducing JCC formulation is essential because the microgrid is reliable *if and only if* all individual CCs are satisfied simultaneously. As shown in Figure 4.2, the procedure for formulating and solving the DRJCC problem includes four steps.

As shown in Figure 4.2, firstly, solar forecast error samples are collected. The framework then summarizes the family of error distributions and builds ambiguity sets. Given the shape and moment assumptions of ambiguity sets, a DR-JCC problem is formulated considering the *joint* risk of the battery power capacity, bus voltage, and network violation. The problem is solved by decomposing and recasting into the SOCP formulation. However, the JCC with the optimized violation rates is an intractable problem. Thus, a novel evolutionary algorithm is proposed to solve this problem.

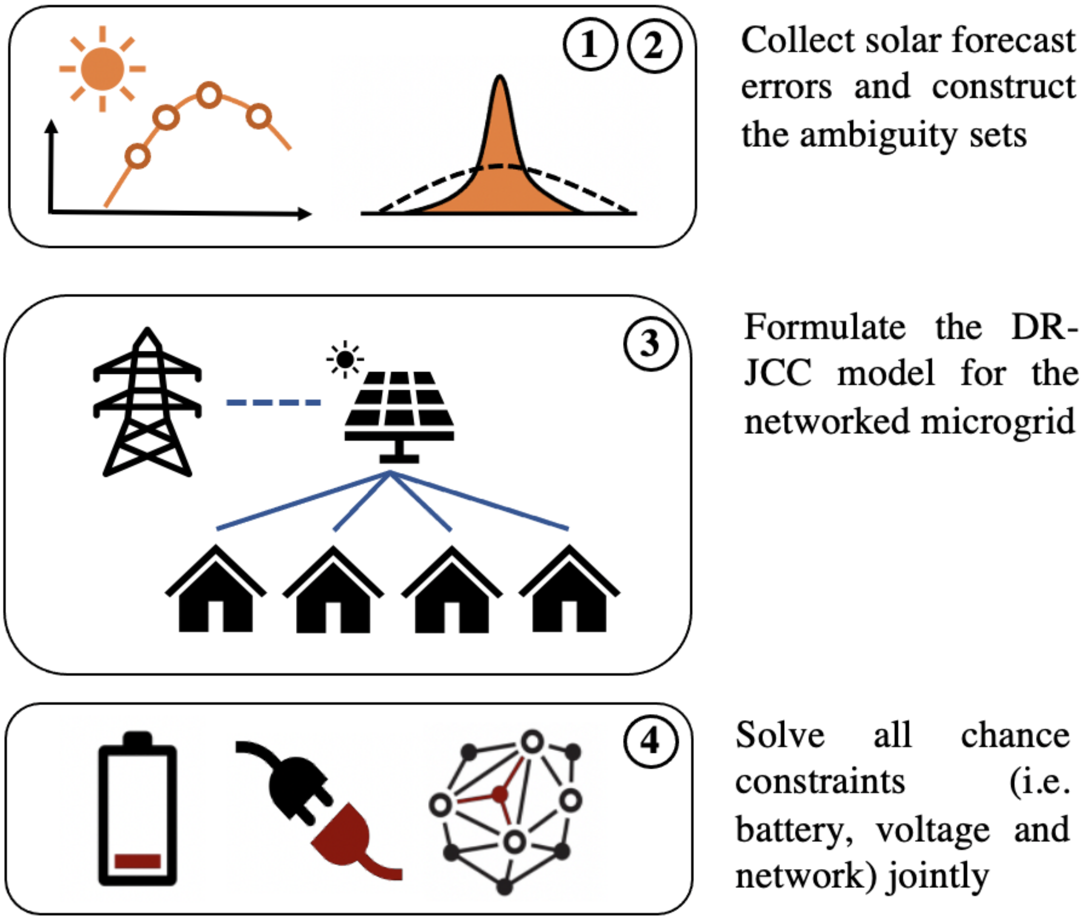

Figure 4.2: Schematic diagram for formulating and solving the JCC model for the networked microgrid

4.4.1 DR-JCC formulation

This subsection presents the DR-JCC formulation considering uncertain solar power forecast error distributions. All six constraints (4.1c) - (4.1f) and (4.1p) involving forecast errors $P_{s,t}^{\tilde{pv}}$ are reformulated in the DR-JCC fashion, while the other constraints (4.1b), (4.1g) - (4.1o), (4.1q) and (4.1r) remain the same. The new problem formulation is given by,

$$\min_{x} \quad \mathbb{E}_{\mathbb{P}}[\mathscr{J}(x)] \tag{4.2a}$$

$$\text{subject to} \quad (4.1b), (4.1g) - (4.1o), (4.1q), (4.1r) \tag{4.2b}$$

$$\inf_{\mathbb{P}\in\mathscr{P}(\mu,\sigma)} \mathbb{P}(\textstyle\bigcap_{i=0}^{N_c}(A_i(x)\zeta \leq b_i(x))) \geq 1-\varepsilon_j \tag{4.2c}$$

The letter x represents the decision variable vector (i.e., droop coefficient, grid power, and distributed asset power outputs), and ζ represents the uncertainty variable vector following the distribution $\mathbb{P} \subset \mathscr{P}$ (i.e., solar power forecast error $\tilde{P^{pv}}$). N_c is the number of individual constraints included, while $A_i(x)$ and $b_i(x)$ are affine functions about the decision variables. This DR-JCC formulation (4.2c) means that given all distributions in the ambiguity set $\mathscr{P}$ built upon moments μ, σ, the violation rate of the JCC is less than ε_j even for the worst-case distribution. Specifically, the DRJCC (4.2c) incorporating constraints (4.1c) - (4.1f) and (4.1p) is formulated as,

$$\inf_{\mathbb{P}\in\mathscr{P}(\mu,\sigma)} \mathbb{P}\begin{pmatrix} d^s_{s,t}\sum_{i=0}^{N_{PV}} \tilde{P^{pv}_{s,t}} \leq R^{up}_{s,t} \\ P^d_{s,t} + R^{up}_{s,t} \leq \overline{P^d} \\ R^{up}_{s,t} \leq P^c_{s,t} \\ E^s_{s,t} - R^{up}_{s,t}\Delta t \geq \underline{SoC} \\ \mathbf{v} \geq \underline{v} \\ \mathbf{v} \leq \overline{v} \end{pmatrix} \geq 1-\varepsilon_j \tag{4.3}$$

This work uses the approximation-based method to solve the JCC problem. In other words, JCC is decomposed into a set of SCCs with the approximated individual violation rates. Assuming that the decomposed SCC has a violation rate ε_i,

$$\inf_{\mathbb{P}\in\mathscr{P}(\mu,\sigma)} \mathbb{P}(A_i(x)\zeta > b_i(x)) \geq 1-\varepsilon_i, \quad i \in \{1,2,...,N_c\} \tag{4.4}$$

Hence, the following equalities and inequalities hold,

$$\inf_{\mathbb{P}\in\mathscr{P}(\mu,\sigma)} \mathbb{P}(\textstyle\bigcap_{i=0}^{N_c}(A_i(x)\zeta \leq b_i(x))) \tag{4.5a}$$

$$= \sup_{\mathbb{P}\in\mathscr{P}(\mu,\sigma)} \mathbb{P}(\bigcup_{i=0}^{N_c}(A_i(x)\zeta > b_i(x))) \tag{4.5b}$$

$$\leq \sup_{\mathbb{P}\in\mathscr{P}(\mu,\sigma)} \sum_{i=0}^{N_c} \mathbb{P}(A_i(x)\zeta > b_i(x)) \tag{4.5c}$$

$$\leq \sum_{i=0}^{N_c} \varepsilon_i \leq \varepsilon_j \tag{4.5d}$$

Constraint (4.5a) is transformed to (4.5b) based on the set operation properties of supremum and infimum. Then the first inequality (4.5c) is due to *Boole's inequality*. Finally, a conservative assumption (4.5d) is made to secure the solution robustness.

4.4.2 Ambiguity set construction and SOCP formulation

Solar forecast error distributions often show strong unimodality or symmetry [140]. This research uses moments and distribution structure assumptions to construct ambiguity sets. For a time interval $t \in \mathscr{T}$, there are N_s error samples $\zeta_{n,t}$. The empirical mean μ_t and covariance σ_t of the error distribution is computed in this time interval,

$$\mu_t = \frac{1}{N_s}\sum_{n=0}^{N_s} \zeta_{n,t} \tag{4.6}$$

$$\sigma_t^2 = \frac{1}{N_s - 1}\sum_{n=0}^{N_s} (\zeta_{n,t} - \mu_t)(\zeta_{n,t} - \mu_t)^\intercal \tag{4.7}$$

Given the solar power correlation between two buses a, b at time t denoted as $\gamma_{a,b,t}$, the mean vector μ_{pv} and covariance matrix Σ_{pv} of forecast errors in the network are,

$$\mu_{pv} := [\mu_{0,t}, \cdots, \mu_{b,t}] \tag{4.8}$$

$$s_{a,b} = \begin{cases} \sigma_{a,t}^2 & \text{if } a = b \\ \gamma_{a,b,t}\sigma_{a,t}\sigma_{b,t} & \text{if } a \neq b \end{cases} \quad (\forall s_{a,b} \in \mathbf{S} \in \mathbb{R}^{N_B \times N_B}) \tag{4.9}$$

$$\Sigma_{pv} := \mathbf{S} - \mu_{pv}\mu_{pv}^\intercal \tag{4.10}$$

Depending on the certainty of the computed moments, one can build the ambiguity set and recast the CC model into either the SOCP or SDP formulation. The former considers no estimation errors for the computed moments, while the latter considers the confidence level of the moment estimation. A detailed demonstration can be consulted in [179]. This research prepares a sufficient number of samples and adopts the SOCP formulation assuming the computed moments are exactly true moments of unknown distributions. Three ambiguity sets are constructed as $\mathscr{D}_\zeta^1$, $\mathscr{D}_\zeta^2$ and $\mathscr{D}_\zeta^3$. This chapter drops indexes t, pv for conciseness.

Ambiguity set 1: (Unimodal, centered at the mean and mode zero.)

$$\mathscr{D}_\zeta^1 := \left\{ \mathbb{P} \in \mathscr{P}^\alpha : \begin{array}{c} \mathbb{E}[\zeta] = \mu, \mathbb{E}[(\zeta - \mu)^2] = \sigma^2 \\ \mathscr{M}[\zeta] = \mu = 0 \end{array} \right\}$$

where $\mathscr{P}^{\alpha}$ denotes all unimodal distributions on $\mathbb{R}^n$, and $\mathscr{M}[\zeta]$ is the mode of distributions.

Ambiguity set 2: (Symmetric, centered at mean zero.)

$$\mathscr{D}_{\zeta}^{2} := \left\{ \mathbb{P} \in \mathscr{P} : \begin{array}{c} \mathbb{E}[\zeta] = \mu, \mathbb{E}[(\zeta-\mu)^2] = \sigma^2 \\ \mathbb{P}[\zeta] = \mathbb{P}[-\zeta] = 0 \end{array} \right\} \tag{4.11}$$

Ambiguity set 3: (Unimodal and symmetric, centered at mean and mode zero.)

$$\mathscr{D}_{\zeta}^{3} := \left\{ \mathbb{P} \in \mathscr{D}_{\zeta}^{1} \cap \mathscr{D}_{\zeta}^{2} \right\} \tag{4.12}$$

Based on the proof in [180], a single DR-CC as (4.4) can be recast into a SOCP constraint given by,

$$\lambda(\varepsilon_i)||\Sigma^{1/2}A_i(x)||_2 \leq b_i(x) - \mu^{\intercal}A_i(x) \tag{4.13}$$

where $\lambda(\varepsilon_i)$ depends on the specific ambiguity set. For $\mathscr{D}_{\zeta}^{1}$, $\mathscr{D}_{\zeta}^{2}$ and $\mathscr{D}_{\zeta}^{3}$, their functions $\lambda_1(\varepsilon_i)$, $\lambda_2(\varepsilon_i)$ and $\lambda_3(\varepsilon_i)$ are,

$$\lambda_1(\varepsilon_i) := \frac{2}{3}\sqrt{\frac{1}{\varepsilon_i}} \quad \forall \varepsilon_i \in (0, \frac{1}{3}) \tag{4.14}$$

$$\lambda_2(\varepsilon_i) := \sqrt{\frac{1}{2\varepsilon_i}} \quad \forall \varepsilon_i \in (0, \frac{1}{2}) \tag{4.15}$$

$$\lambda_2(\varepsilon_i) := \sqrt{\frac{2}{9\varepsilon_i}} \quad \forall \varepsilon_i \in (0, \frac{1}{6}) \tag{4.16}$$

Equations (4.14) - (4.15) are proven in [181] and [182] based on *Gauss's inequality* and *Chebyshev's inequality*. Equation (4.16) is proven in Appendix A.1.1.

4.4.3 Approximation of individual violation rates

Since the model only knows a joint violation rate ε_j, single violation rates ε_i need to be computed. The *Bonferroni Approximation* is set as the baseline case. It assumes all individual violation rates to be equal, and the sum of individual violation rates is the exact joint violation rate.

$$\varepsilon_i := \frac{\varepsilon_j}{N_c} \tag{4.17}$$

Nevertheless, *Bonferroni Approximation* leads to a very conservative solution, as this approximation assumes all constraints have the same chance of being violated. This research, therefore, proposes an improved method using *Optimized Bonferroni Approximation*. It considers the individual violation rate ε_i as a variable rather than a fixed priori, which is solved simultaneously with the original optimization problem [134].

$$\mathbf{H} := -\mathbf{C}^s \mathbf{d}_t^s + \mathbf{C}^{pv} \tag{4.18}$$

$$\mathbf{G} := \mathbf{diag}(\mathbf{V}_o)\mathbf{Y} + \mathbf{diag}(\mathbf{I}_o) \tag{4.19}$$

$$\lambda(\varepsilon_i)||\mathbf{d}_t^s \mathbf{1} \Sigma_{pv,t}^{1/2}||_2 \leq \mathbf{R}_t^{up} - \mu_{pv,t}^{\mathsf{T}} \mathbf{1} \mathbf{d}_t^s \tag{4.20}$$

$$\lambda(\varepsilon_i)||\mathbf{d}_t^s \mathbf{1} \Sigma_{pv,t}^{1/2}||_2 \leq \overline{\mathbf{P}^d} - \mathbf{P}_t^d - \mu_{pv,t}^{\mathsf{T}} \mathbf{1} \mathbf{d}_t^s \tag{4.21}$$

$$\lambda(\varepsilon_i)||\mathbf{d}_t^s \mathbf{1} \Sigma_{pv,t}^{1/2}||_2 \leq \overline{\mathbf{P}^c} - \mu_{pv,t}^{\mathsf{T}} \mathbf{1} \mathbf{d}_t^s \tag{4.22}$$

$$\lambda(\varepsilon_i)||-\mathbf{d}_t^s \mathbf{1} \Delta t \Sigma_{pv,t}^{1/2}||_2 \leq \mathbf{E}_t^s - \underline{\mathbf{SoC}} - (-\mu_{pv,t}^{\mathsf{T}} \mathbf{1} \Delta t \mathbf{d}_t^s) \tag{4.23}$$

$$\lambda(\varepsilon_i)||\mathbf{G}^{-1}\mathbf{H}\Sigma_{pv,t}^{1/2}||_2 \leq \overline{\mathbf{v}} - \mathbf{G}^{-1}\{\mathbf{diag}(\mathbf{I}_o)\mathbf{V}_o + \mathbf{P}_t^{inj}\} \tag{4.24}$$

$$\lambda(\varepsilon_i)||\mathbf{G}^{-1}\mathbf{H}\Sigma_{pv,t}^{1/2}||_2 \leq -\underline{\mathbf{v}} + \mathbf{G}^{-1}\{\mathbf{diag}(\mathbf{I}_o)\mathbf{V}_o + \mathbf{P}_t^{inj}\} \tag{4.25}$$

Based on *Optimized Bonferroni Approximation* and the reformulation in (4.13) - (4.16), the DR-JCC inequality (4.3) is recast into individual SOCP constraints (4.20) - (4.25), in addition to a new variable ε_i and constraint (4.5d). $\mathbf{1}$ represents the unit vector with the dimension of buses, and $\mu_{pv,t}^{\mathsf{T}}\mathbf{1}$ thus is the sum of total forecast error across all buses. However, introducing the variable ε_i destroys the convexity of this problem. One can observe that each constraint has a multiplication of variables. The problem has been proven as strongly NP-hard [134].

4.4.4 Evolutionary algorithm for the JCC problem

To tackle the aforementioned NP-hard problem, this research proposes an evolutionary algorithm to approximate solutions to this intractable JCC problem, including the optimized individual violation rate for each CC and total operation cost. In the subfield of meta-heuristic optimization, the evolutionary algorithm is a bio-inspired algorithm analogous to the natural evolution process [183]. The essence of an evolutionary approach to solving a problem is to equate possible solutions to individuals in a population and to introduce a notion of fitness based on solution quality [184].

Algorithm 1: Population-based evolutionary algorithm

```
Initialise the population with random individual solutions;
Evaluate each individual solution;
while termination condition is not satisfied do
    Perform competitive selection;
    Apply pair, breed and mutation procedures;
    Evaluate the new pool of individual solutions;
    Apply replacement to form the new population;
    Find current best solution;
end
Output overall best solution;
```

Algorithm 1 provides the pseudo-code of the proposed population-based evolutionary algorithm. First, a group of N_p individuals $\mathbf{P}$ as the first-generation population is created,

$$\mathbf{P} := [\varepsilon_0, ..., \varepsilon_i] \quad \forall i \in [0, N_c], \forall \varepsilon_i \in [\underline{\varepsilon}, \overline{\varepsilon}] \tag{4.26}$$

Each individual $\mathbf{P}$ has six parameters, which are the optimized variables for each CC violation rate (4.26). The sum of individuals' parameters should satisfy constraints (4.5d). This research uses the voltage violation rate of a feeder in reality (e.g., 0.1% [185]) to set the lower bound of parameters.

$$\mathscr{F}(\mathbf{P}) := \mathbb{E}_{\mathbb{P}}[\mathscr{J}(\mathbf{P}, x)] \tag{4.27}$$

Each individual is evaluated by the fitness value, defined as the objective function value of the JCC problem (4.27), where x is the decision variable in the original optimization problem to be solved together. The evolution of the population is conducted iteratively based on competitive selection. For each iteration, only the first half of individuals with lower fitness values are selected as the *elite* for the next generation.

$$\varepsilon_i^s := \frac{\varepsilon_i^{cm} + \varepsilon_i^{cn}}{2} \quad \forall \varepsilon_i^{cm} \in \mathbf{P}_m, \forall \varepsilon_i^{cn} \in \mathbf{P}_n \tag{4.28}$$

$$\varepsilon_i^{s'} := \varepsilon_i^s + \max\{\theta, 0\} \quad \theta \sim \mathcal{N}(0, \sigma^2) \tag{4.29}$$

The *elite* pairs with each other and generates offspring. Specifically, the offspring is generated by taking the average value of the parents' parameters (4.28). The mutation of offspring is necessary. Otherwise the solution might be trapped into a local minimum. That is to add a random number from a normal distribution to six offspring' parameters (4.29). The parameters of the mutated offspring should be normalized so that the sum is always equal to the joint violation rate. The mutation only happens when all parents are different. Otherwise, the fitness values over generations are hard to converge.

The evolutionary iteration will stop until these two termination criteria are reached, the maximum number of iterations (i.e., N_{it} = 10), and the ratio to measure the solution convergence. It is defined as the ratio of the maximum and the average value of fitness functions in each generation as below.

$$r := \frac{\max\{\mathscr{F}(\mathbf{P}_0), ..., \mathscr{F}(\mathbf{P}_n)\}}{\frac{1}{N_p}\sum_{n=0}^{N_p} \mathscr{F}(\mathbf{P}_n)} - 1 \leq r_{thr} \tag{4.30}$$

The algorithm will stop when the ratio (4.30) is lower than the threshold value, r_{thr}. Empirical evidence such as [184], [186], and [187] shows that if the evolutionary algorithm, as one of the global optimization methods, is repeated many times with the random initial guesses in the first generation and still obtains the same solution, this solution is considered as an acceptable approximation of the global optimum.

This method can be readily transferable from one problem to another, as only two parts of the algorithm are problem-dependent, the initial values of the first generation (i.e., the initial guess of individual violation rates) and fitness function (i.e., the objective function). Moreover, one can adjust the convergence ratio (4.30) of the algorithm to decide the trade-off between the fidelity of the optimum approximation and the computation time. However, if a problem has a very small joint violation rate divided by a great number of CCs, some of the initial guesses for the optimized violation rates could result in infeasible solutions. A successful evolution process would require a large population and significant computation efforts.

4.5 Data-driven solar power forecast

A Light GBM in ref. [57] is used to predict solar power. This research does not consider the demand uncertainty since its forecast errors are generally much smaller than solar power forecast errors for a microgrid with a high solar self-consumption. The dataset used is two-year 15-min weather measurements in Gitaru dam, Kenya, including the solar irradiance, air temperature, and wind speed for the prediction features. The prediction horizon is set to be one day, and the prediction is updated every time interval in the receding horizon for a whole year.

This model calculates solar power forecast errors over the year, aggregates them based on time intervals, and labels the mean and mode, as in Figure 4.3 (a). For each time interval with daylight, there are 365 error samples from each day in a year. The maximum value is around 150% in the early morning when the solar power is too small (around 1W) to be accurately predicted. One can also observe several phenomena in Figure 4.3. First, those error distributions in each time interval are highly symmetric and centered at their means. Second, the mean and mode are around zero and relatively close except for early morning and late afternoon when the solar irradiance is very low. Third, the distribution is unimodal but not necessarily normal, as shown in Figure 4.3 (b).

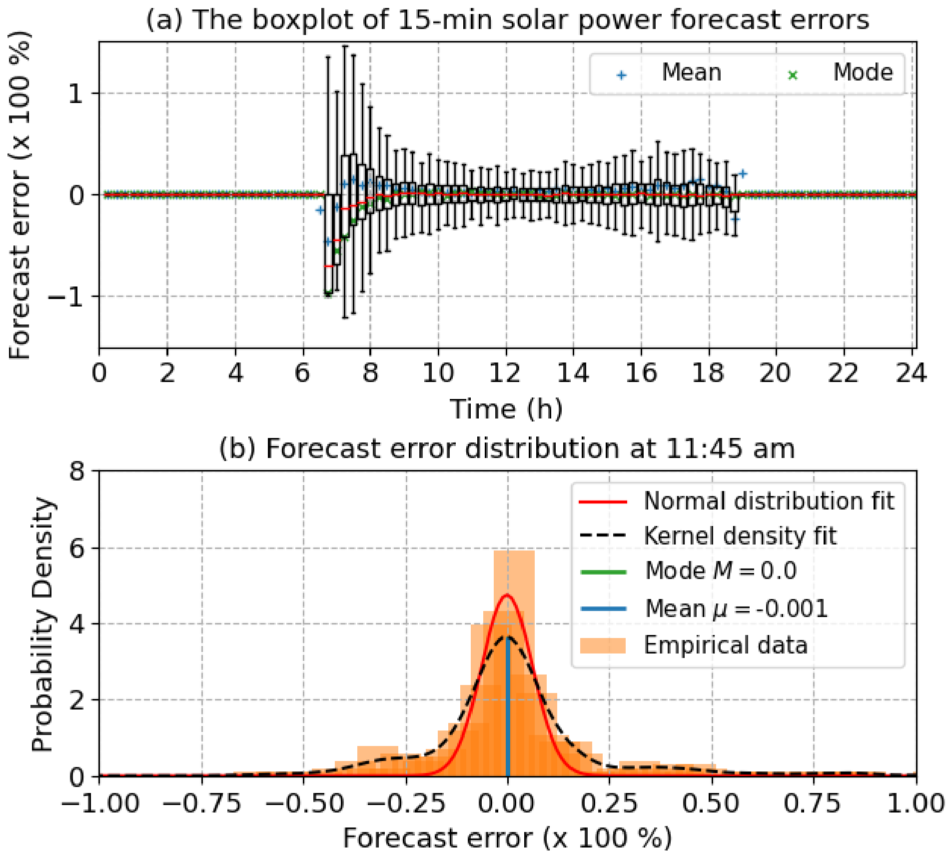

Figure 4.3: Solar power forecast error distributions in a year (a) the boxplot of 15-min solar power forecast error (modes and means labeled with black and green markers) (b) Forecast error distribution at 11:45 am (with non-parametric kernel density estimation in black dash line and parametric Gaussian fitting in solid red line)

4.6 Case Study

The section presents a case study of a networked DC microgrid in rural Kenya. This microgrid has ten households connected to the main busbar radially, and each has local PV panels and 60Wh, 20W home batteries. Their power outputs are considered to be independent. Households' load profiles are constructed using weather data, detailed in Appendix B.2. Specifically, inflexible (i.e., light), and flexible (i.e., fan and phone charger) loads are considered. The utility for using them is modeled as a quadratic function of power, temperature, and solar irradiance based on welfare economics [165]. The parameter m_l is a coefficient of the utility function to reflect users' comfortness to achieve smart load shedding. For example, the users' utility when using lights is inversely proportional to the solar irradiance, meaning that more utility users can get from lights when solar irradiance is lower.

Table 4.1: Case study parameters

m_u	\$0.023/(kWh)^2	η_{dis}, η_{ch}	0.95	N_p	6
m_s	\$1.00/(kWh)^2	$\underline{SoC}, \overline{SoC}$	0.2, 1	N_{it}	10
m_v	\$0.23/(kWh)^2	$\underline{v_b}, \overline{v_b}$	0.95, 1.05p.u.	r_{thr}	2%
m_d	\$0.27/kWh	R_{line}	8Ω/km	σ_m	0.1

The power capacity of PV panels and line length from the main bus to households are drawn from uniform distributions $\mathscr{U}_1(20,40)$ [W] and $\mathscr{U}_2(50,200)$ [m]. Table 4.1 shows the model parameters in the first two columns and the evolutionary algorithm parameters in the third column. The model parameters, such as line resistance, are from the manufacturer information [188], and m_s is set to be large to encourage self-consumption. The evolutionary algorithm parameters are optimized from empirical experiments. For example, we increase the number of individuals N_p by two each time and observe the diminishing marginal improvements in computation time and results until there are no significant improvements. The model framework is built using CVXPY package [189] in Python and runs on an Apple iMac with a processor of 3.1GHz Intel Core i5 and a memory module of 8 GB 2133 MHz LPDDR3.

4.6.1 Solving DRJCC problem with evolutionary algorithm

This research first solves the DR-JCC model with different joint violation rates using the proposed evolutionary algorithm and compares results with the baseline case, *Bonferroni Approximation*. In each experiment with one joint violation rate, the evolutionary algorithm is run for 10 times and the solutions converge to the same optimum. Those solutions are thus considered an acceptable approximation of the global optimum.

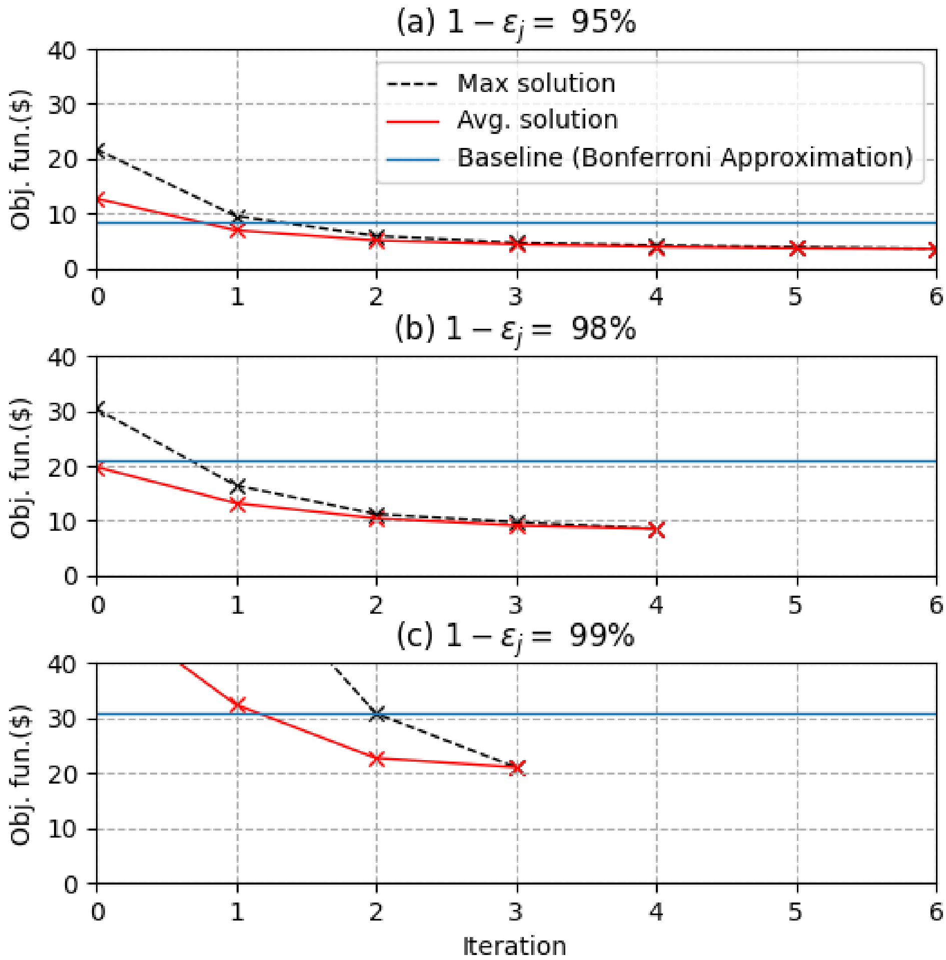

Figure 4.4: Computation processes for DR-JCC under three joint violation rates

The computation processes for three joint violation rates (i.e., 0.05, 0.02, 0.01) with the unimodal ambiguity set $\mathscr{D}_{\zeta}^{1}$ are shown in Figure 4.4. First, solutions converge faster in the higher system reliability (i.e. $1-\varepsilon_j$) cases. This is because the numerical range of individuals' parameters to explore becomes smaller. Second, the substantial reduction of the objective function is often achieved in the first two or three iterations. This means one can decide the trade-off between computation time and solution conservativeness by changing the termination condition such as the ratio r_{thr}.

Detailed results and computation times are listed in Table 4.2. Significant cost reductions are achieved in all runs compared with the baseline method. For three cases, the reduction is 58.50%, 60.35% and 31.75% respectively. Parameters ε_1 - ε_6 are the

Table 4.2: Summary of results and computation times (BL.: Baseline; PPSD.: Proposed

Joint vio. rates	**0.05**		**0.02**		**0.01**	
Methods	BL.	PPSD.	BL.	PPSD.	BL.	PPSD.
ε_1		0.023		0.008		0.003
ε_2	0.0083	0.022	0.0033	0.008	0.0016	0.003
$\varepsilon_3 - \varepsilon_6$		0.001		0.001		0.001
Time (s)	11.02	464.02	11.69	351.82	12.55	269.14
Obj. func. ($)	8.54	3.54	21.54	8.54	30.86	21.06

optimized individual violation rates of CCs (4.20) - (4.25). These constraints are voltage droop regulation, battery power discharging and charging limits, battery energy limit, and voltage upper and lower regulations. Among all six violation rates, parameters ε_1 and ε_2 are optimized to have the higher values, while parameters ε_3 - ε_6 are optimized to have the lower values (i.e., 0.001). To check how these parameters change with the power flow conditions, the line resistance R_{line} is changed from 8 to 12 Ω/km and repeat experiments. The operation cost slightly increases, but the value of parameters remains unchanged.

The results indicate the first two constraints are the most critical to the operation costs. They are droop regulation and battery discharging power limit. This means the operation cost is mainly determined by the droop provision and solar power uncertainty because of the high value for cost coefficients for energy reserve and solar curtailment. On the other hand, tightening the last four constraints, battery power charging limit, bus voltage, and battery energy limits, will not increase the cost significantly, even with a high line resistance. This is because the objective function does not include the monetary term for power loss about bus voltages.

4.6.2 DRJCC framework performances

As this model framework aims to address renewable uncertainty in microgrids and tackle utility contingencies, this research conducts simulations with a progressively smaller joint violation rate ranging from 0.2 to 0.01. The utility blackout is simulated, ranging from one hour to a day.

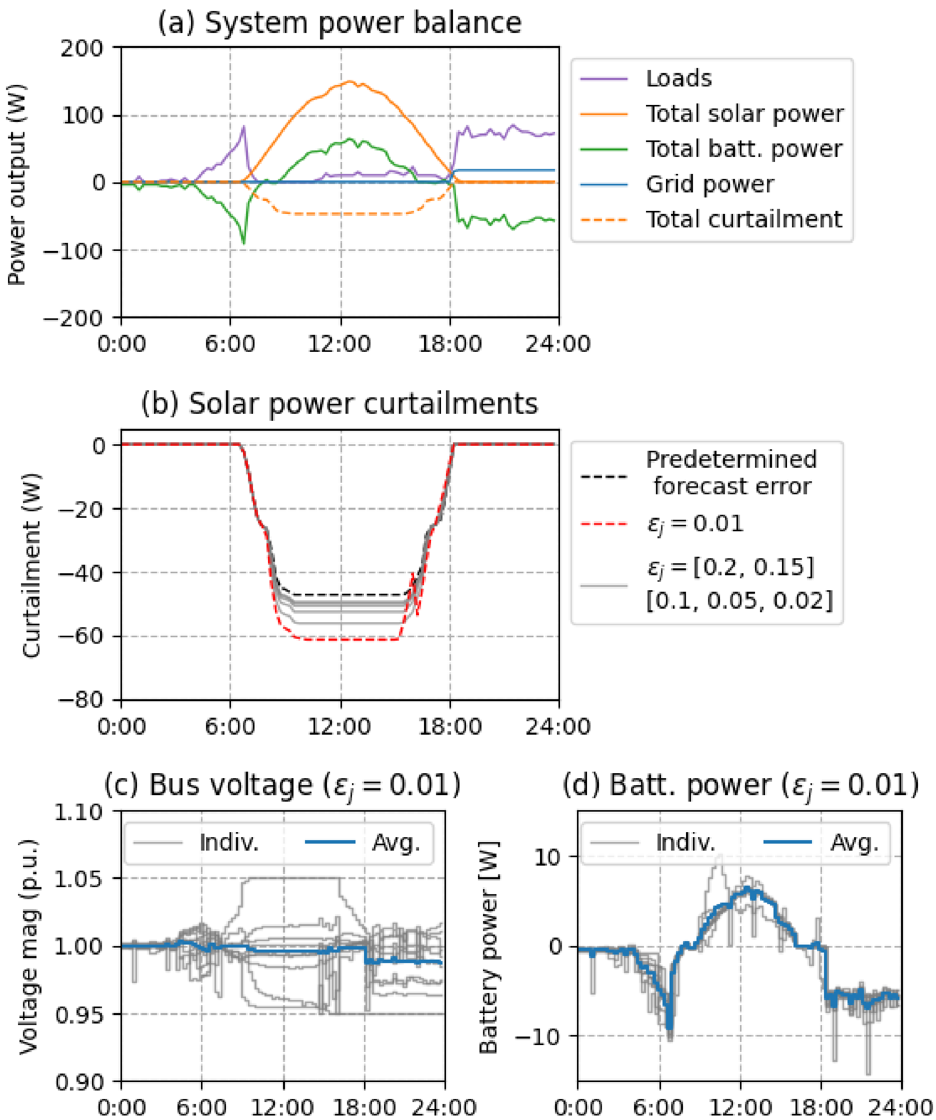

Figure 4.5: (a) System power balance, (b) solar power curtailment, (c) bus voltages when $\varepsilon_j = 0.01$ and (d) batteries' power outputs when $\varepsilon_j = 0.01$ in the DR-JCC framework. The average values are plotted in thick blue lines in (c) and (d) (with 50% opacity to show individual values)

Figure 4.5 showcases how the microgrid system addresses solar generation uncertainty under different joint violation rates ε_j, including (a) overall power flow, (b) solar power curtailment, (c) bus voltages and (d) battery powers when $\varepsilon_j = 0.01$. For battery power outputs, negative values represent the battery discharging. In the DR-JCC framework, the microgrid system curtails the increasing amount of solar power with a progressively tighter joint violation rate in Figure 4.5 (b). Figure 4.5 (c) and (d) show bus voltages and battery power of individual households with the average value. Bus voltages are regulated within a range of ± 0.05 p.u..

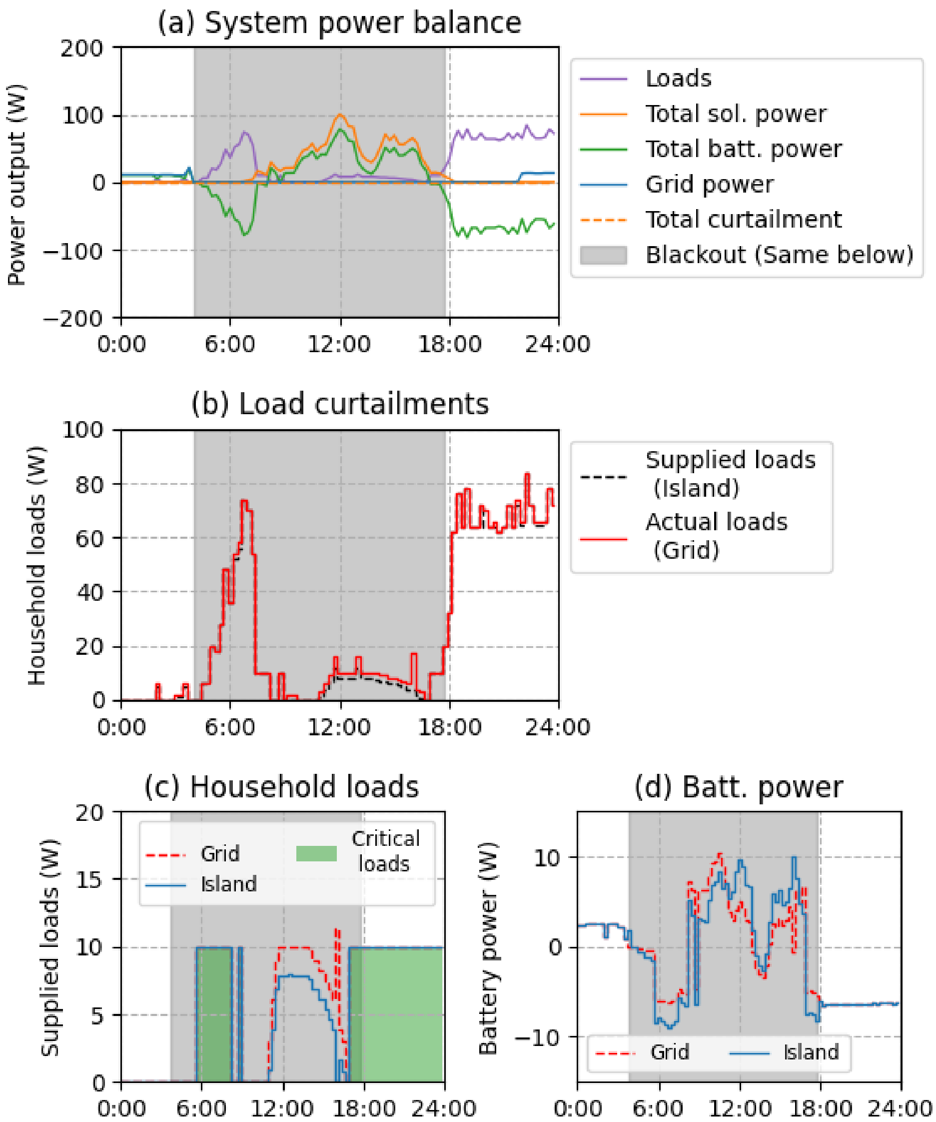

Figure 4.6: (a) System power balance, (b) load curtailments, (c) one household load supply, and (d) one household battery power output when there is a blackout. In (b), (c), and (d), two scenarios are presented, the grid-connected one under the normal condition and the island one for the blackout.

The islanding operation of the microgrid is simulated for an overcast day with a blackout between 4 a.m. - 6 p.m. Figure 4.6 shows the simulation result. In Figure 4.6

(b), when a blackout happens, flexible loads are curtailed, mainly phone changer (i.e., load spike at 5 p.m.) and fan at noon. Battery control reference follows the island mode solution during a blackout, as solid blue lines in Figure 4.6 (d).

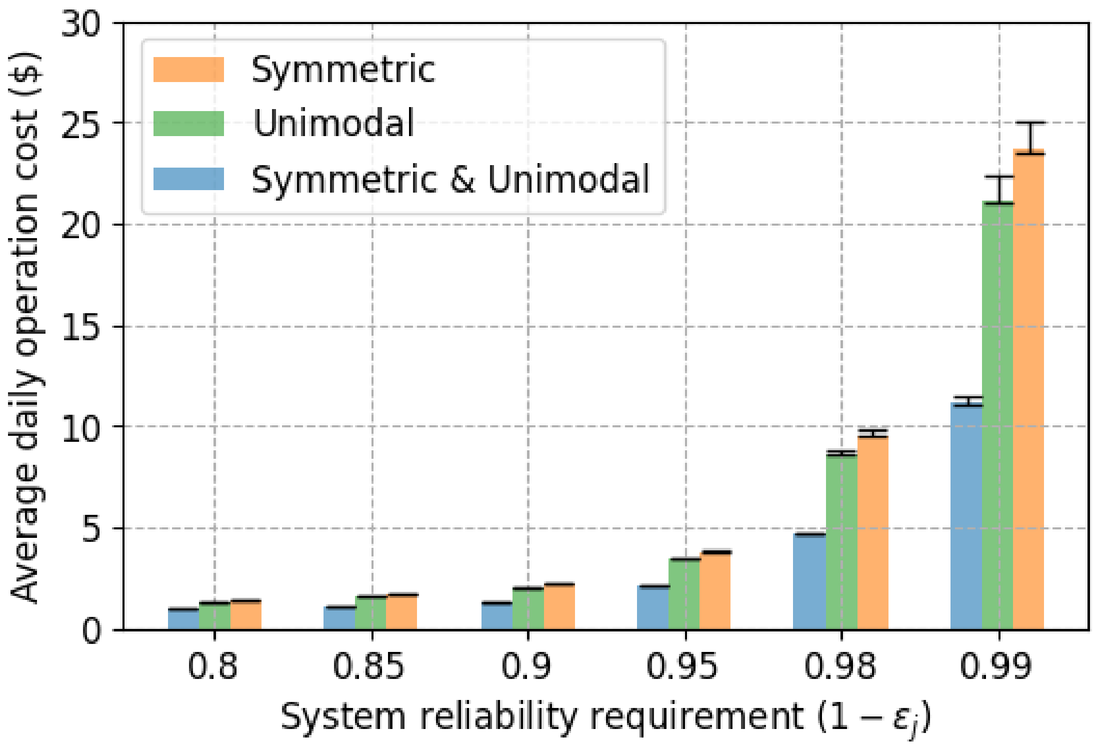

Figure 4.7: Daily operation costs with the three ambiguity sets (with the error bar to show the cost deviation due to the island hour from no blackout to a full-day blackout)

Daily operation costs are computed under six different reliability requirements (i.e., $1-\varepsilon_j$) and three ambiguity sets with a 12-hour island period. The error bars show the cost deviations due to the island time from no blackout to a full-day blackout. In summary, in Figure 4.7, daily operation costs increase exponentially along with system reliability requirements for all three ambiguity sets, mainly due to solar power curtailment penalties and the increasing reserved energy. Extending the island duration also increases the operational cost due to battery degradation costs.

4.6.3 Solution robustness and system reliability

Following demonstrations of framework performance, this section presents solution robustness tests for the proposed evolutionary algorithm and system reliability outcomes. This research uses forecast error samples, excluding those used to construct the ambiguity set. The simulations fix the solution obtained, run the power flow under those forecast error samples, and then count the case when all the constraints are met (i.e., no-violation case). The percentage of no-violation time intervals out of total time intervals in a day is defined as the daily reliability outcome.

This research tests three joint violation rates, which are 0.05, 0.02, and 0.01 (i.e., the corresponding daily reliability requirements are 0.95, 0.98, and 0.99), and considers three aforementioned ambiguity sets, symmetric, unimodal and symmetric & unimodal sets. For comparison, this research includes another two methods, SCC with all violation rates equal to the set reliability level, and the baseline case, JCC with *Bonferroni Approximation*. A full factorial experiment of these settings gives a total number of $3 \times 3 \times 3 = 27$ robustness tests. Each test uses 10,800 forecast error samples (i.e., 30 samples for each 15-min time interval) from the historical sample pool. Samples in each time interval are from different days and independent. Thus the power flow tests conducted in each time interval are considered to be independent.

As the perfect ambiguity set is unattainable (i.e., the exact shape of uncertainty distributions is unknown or cannot be modeled exactly), the ambiguity set will be either larger or smaller than exact distributions. We use three ambiguity sets in this work. The unimodal and symmetrical sets are both larger than the exact distribution, and the unimodal & symmetrical set is smaller than the exact distribution. The proposed JCC is effective because the tested reliability is higher than the required value when using the overly-conservative / small ambiguity set (i.e., the unimodal & symmetrical set). The reliability is lower than the required value when using the large ambiguity

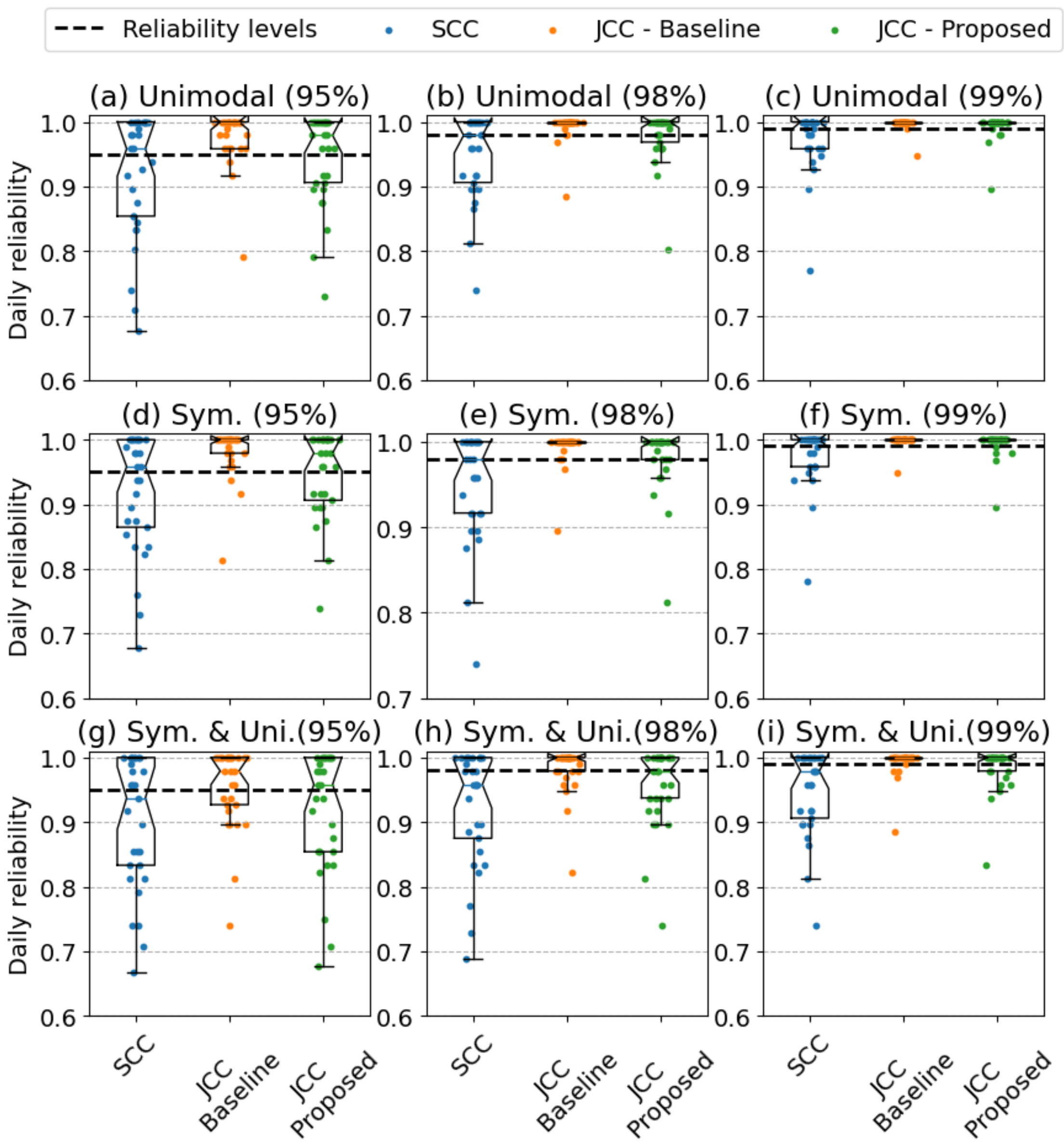

Figure 4.8: Summary of the full factorial robustness tests for three formulation methods. (Titles of each plot show the experiment setting; Black dash lines show the reliability levels; Colored markers in the box plots are daily results in each experiment)

set. (i.e., the unimodal and symmetrical sets). In this case, it could be deduced that if there exists a perfect ambiguity set, the proposed JCC will give the reliability which is closest to the required value.

Figure 4.8 shows results from all robustness tests. Each subplot represents a combination of a daily reliability requirement and an ambiguity set. Three box plots in each experiment from left to right show results from SCC (blue circles), JCC using *Bonferroni*

Approximation (orange triangles) and the proposed algorithm for JCC (green squares), respectively. Colored markers in each box plot are daily results. Table 4.3 summarizes the average daily reliability in all robustness tests, and the reliability values which are closets to the set requirements are highlighted in **bold**.

Table 4.3: Summary of the average daily reliability (%) in all robustness tests

Reliability	**95%**			**98%**			**99%**		
Cases	(a)	(d)	(g)	(b)	(e)	(h)	(c)	(f)	(i)
SCC	91.5	91.9	89.9	94.6	94.9	92.3	97.4	97.5	94.6
JCC-B	97.8	98.1	**95.3**	99.3	99.4	**98.2**	99.8	99.8	**99.3**
JCC-P	**94.3**	**94.6**	91.7	**97.9**	**98.1**	95.4	**99.4**	**99.4**	98.4

SCC: Single chance constraints; JCC-B: Joint chance constraints (Baseline); JCC-P: Joint chance constraints (Proposed)

The SCC method fails to meet reliability requirements in all tests. More than half of its reliability outcomes are below the set requirements (i.e., black dash lines in Figure 4.8), and the average daily reliability is also far lower than requirements (Table 4.3). In contrast, JCC with *Bonferroni Approximation* gives an overly-conservative solution. In cases except (g) and (h), the majority of daily reliability outcomes are exactly or nearly 100%. This makes its average reliability outcomes are highest and above the set requirements in all cases. The proposed algorithm can decide the optimal trade-off between the system reliability and operation cost. In each test, more than half of its daily reliability meets the requirements (i.e., the median of the box plot is all above the reliability threshold), and its average reliability values in six of nine cases are closest to the set reliability levels. In cases (g), (h), and (i), the underlying uncertainty distribution assumption is based on the unimodal & symmetrical set, which is the smallest set among all three. The set can not fully incorporate all true distributions and impacts the reliability performance. The average values are below the reliability requirements.

For three ambiguity sets, the rank based on the system reliability from the highest to the lowest is symmetric, unimodal and unimodal & symmetric sets, respectively. This result is aligned with the operation cost (Figure 4.7). When the set gets less constrained, the solution becomes less reliable, and the operation cost becomes lower, and vice versa. In our case, the unimodal & symmetric set is overly-constrained to describe the true distribution of solar forecast errors. However, even using the other two sets, there are a few discrete outliers on particular days that are costly to address unless using an ambiguity set including all possible distributions. A feasible method for system operators could be to predict those particular days using the multi-year data and prepare the extra storage for those times.

4.7 Concluding remarks and discussion

The chapter proposes a DR-JCC framework for microgrids considering solar generation uncertainty and utility contingencies. The framework models a networked microgrid with an islanding capability and smart load shedding during blackouts. Under imperfect solar forecasts, it optimizes CCs pertaining to the power flow, voltage control, and battery limits *jointly* to decide the optimal trade-off between the operation cost and system reliability.

To find the optimized individual violation rates for the JCC problem, this research proposes a novel population-based evolutionary algorithm to optimize the decision variables and individual violation rates simultaneously. Results clearly show the proposed algorithm can effectively solve the problem non-conservatively. It can reduce the operation cost by around 50% compared to the benchmark case, which has evenly individual violation rates (i.e., *Bonferroni Approximation*). Moreover, the individual violation rates from the proposed method indicate the cost sensitivities of constraints. A higher violation rate means this constraint is more influential to the total cost.

For solution robustness, this research considers three ambiguity sets for solar power forecast error distributions, unimodal, symmetric and unimodal & symmetric sets, based on empirical samples. The model is solved based on these set assumptions and then test the constraint violations with new forecast errors. Under the well-fitting ambiguity set assumption, the solution from the proposed method can control the system to meet the reliability requirements closely. The SCC formulation widely used in current research and practices, however, shows poor performance in securing reliability.

Although this research proposes a lightweight heuristic optimization method to approximate its solutions, a conservative assumption has been made beforehand. That is, individual violation rates can be directly split from the joint violation rate (i.e., $\sum_{i=0}^{N_c} \varepsilon_i = \varepsilon_j$), which satisfies the upper bound in Boole's inequality. This assumption neglects the intersections of probabilities of individual violation events. Theoretically, the more individual CCs are, the larger their intersection and the more conservative solution is. Improvements can be made by considering a less conservative case when the sum of individual violation rates is great than the joint violation rate (i.e., $\sum_{i=0}^{N_c} \varepsilon_i \geq \varepsilon_j$)). However, to secure the solution robustness in this case, the reliability test will need to be incorporated into the solution approximation procedures. This requires a more comprehensive solving framework.

Following discussions, future research will investigate the implementation of other solving methods and compare them with the proposed genetic algorithm in reducing the solution conservativeness and solution time. Another important area is to implement this approach in the transmission network operation and planning (e.g., [169], [190]), which can shift the paradigm in risk and reliability management, especially under extreme weather and create greater value by saving million-scale reserve procurement costs.

5

Joint Chance-constrained Game for Coordinating Microgrids in Energy and Reserve Markets

Contents

Chapter Summary: *Networked microgrids incorporate distributed energy resources and flexible loads, which can provide energy and reserve services for the main grid. However, due to uncertain renewable generations such as solar power, microgrids might breach the day-ahead contracts and under-deliver reserve services in the real-time market. If multiple microgrids fail to deliver services as promised, this could lead to a severe contingency. This chapter designs a DRJCC market game simulating risk-aware bidding and system-wide reserve policy. Leveraging historical error samples, the reserve bidding*

strategy of each microgrid is formulated into a two-stage Wasserstein-metrics DRO model. A JCC regulates the under-delivered reserve capacities of all microgrids in a non-cooperative game. Considering the unknown correlation among players, a novel Bayesian optimization method approximates the optimal individual violation rates of microgrids and market equilibrium. The proposed game framework with the optimal rates is simulated with up to 14 players in a 30-bus network. Case studies are conducted using the California power market data. The proposed Bayesian method can effectively regulate the joint violation rate of the under-delivered reserve and secure the profit of microgrids in the reserve market.

5.1 Nomenclature

A.	*Set and indices*
$t \in \mathscr{T}$	Set, index of time steps
$n \in \mathscr{N}$	Set, index of network buses
$i \in \mathscr{I}$	Set, index of microgrids
$j \in \mathscr{J}$	Set, index of net loads
$k \in \mathscr{K}$	Set, index of thermal generators
$s \in \mathscr{S}$	Set, index of forecast error samples

B.	*Parameters*
N_B	Number of buses
N_S	Number of error samples for constructing ambiguity sets
N_{LN}	Number of lines in the network
N_L	Number of passive loads
N_{MG}	Number of active microgrids
N_{OS}	Number of out-of-sample tests
N_{IT}	Number of iterations in Bayesian optimization
m^g	Fuel cost of thermal generations [\$/(MWh)^2]
m^s	Battery degradation cost of microgrids [\$/(MWh)^2]
m^{mgr}	Reserve provision cost of microgrids [\$/MWh]

m^{gr}	Reserve provision cost of thermal generators [$/MWh]
R^{req}	System reserve requirement [MW]
$\overline{P^g}$	Power capacity of thermal generators [MW]
P^{nl}	Net loads at buses [MW]
V	VoLL of microgrids [$/MWh]
$\overline{R^{fup}}$	Maximum energy capacity of the under-delivered reserve of each microgrid [MWh]
$\overline{P^s}$	Lower power limit of flexible renewable generations of each microgrid [MW]
$\underline{P^s}$	Upper power limit of flexible renewable generations of each microgrid [MW]
$\overline{L}$	Line limit of the transmission network [MW]
ν	Radius of the Wasserstein-metrics ambiguity set
γ	Fraction of flexible generations in microgrids that can be used for reserve provision

C.	*Decision variables*
π^r	Reserve bidding prices [$/MWh]
π^e	Energy bidding prices [$/MWh]
P^g	Power output of thermal generators [MW]
P^s	Power output of flexible generations of microgrids [MW]
P^{inj}	Power injection at each bus [MW]
R^g	Reserve capacity offers of thermal generators [MW]
R^{up}	Reserve capacity offers of microgrids [MW]
P^{ex}	Energy capacity offers of microgrids [MW]
ε^{jot}	Joint violation rate
ε^{ind}	Individual violation rate
ζ	Uncertain variable for solar forecast errors

5.2 Introduction

The transition to the net-zero emission energy system drives the rise of BTM assets, including solar PV panels, heat pumps, and electric vehicles [191]. These distributed energy resources can be aggregated at the distribution level as a microgrid to provide energy services such as reserve. However, these microgrids do not directly receive price signals from the wholesale day-ahead or real-time markets. Furthermore, they have no explicit energy and power capacity, and their service provisions are uncertain due to the stochastic nature of renewable generations and demands.

To coordinate their energy service provisions, previous works have proposed different market designs, including distribution locational marginal prices (DLMP) [192, 193], bilateral matching [194], decentralized negotiation [195] and game-theoretical framework [196, 197]. In the conventional LMP and bilateral matching frameworks, microgrids are considered price-takers, meaning they do not influence the market price and other players' strategies. While in the game-theoretical framework, these microgrids are interested in maximizing their own profits. Their biddings or offerings change based on other players' strategies and the market price until market equilibrium is reached. This market dynamics can be modeled mainly by either of two methods. They are the iterative or heuristic algorithm to search for the equilibrium, such as the best response algorithm, and mathematical programming with equilibrium constraints (MPEC) based on Karush-Kuhn-Tucker (KKT) rule [198].

The first method simulates each player's strategy iteratively and searches a market equilibrium until all players' solutions converge to a global optimum [199, 126]. However, this approach only guarantees convergence to an efficient equilibrium for certain types of utility functions, and it is sensitive to the starting point [200]. MPEC method secures the equilibrium when at least a feasible solution exists, but it results in the non-convex optimization problem requiring additional mathematical reformulations [201].

Modeling the uncertainty in the market game is another challenge. Deterministic formulations consider certain players' strategies and their service provisions [196], which neglects the stochasticity of market strategies. The scenario-based approach is a straightforward way to model uncertainty. For example, ref. [202] uses a Stackelberg leader-follower game to model strategic behaviors of distribution companies in the wholesale energy and reserve markets. The uncertainty of wind and solar power generation is modeled by scenarios. In ref. [197], a distribution and transmission coordination framework is also formulated as a leader-follower game. The uncertainty of the reserve response for a service call and the reserve service delivery failure is modeled by two probability indexes. A sensitivity analysis is conducted by sampling within their ranges and calculating for each scenario.

However, the computation cost of the scenario-based approach increases with the number of scenarios, which becomes a significant bottleneck when involving more players in the game. An alternative method is the analytical approaches, such as CC [203], CVaR [204], and distributionally robust formulations [205, 136]. These formulations assume an underlying distribution or a family of distributions for uncertain variables and model the expected market outcomes at a certain chance. This allows the introduction of risk-aware metrics such as the confidence level of power generation and energy service provision failure rates [203, 206]—for example, refs. [144, 136] model uncertain energy and reserve markets using a DRCC formulation. The DRCC models envisage reserve service from thermal power plants is compulsory, and the operator only manages the chances of different technical violation events such as line violations. However, it does not simulate regulations for contract-based market players such as distributed renewable generations. As the players' strategies are interlinked in a game, the market equilibrium is the joint solution of all players. Therefore, uncertainties from all players' strategies should be modeled jointly, forming a JCC problem.

In terms of modeling renewable energy sources in the CC market game, ref. [203] uses the CC formulation in a Nash-Cournot game to simulate the impact of uncertain wind power outputs in the electricity market considering the congestion arbitrage. This work shows that the profits of wind farms largely depend on the confidence level of wind power generation predictions. This work found that the profits of wind farms largely rely on the confidence level of wind power generation predictions. However, it simply assumes all the wind farms are independent and have the same confidence levels in predictions. Recent work [126] uses a DRJCC formulation to model the generalized Nash equilibrium problem to coordinate renewable energy aggregators in the local electricity market. The work considers the maximum power outputs of aggregators as an uncertain distribution.

Then the model solves their bidding strategies and market equilibrium using the best response algorithm. Nevertheless, all the renewable energy aggregators are assumed to be identical in bidding and have the same best response function. Furthermore, the paper does not discuss the correlation between players' decisions and their uncertainties, although the game is simulated with varying numbers of players.

DRJCC problems generally cannot be solved directly. However, their solutions can be approximated by decomposing the JCC into tractable single CCs and enforcing *Boole's inequality*, meaning that the chance that at least one of the events happens is no greater than the sum of the possibilities of all individual events. A particular case is the *Bonferroni approximation*, which assumes that these CCs are entirely independent and the sum of individual violation rates is equal to the joint violation rate [205, 207]. However, CCs with the same uncertainty source (e.g., renewable power sources within an area) are usually positively correlated. This assumption neglects their correlation and leads to high solution conservativeness. Ref. [207] proves the solution conservativeness of the JCC problem will increase with the number of decomposed CCs if their violation rates are obtained using the *Bonferroni approximation*. Ref. [132] proves that if these constraints are completely correlated, the efficacy of the *Bonferroni approximation* will diminish.

To identify the correlation between CCs and reduce solution conservativeness, a common method is pre-solving the problem with a conservative guess, then progressively tightening the worst-case bounds. For example, ref. [132] proposed a tractable CVaR formulation to model the worst-case bound of the JCC problem. An auxiliary variable is introduced to tighten the bound and solved iteratively with decision variables. This optimized CVaR method has been applied in the energy and reserve joint dispatch problem [144]. Refs. [124] and [125] solve the OPF problem using JCC. To obtain a low-conservative solution, they estimate analytically and eliminate the overlapped violation events due to the correlated CCs in an iterative framework. These works effectively reduce solution conservativeness but require a comprehensive analysis of violation events in the network. Recent work [208] incorporates the correlation in power generations of wind farms in the ambiguity set as a constraint in the DRO formulation. However, this requires the specified correlation that is hard to obtain in a changing power market environment.

This chapter designs a DRJCC game framework to coordinate solar-powered microgrids in the energy and reserve markets and proposes a novel Bayesian approach to approximate the optimal solution. Compared with stochastic game designs in the existing literature, this game framework presents the following novel features,

1) A two-stage DRO formulation is employed to schedule the energy and reserve for microgrids, considering the worst-case reserve service performance. This game framework uses renewable forecast errors and VoLL to quantify the load-shedding penalty due to the under-delivered reserve. The interactions of microgrids in the real-time energy and reserve markets are simulated in a non-cooperative game.

2) A JCC regulates the under-delivered reserve of all microgrids jointly. To avoid an over-conservative reserve regulation, a novel Bayesian optimization method is proposed to approximate individual violation rates of decomposed DRCCs, considering the changing correlation in the market game. Compared with the latest work [209], it is for the first time using Bayesian optimization to solve the DRJCC problem.

3) Case studies in the 30-bus network use California power market data. The economic analysis of simulations with the varying number of players shows the trade-off between energy and reserve capacity offers of microgrids. It also shows that the capacity range where the reserve offer is profitable.

5.3 DRJCC market design considering uncertain reserve provision

This game framework is demonstrated under the two-settlement electricity market designs used in the U.S. and Europe.

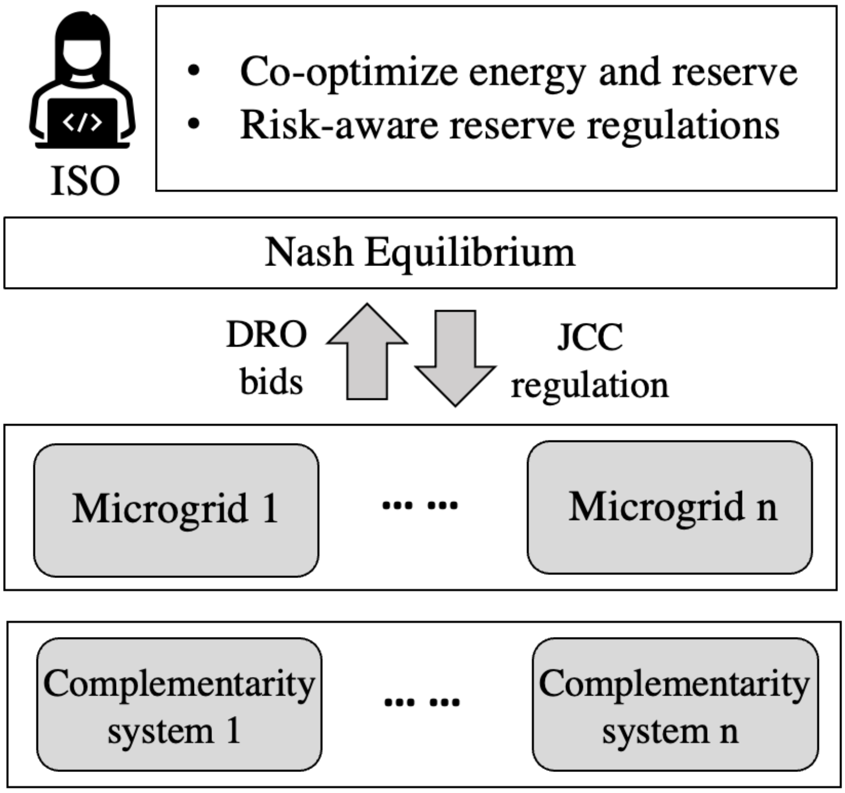

Figure 5.1: Bi-level game framework for the energy and reserve market clearing

As presented in Fig.5.1, the proposed DRJCC market design is used for the joint dispatch of energy and reserve. An independent system operator (ISO) manages the transmission network and acquires probabilistic net load forecasts (i.e., power demands subtracting solar power generations) at all network nodes. In the day-ahead market, ISO sets energy and reserve requirements. It receives offers from thermal generators and microgrids, while microgrids submit bids incorporating costs and service provision capacities. Due to renewable generation uncertainty, microgrid reserve services might under-deliver in the real-time market, leading to unexpected load shedding. Instead of being a price-taker, each microgrid can leverage the historical forecast data and adjusts bids strategically to avoid losses. Meanwhile, ISO regulates reserve provision performances to ensure system reliability.

5.4 Model formulation

This market design in Fig. 5.1 is simulated as a one-leader-multiple-follower stochastic game. Under this model setting, ISO knows all the bidding curves submitted by microgrids, but each microgrid does not learn the upper-level market clearing and others' bidding strategies. The DRJCC formulation models uncertain reserve provisions performance due to solar forecast errors. This game framework includes two types of players. ISO as a leader, co-optimizes energy and reserve and regulates uncertain reserve service provisions. As followers, microgrids schedule and conduct strategic bidding. To model interactions of all these market players, these models are reformulated into a complementarity model that can represent the simultaneous optimization problems of one or several decision-makers [210].

5.4.1 Wasserstein-metric ambiguity set considering the worst-case reserve performance

Assuming that microgrids are mainly supplied by renewable power sources such as solar panels or wind farms, their real-time reserve provision performances are modeled based on historical renewable forecast error samples.

Fig. 5.2 shows the distribution of errors between the day-ahead and real-time solar power forecasts from California independent system operator (CAISO) from 2021 to 2022 [6]. Two real-time market intervals at 15 and 45 mins in each hour are chosen to calculate these errors. I build a Wasserstein-metric ambiguity set $\mathscr{P}$ using these samples to model the worst-case reserve performance, which is moment-free and built upon historical error samples. The error samples of the Wasserstein-metrics ambiguity set could be updated depending on the available information. This will account for the impacts of solar power generation and forecast techniques. .

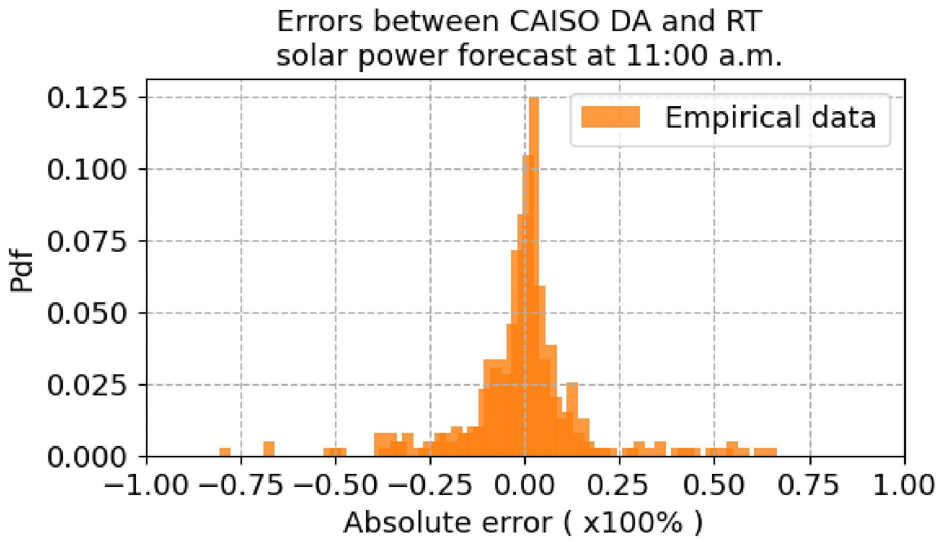

Figure 5.2: The distribution of errors between the day-ahead and real-time solar power forecasts at 11:00 a.m.

Let the error sample vector $s := \{s_1, s_2, ..., s_n\}$ be a vector including n random samples of the uncertain variable ζ at time interval t. If $s_1 \sim \mathbb{P}_1$ and $s_2 \sim \mathbb{P}_2$, the Wasserstein metric $W(\mathbb{P}_1, \mathbb{P}_2)$ between two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined as [211],

$$W(\mathbb{P}_1, \mathbb{P}_2) = \inf\{\int_{\mathscr{R}(\Xi)} ||s_1 - s_2|| \mathscr{Q}(ds_1, ds_2)\} \tag{5.1}$$

The term $||s_1 - s_2||$ is the distance between two samples, and $\mathscr{Q}(ds_1, ds_2)$ is the joint distribution with marginal distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ on the support $\mathscr{R}(\Xi)$. The Wasserstein-metric ambiguity set $\mathscr{P}$ is then defined as a ball space of radius $\nu \geq 0$ concerning Wasserstein distance, centered at a prescribed reference distribution $\mathbb{P}_\zeta$.

$$\mathscr{P} := \{\mathbb{P} : \mathbb{P} \in \mathscr{R}(\Xi), W(\mathbb{P}, \mathbb{P}_\zeta) \leq \nu\} \tag{5.2}$$

where $\mathbb{P}_\zeta$ is constructed based on samples s.

5.4.2 ISO's optimization problem for the energy and reserve joint dispatch

This work considers a power network with N_B buses and N_{LN} lines. Its bus susceptance matrix denotes as $\mathbf{B}_{bus} \in \mathbb{R}^{N_B \times N_B}$. Flow susceptance matrix denotes as $\mathbf{B}_{line} \in \mathbb{R}^{N_{LN} \times N_B}$

to calculate the difference in voltage angle across each line. Thermal generations, loads, and microgrids are located on different buses. Their positions in the network are mapped by sparse matrices $\mathbf{C}^g \in \mathbb{R}^{N_B \times N_G}$, $\mathbf{C}^l \in \mathbb{R}^{N_B \times N_L}$ and $\mathbf{C}^{mg} \in \mathbb{R}^{N_B \times N_{MG}}$. All the vectors in **bold** without specification represent the vector of a variable or a parameter across the network over the optimization horizon (e.g., $\mathbf{P}_{nl}$ is all the net loads over the period $\mathscr{T}$). Given a real-time market clearing interval $\Delta t \in \mathscr{T}$, the multi-period power dispatch considering the network is formulated as follows.

1) *The objective function:* The cost function of ISO's model includes energy and reserve capacity offers from thermal generations and microgrids. The energy cost of thermal generations is represented as a quadratic cost function, and the reserve cost is a linear function of their reserve capacity offers. The cost function also includes energy and reserve capacity offers from microgrids.

$$\begin{aligned}(M1)\min \mathscr{J}^H = \min\{\sum_{t\in\mathscr{T}}\sum_{k\in\mathscr{K}}(m_k^g(\Delta t P_{t,k}^g)^2 + \Delta t m_k^{gr} R_{t,k}^g)\\ + \sum_{t\in\mathscr{T}}\sum_{i\in\mathscr{I}}\Delta t(\pi_i^r R_{t,i}^{up} - \pi_i^e P_{t,i}^{ex})\}\end{aligned} \tag{5.3}$$

2) *Power and reserve balances*: Power balance is required, and the total reserve capacity offer must be greater than the reserve requirement.

$$\sum_{k\in\mathscr{K}} P_{t,k}^g + \sum_{i\in\mathscr{I}} P_{t,i}^{ex} = \sum_{j\in\mathscr{J}} P_{t,j}^{nl} \quad t\in\mathscr{T} \tag{5.4}$$

$$\sum_{i\in\mathscr{K}} R_{t,k}^g + \sum_{i\in\mathscr{I}} R_{t,i}^{up} \geq R_t^{req} \quad t\in\mathscr{T} \tag{5.5}$$

Energy and reserve provisions are non-negative, and the sum should be less than the power capacity.

$$\mathbf{P}^g + \mathbf{R}^g \leq \overline{\mathbf{P}^g} \tag{5.6}$$

$$[\mathbf{P}^g, \mathbf{R}^g] \geq 0 \tag{5.7}$$

3) *Power injections and line limits*: The net power injection is the sum of thermal power generations, reserve provisions, load consumptions, or power exchanged with microgrids at each node.

$$\mathbf{P}^{inj} = \mathbf{C}^{g}(\mathbf{P}^{g} + \mathbf{R}^{g}) - \mathbf{C}^{mg}(\mathbf{P}^{ex} - \mathbf{R}^{up}) - \mathbf{C}^{l}\mathbf{P}^{nl} \tag{5.8}$$

Power flow is regulated within the thermal power capacity of each line.

$$\mathbf{B}_{line}\mathbf{B}_{bus}^{-1}\mathbf{P}^{inj} \leq \overline{\mathbf{L}} \tag{5.9}$$

$$\mathbf{B}_{line}\mathbf{B}_{bus}^{-1}\mathbf{P}^{inj} \geq -\overline{\mathbf{L}} \tag{5.10}$$

4) *Reserve performance regulations based on day-ahead contracts*: ISO regulates the reserve performance of all activated microgrids for a reserve call based on their day-ahead contracts. If multiple microgrids fail to deliver the reserve simultaneously, the system might have a severe contingency. Considering the Wasserstein-metric ambiguity set $\mathscr{P}$, the total amount of under-delivered reserve of each microgrid in the worst case over the period $\mathscr{T}$ should not exceed the limit $\overline{R^{fup}}$ at $1-\varepsilon^{jot}$ chance, as agreed in the day-ahead market. The regulation based on these day-ahead contracts can be formulated as a set of DRCCs where ε^{jot} represents the joint risk of breaching contracts,

$$\inf_{\mathbb{P}\in\mathscr{P}} \mathbb{P}\Big(\bigcap_{i\in\mathscr{I}} \sum_{t\in\mathscr{T}} \Delta t \zeta R_{i,t}^{up} \leq \overline{R^{fup}}\Big) \geq 1-\varepsilon^{jot} \tag{5.11}$$

The DRJCC (5.11) is first decomposed into a set of single DRCCs. Let the individual violation rate of each microgrid is ε_i^{ind}, the decomposed CC is given by,

$$\inf_{\mathbb{P}\in\mathscr{P}} \mathbb{P}\Big(\sum_{t\in\mathscr{T}} \Delta t \zeta R_{i,t}^{up} \leq \overline{R^{fup}}\Big) \geq 1-\varepsilon_i^{ind} \quad i\in\mathscr{I} \tag{5.12}$$

As the proof (Theorem 2.2) in ref. [177], each DRCC can be approximated as a CVaR constraint. This CVaR constraint is then reformulated into the following convex constraints based on refs. [212, 136], for each microgrid $i \in \mathscr{I}$.

$$-\eta_i^u - \frac{1}{\varepsilon_i^{ind}}\Big(\nu l_i^u + \frac{1^\top \mathbf{W}_i^u}{N_S}\Big) \geq 0 \tag{5.13}$$

$$\mathbf{W}_i^u - \Delta t \mathbf{s}(\mathbf{R}_i^{up})^\top + \overline{R^{fup}} + \eta_i^u \geq 0 \tag{5.14}$$

$$\mathbf{W}_i^u \geq 0 \tag{5.15}$$

$$||\Delta t \mathbf{s}(\mathbf{R}_i^{up})^\top|| \leq l_i^u \tag{5.16}$$

Where ν is the radius of the Wasserstein-metric ambiguity set. The matrices $\mathbf{R}_i^{up} \in \mathbb{R}^{1\times N_T}$ and $\mathbf{s} \in \mathbb{R}^{N_S \times N_T}$ denote the reserve vector of microgrid i and the empirical reserve forecast error matrix. The variables η^u, l^u, $\mathbf{W}^u \in \mathbb{R}^{N_S\times 1}$ are auxiliary variables. For the unknown ε_i^{ind}, instead of enforcing any conservative assumption, its range is set as below.

$$\varepsilon_i^{ind} \in [\frac{\varepsilon^{jot}}{N_{MG}}, \varepsilon^{jot}] \tag{5.17}$$

Two bounds represent two extreme scenarios. The lower bound (i.e., $\frac{\varepsilon^{jot}}{N_{MG}}$) is based on the *Bonferroni approximation* where all the single CCs have equal violation rates divided from the joint violation rate. It assumes no positive correlation in the reserve provision performance of microgrids. The upper bound (i.e., ε^{jot}), on the contrary, assumes all the microgrids have the same reserve provision performance.

5.4.3 Data-driven bidding considering the expected reserve performance under the worst error distribution

Microgrids schedule their energy and reserve bids in a risk-aware manner using historical solar forecast error data. A two-stage DRO model is thus formulated considering the worst-case reserve provision performance. The following subsections present the optimization model for each microgrid $i \in \mathscr{I}$.

1) *The objective function:* The cost of each microgrid includes three parts. The first part is the cost of energy and reserve provisions, and the second is the revenue from energy and reserve service trading. The final part models the economic penalty due to the unavailable reserve service. Here we assume the penalty as the expected load-shedding loss in the worst case, using the DRO formulation based on the Wasserstein-metric ambiguity set $\mathscr{P}$.

$$
\begin{aligned}
(M2)\min \mathscr{J}^L = \min\{\sum_{t\in\mathscr{T}} (m_i^s(\Delta t P_{i,t}^s)^2 + \Delta t m_i^{mgr} R_{i,t}^{up} \\
+\Delta t \pi_i^e P_{i,t}^{ex} - \Delta t \pi_i^r R_{i,t}^{up}) \\
+\mathbb{E}_{\mathbb{P}\in\mathscr{P}}[\Delta t \zeta R_{i,t}^{up} V_i]\}
\end{aligned}
\tag{5.18}
$$

The load-shedding loss due to the under-delivered reserve is calculated by multiplying the under-delivered reserve capacity and the per-unit VoLL (i.e.,V_i) of each microgrid.

2) *Convex reformulation of DRO:* Similar to the reformulation in Section *3.2*, the DRO term in the objective function can be reformulated, as Corollary 5.4 in ref. [212].

$$\mathbb{E}_{\mathbb{P}\in\mathscr{P}}[\sum_{t\in\mathscr{T}} \Delta t \zeta R_{i,t}^{up} V_i] = \nu l_i^l + \frac{1^\top \mathbf{W}_i^l}{N_S} \tag{5.19}$$

$$(\mu^{dro1}) \quad \mathbf{W}_i^l - \Delta t \mathbf{s}(\mathbf{R}_i^{up})^\top V_i + \eta_i^l \geq 0 \tag{5.20}$$

$$(\mu^{dro2}) \quad \mathbf{W}_i^l \geq 0 \tag{5.21}$$

$$(\mu^{dro3}) \quad \Delta t \mathbf{s}(\mathbf{R}_i^{up})^\top V_i + l_i^l \geq 0 \tag{5.22}$$

$$(\mu^{dro4}) \quad \Delta t \mathbf{s}(\mathbf{R}_i^{up})^\top V_i - l_i^l \leq 0 \tag{5.23}$$

where l^l, η^l and $\mathbf{W}^l \in \mathbb{R}^{N_S\times 1}$ are auxiliary variables in the reformulation. The dual variables in the bracket, μ^{dro1},μ^{dro2}, μ^{dro3} and μ^{dro4} all have the same dimension $\mathbb{R}^{N_S\times 1}$.

3) *Power balance and provision*: Each microgrid requires power balance. Power sources in microgrids are considered flexible power generations, which consist of renewable power sources and ES. Microgrids can thus import and export power. In the power markets, ISO can only know energy and reserve offers from microgrids instead of their dispatch. Therefore, this model does not include the energy constraints of ES and the power flow for each microgrid.

$$(\mu^{pb}) \quad P_{i,t}^{ex} + P_{i,t}^s = P_{i,t}^{nl} \quad t\in\mathscr{T} \tag{5.24}$$

$$(\mu^{dnpl}, \mu^{uppl}) \quad -\underline{P^s} \leq P_{i,t}^s \leq \overline{P^s} \quad t\in\mathscr{T} \tag{5.25}$$

4) *Reserve provision*: The upper limit of reserve provision is set as a fraction of the flexible power generation level.

$$(\mu^{dnrl}, \mu^{uprl}) \quad 0 \leq R^{up}_{i,t} \leq \gamma P^{s}_{i,t} \quad t \in \mathscr{T} \tag{5.26}$$

5.4.4 Complementarity model for market equilibrium

lSO and microgrids' optimization are interlinked in the quantity of energy and reserve capacity offers (i.e., $\pi^r R^{up} - \pi^e P^{ex}$). To form a one-leader-multiple-follower problem, the MPEC technique is employed to solve two types of optimizations simultaneously and obtain market equilibrium. Based on the strong duality and KKT rule, the dual problem of model M2 is formulated and integrated into the upper-level model M1. This game considers N_{MG} of microgrids over the period $\mathscr{T}$. The indexes i,t are omitted for conciseness. $\mathbf{1}^s$ represents a unit vector with the same dimension as the error sample vector s.

First, stationary constraints are formulated for six decision variables $P^s, P^{ex}, R^{up}, l^l, W^l$ and η^l in model M2.

$$2\Delta t m^s P^s + \mu^{pb} - \mu^{dnpl} + \mu^{uppl} - \gamma \mu^{uprl} = 0 \tag{5.27}$$

$$\pi^e + \mu^{pb} = 0 \tag{5.28}$$

$$\begin{aligned} \Delta t(m^{mgr} - \pi^r) + \Delta t s^\top (\mu^{dro1} - \mu^{dro3} + \mu^{dro4})V \\ - \mu^{dnrl} + \mu^{uprl} = 0 \end{aligned} \tag{5.29}$$

$$\nu - (\mathbf{1}^s)^\top \mu^{dro3} - (\mathbf{1}^s)^\top \mu^{dro4} = 0 \tag{5.30}$$

$$\frac{1}{N_S} - \mu^{dro1} - \mu^{dro2} = 0 \tag{5.31}$$

$$(\mathbf{1}^s)^\top \mu^{dro1} = 0 \tag{5.32}$$

Complementary slackness constraints are,

$$0 \leq (\mathbf{W}^l - \Delta t \mathbf{s}(\mathbf{R}^{up})^\top V + \eta^l) \perp \mu^{dro1} \geq 0 \tag{5.33}$$

$$0 \leq \mathbf{W}^l \perp \mu^{dro2} \geq 0 \tag{5.34}$$

$$0 \leq (\Delta t \mathbf{s}(\mathbf{R}^{up})^\top V + l^l) \perp \mu^{dro3} \geq 0 \tag{5.35}$$

$$0 \leq (l^l - \Delta t \mathbf{s}(\mathbf{R}^{up})^\top V) \perp \mu^{dro4} \geq 0 \tag{5.36}$$

$$0 \leq (\overline{P^s} - P^s) \perp \mu^{uppl} \geq 0 \tag{5.37}$$

$$0 \leq (\underline{P^s} + P^s) \perp \mu^{dnpl} \geq 0 \tag{5.38}$$

$$0 \leq R^{up} \perp \mu^{dnpl} \geq 0 \tag{5.39}$$

$$0 \leq (\gamma P^s - R^{up}) \perp \mu^{uppl} \geq 0 \tag{5.40}$$

The big-M technique is used to solve non-linear constraints (5.33) - (5.40). Assuming a and b are variables, the non-linear constraint on the left-hand side can be transformed into the four convex constraints on the right-hand side.

$$0 \leq a \perp b \geq 0 \Rightarrow a \geq 0, b \geq 0, a \leq MU, b \leq M(1-U) \tag{5.41}$$

where M is a number far larger than a and b, and U has the same dimension as a and b.

The objective function of ISO's optimization problem is nonlinear after reformulation, as it has a multiplication of price and capacity variables. The objective function is linearized based on the strong duality theory. The reformulated model for solving the market equilibrium of ISO and N_{MG} of microgrids is as follows.

$$\begin{aligned} (M3) \quad \min_{\Xi_{bi}} \mathscr{J} = \min\{ & \sum_{t\in\mathscr{T}} \sum_{k\in\mathscr{K}} (m_k^g (\Delta t P_{t,k}^g)^2 + \Delta t m_{t,k}^{gr} R_{t,k}^g) \\ & + \sum_{t\in\mathscr{T}} \sum_{i\in\mathscr{I}} [m_i^s (\Delta t P_{t,i}^s)^2 + \Delta t m_i^{mgr} R_{t,i}^{up} + (\nu l_i^l + \frac{1^\top \mathbf{W}_i^l}{N_S}) \\ & - \mu_{t,i}^{pb} P_{t,i}^l + \mu_{t,i}^{uppl} \overline{P^s} + \mu_{t,i}^{dnpl} \underline{P^s}]\} \end{aligned} \tag{5.42}$$

$$\text{s.t.} \quad \text{equations} \quad (5.4)-(5.16) \quad \text{and} \quad (5.27)-(5.40)$$

where the variable set Ξ_{bi} includes P^g, R^g, P^s, R^{up}, R^{ex}, π^r, π^e, l^l, η^l, $\mathbf{W}^l$, η^u, l^u, $\mathbf{W}^u$, ε^{ind}, μ^{dro1}, μ^{dro2}, μ^{dro3}, μ^{dro4}, μ^{pb}, μ^{dnpl}, μ^{uppl}, μ^{dnrl}, μ^{uprl}.

5.5 Approximating the optimal solution using Bayesian optimization

The reformulated model M3 is intractable with unknown optimal individual violation rates ε^{ind}. Bayesian optimization is introduced to approximate the optimal solution iteratively. Bayesian optimization is a global approximation approach to optimize the black-box function. Its applications include hyper-parameter tuning for deep learning models [213] and solving difficult combinatorial optimizations [214]. Bayesian optimization best suits the expensive-to-evaluate functions and tolerates the stochastic noise in function evaluations [214]. In this work, the function of the individual violation rates is unknown, and the observations from stochastic out-of-sample tests are considered noisy. The following subsections summarize the algorithm, the design of the evaluation, and acquisition functions.

5.5.1 Summary of the algorithm

Algorithm 2: Solving M3 using Bayesian optimization

Initialize model M3 and define the exploration space;
Evaluate the initial samples ε_o;
for $m = 1, \ldots.., N_{IT}$ **do**
  Update the surrogate model with observations $\mathbf{O}_m$;
  Identify new samples $\varepsilon_{m+1} = \underset{\varepsilon}{\operatorname{argmax}} \mathscr{G}(\varepsilon_m; \mathbf{O}_m)$;
  Query the evaluation function $\mathscr{H}(\varepsilon_{m+1})$;
  Augment observations $\mathbf{O}_{m+1} := \{\mathbf{O}_m; \mathscr{H}(\varepsilon_{m+1})\}$;
end

As in Algorithm 1, I first initialize the problem to solve, model M3, and define the exploration space of the individual violation rates $\varepsilon_m \in [\frac{\varepsilon^{jot}}{N_{MG}}, \varepsilon^{jot}]$. The dimension of the sample vector is N_{MG}, and the bounds ensure that the evaluated black-box function is feasible and continuous over the space. Then, a Gaussian process regressor (GPR) is used as the surrogate model for fitting the unknown block-box function within N_{IT}

iterations. GPR assumes a prior distribution for the approximated value $\mathbb{P}(\varepsilon_m)$. Given the distributions of observations $\mathbb{P}(\mathbf{O}_m)$ and the likelihood $\mathbb{P}(\mathbf{O}_m|\varepsilon_m)$ (i.e., the maximum difference between the estimation and the ground truth), the posterior of GPR (i.e., the probability distribution of the approximated value) is defined as,

$$\mathbb{P}(\varepsilon_m|\mathbf{O}_m) = \frac{\mathbb{P}(\mathbf{O}_m|\varepsilon_m)\mathbb{P}(\varepsilon_m)}{\mathbb{P}(\mathbf{O}_m)} \tag{5.43}$$

In each iteration m, the posterior is updated using all observations $\mathbf{O}_m$. After that, new samples are proposed for the next iteration using the acquisition function $\mathscr{G}(\varepsilon_m;\mathbf{O}_m)$. These samples are evaluated in the function $\mathscr{H}(\varepsilon_{m+1})$ and added to the observation $\mathbf{O}_{m+1}$. This procedure is executed for N_{IT} iterations.

5.5.2 Evaluation of the black-box function

The optimal reserve regulation policy (5.11) aims to find the optimal individual violation rates ε^{ind} so that the empirical outcome ε^e_m is as close as to the predefined requirement ε^{jot}. The evaluated function $\mathscr{H}(\varepsilon_m)$ is thus defined as the absolute difference between the empirical and predefined joint violation rate.

$$\mathscr{H}(\varepsilon_m) = |\varepsilon^e_m - \varepsilon^{jot}| \tag{5.44}$$

The empirical joint violation rate is obtained from $N_{OS} = 150$ independent out-of-sample tests. In each test, an error sample s different from those in the ambiguity set is used to simulate the under-delivered reserve $R^{under}_s := \Delta ts(\mathbf{R}^{up})^\top$ for each microgrid. A violation event is when any microgrid violates the reserve regulation (i.e., constraint (5.12)). The empirical joint violation rate is calculated as the ratio of the total number of violation events to the total number of tests.

$$\varepsilon^e_m := \frac{\sum_{s=0}^{N_{OS}} 1_{(R^{under}_s > \overline{R^{fup}})}}{N_{OS}} \tag{5.45}$$

5.5.3 Sample selection using the acquisition function

The acquisition function decides the new sample in the iteration m. Different acquisition functions include the lower confidence bound, the probability of improvement, and the expected improvement [215]. We choose the expected improvement after experiments considering the explore-exploit tradeoff. The function is the maximum expected improvement at the sample ε_m estimated from all out-of-sample tests.

$$\mathscr{G}(\varepsilon_m; \mathbf{O}_m) = \mathbb{E}[\mathscr{H}^{'} - \mathscr{H}(\varepsilon_m)]^+ \tag{5.46}$$

Where $\mathscr{H}^{'}$ denotes the minimum value obtained in the evaluation function so far. The new sample in the next iteration gives the maximum value of the acquisition function.

$$\varepsilon_{m+1} = \underset{\varepsilon}{\text{argmax}} \mathscr{G}(\varepsilon_m; \mathbf{O}_m) \tag{5.47}$$

5.6 Case studies

Two case studies are developed using the IEEE 30-bus network [216]. As shown in Fig. 5.3, the power network consists of passive loads and proactive microgrids, marked in black arrows and orange stars, respectively. Each microgrid has a 2MW load on average, a 2MW solar PV generation, and 3MW/6MWh ES system. The parameters of the game framework and Bayesian optimization are presented in Table 5.1.

Table 5.1: The parameters of the game framework and Bayesian optimization

m_s	\$1/(MWh)^2	$\overline{R^{fup}}$	1.5 MWh	Δt	0.5 hour
m_g	\$1/(MWh)^2	$\overline{P^s}$	3 MW	N_{OS}	150
m_{gr}	\$15/MWh	γ	0.5	N_{IT}	20
m_{mgr}	\$5/MWh	ε^{jot}	0.2	N_S	200

Net-load profiles are taken from the UK power networks trial [217] and the deep learning model in ref. [218] generates their probabilistic forecasting profiles for reserve requirements. Fig. 5.4 shows the net load forecast and system reserve requirements. Due to solar power and peak loads, reserve requirements are higher from 6 a.m. to 6 p.m. We choose the reserve requirement R^{req} at the reliability level of 90%.

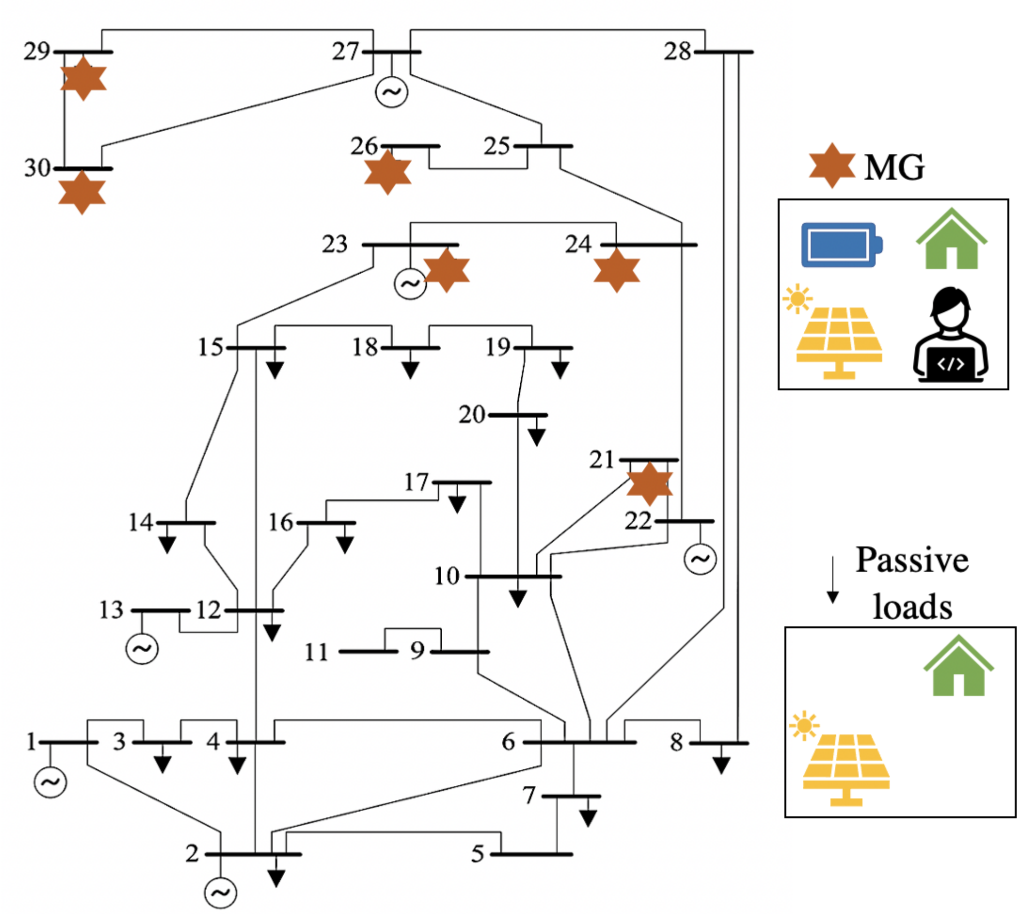

Figure 5.3: The 30-bus power network with microgrids

The first case study simulates the game framework with a fixed number of six microgrids. As shown in Fig. 5.3, microgrids 1-6 are located at buses 21, 23, 24, 26, 29, and 30, respectively. Microgrids 1 and 2 at buses 21 and 23 have a higher VoLL of \$20/MWh, and the other microgrids have a VoLL of \$1/MWh. We conduct the simulations with different radii of the Wasserstein-metrics ambiguity set $\mathscr{P}$ and aim to find the optimal radius given the predefined requirement ε^{jot}. Based on the optimal radius, the second case study simulates the game framework under the varying number of microgrids, ranging from 2 to 14. The correlation of these microgrids changes as the

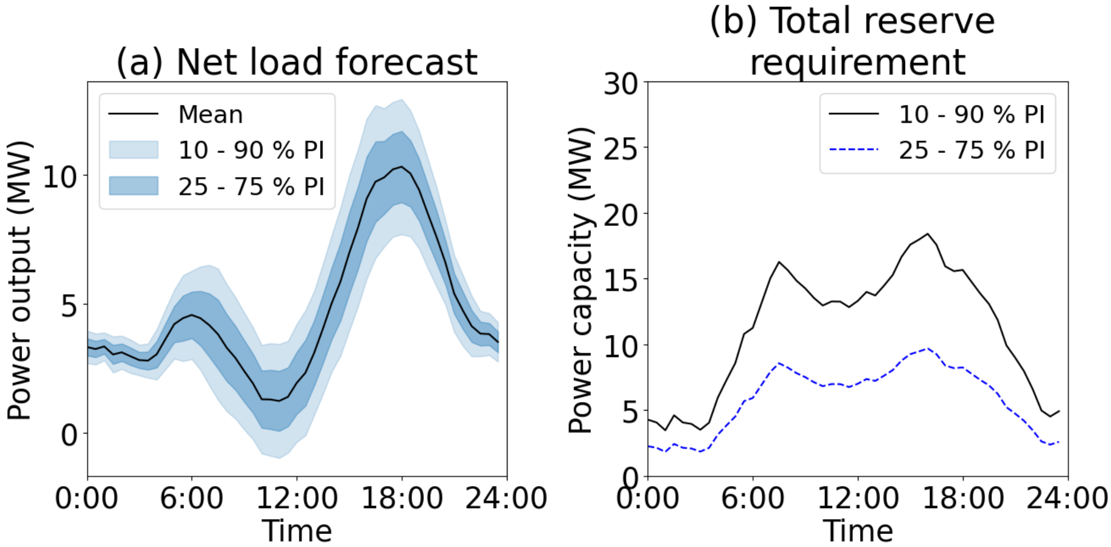

Figure 5.4: (a) An example of probabilistic net load forecast at one bus and (b) total reserve procurement requirements at the reliability levels of 75% and 90%

number of players changes. All the microgrids have the same VoLL of \$1/MWh.

5.7 Numerical results

The proposed DRJCC game framework is simulated for six hours from 6:00 a.m. to 12:00 a.m. This game framework is built using CVXPY package [219] in Python, and the Bayesian optimization is implemented using the scikit-optimize package [220]. The following subsections present results and analysis of two case studies.

5.7.1 The optimal radius of Wasserstein-metrics ambiguity sets

The radius of the Wasserstein-metrics ambiguity determines solution conservativeness, in other words, the conservativeness of reserve regulations. The radii of ambiguity set are the choice of ISO and microgrids, which decides the conservativeness of system reserve regulation. ISO can choose the radii based on grid conditions and system security requirements. In the first case study, simulations are conducted with an increasing radius ranging from 10^{-5} to 10^{-1} to find the optimal radius at the given joint violation rate. For comparison, the benchmark case is simulated without the system-wise reserve regulation, in other words, disabling constraints (5.13) - (5.16).

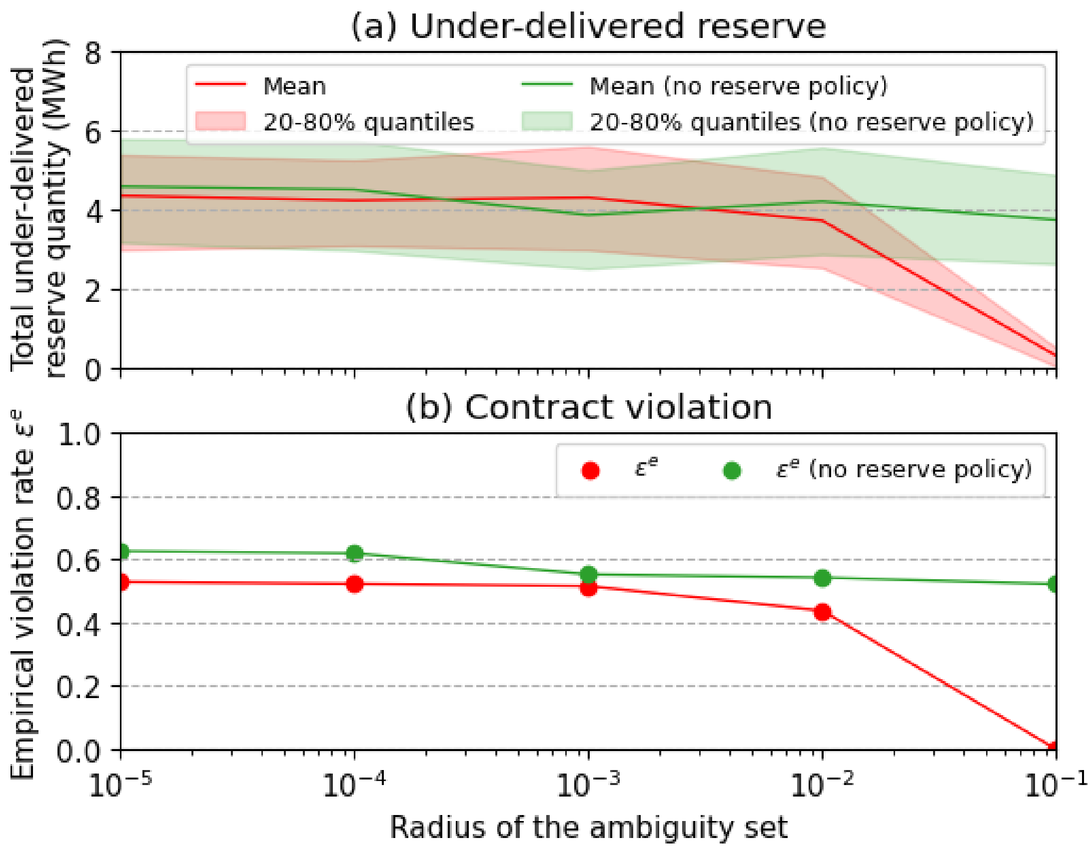

Figure 5.5: The total amount of under-delivered reserve and empirical joint violation rates with an increasing radius of the ambiguity set

Fig. 5.5 (a) and (b) show the under-delivered reserve and contract violation rates in $N_{OS} = 150$ out-of-sample tests. The line is the mean value of the under-delivered reserve from all tests, and the shade shows the 20-80% quantile. The DRCC reserve policy is effective when the radius is between 10^{-2} and 10^{-1}, where the game results differ from the benchmark case (i.e., no reserve policy) significantly. I choose an optimal radius of $\nu = 0.035$ for a joint violation rate requirement of 0.2.

Using the mean values in out-of-sample tests, the average costs of ISO and six microgrids are calculated as the objective functions of models M1 and M2, respectively. Fig. 5.6 summarizes results in five simulations. The costs of microgrids 1 and 2 reduce

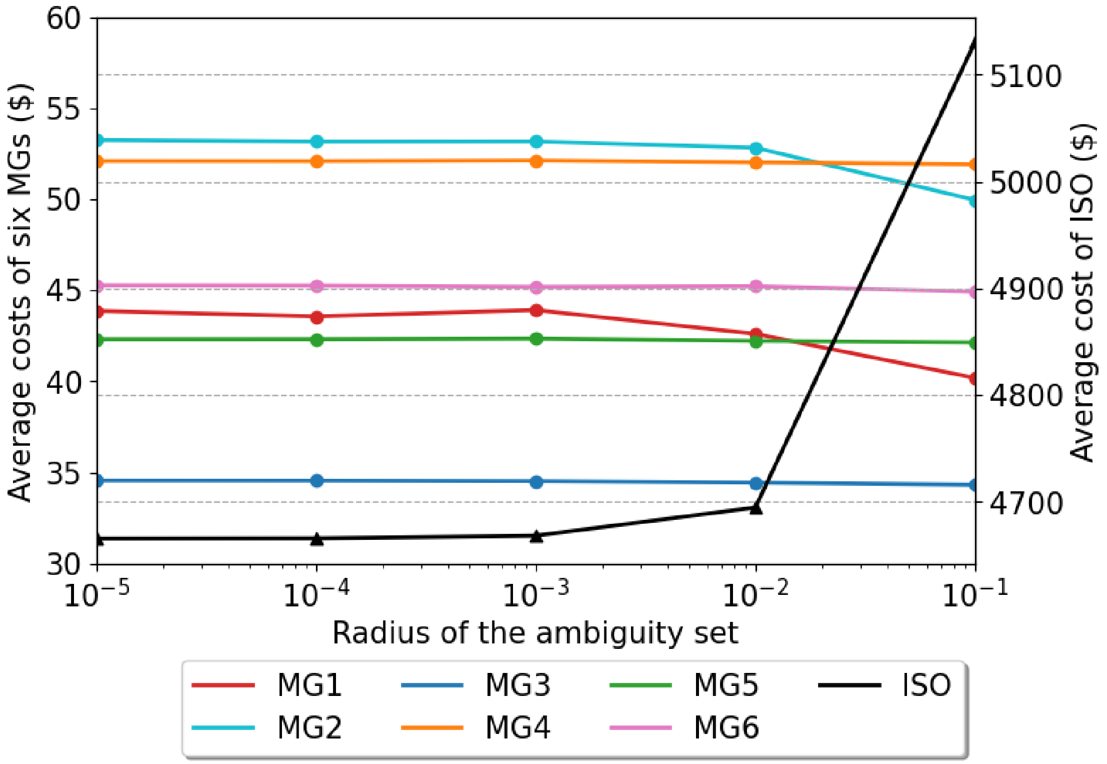

Figure 5.6: Average costs of ISO and microgrids with an increasing radius of the ambiguity set

when the radius increases from 10^{-3} to 10^{-1}, while the costs of the other four microgrids almost remain constant with the radius increases. This is because microgrids 1 and 2 have the higher VoLL and tend to be more risk-aware (i.e., bid less) when the ambiguity set becomes more conservative (i.e., larger). The average cost of ISO increases rapidly as the ambiguity set becomes larger, as ISO needs more expensive reserve services from thermal generators.

5.7.2 Game results with the increasing number of players

Based on this optimal radius, this game framework is simulated with a varying number of players in the second case study. For each simulation, the game increases two microgrid players (from passive net loads at random buses) until the number of players reaches 14. Since microgrids have the same VoLL, their violation rates are assumed to be the same (i.e., $\varepsilon_0^{ind} = \varepsilon_1^{ind} =......= \varepsilon_n^{ind}$). For comparison, two scenarios are included considering the upper and lower bounds of the equation (5.17). The scenario for the lower bound (i.e., $\varepsilon^{ind} := \frac{\varepsilon^{jot}}{N_{MG}}$) denotes C1, and the scenario for the upper bound (i.e., $\varepsilon^{ind} := \varepsilon^{jot}$) denotes C2. The proposed method C3 is between these two scenarios.

Table 5.2: The optimized individual violation rates

Players	2	4	6	8	10	12	14
$\varepsilon^{ind}(\%)$	20.00	15.02	13.73	12.54	11.10	10.42	9.82

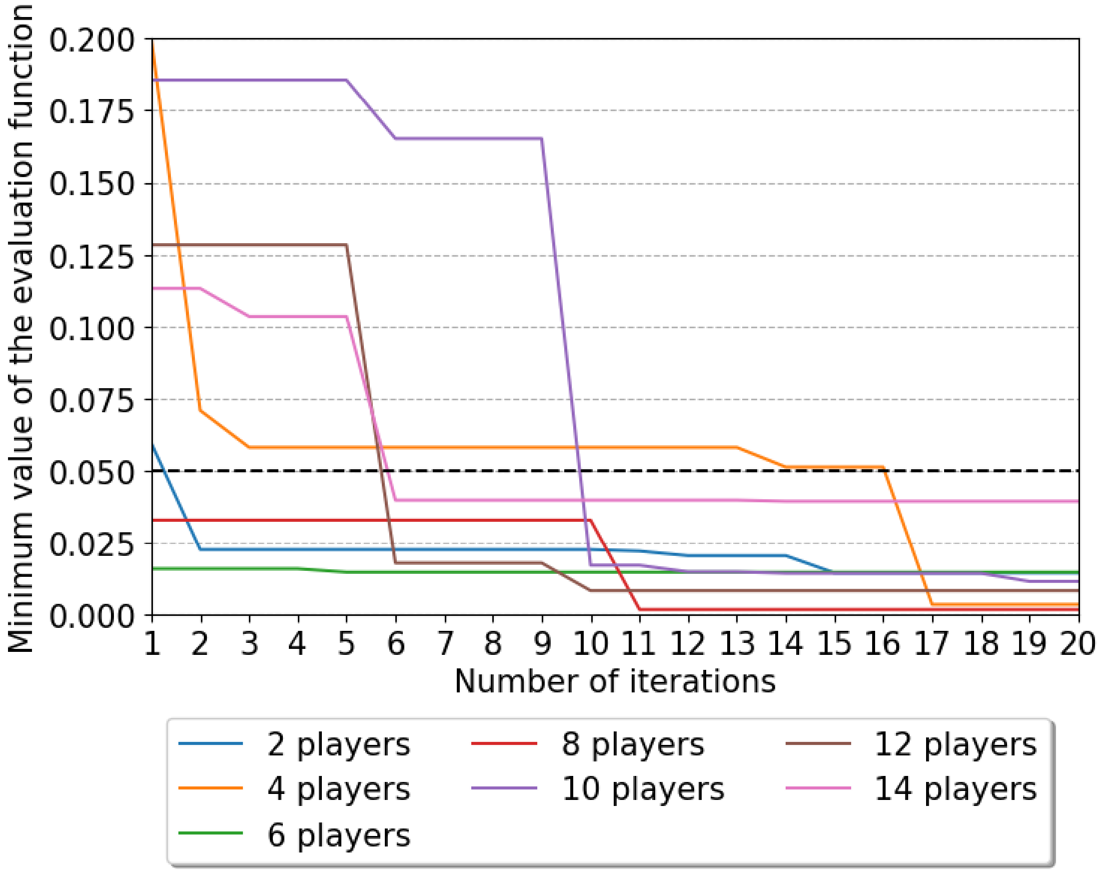

Figure 5.7: The convergence of Bayesian optimization in games

Table 5.2 shows the optimized individual violation rates ε^{ind} for a different number of players. The value of ε^{ind} decreases when the number of players increases since the correlation of microgrids' reserve provisions increases. Fig. 5.7 shows the convergence of Bayesian optimization in different games. In all experiments, Bayesian optimization can converge to the optimal individual violation rates. These values lead to the empirical joint violation rate only deviating from the predefined requirement of less than 0.05 in 20 iterations. The computation time of these games is less than 20 mins.

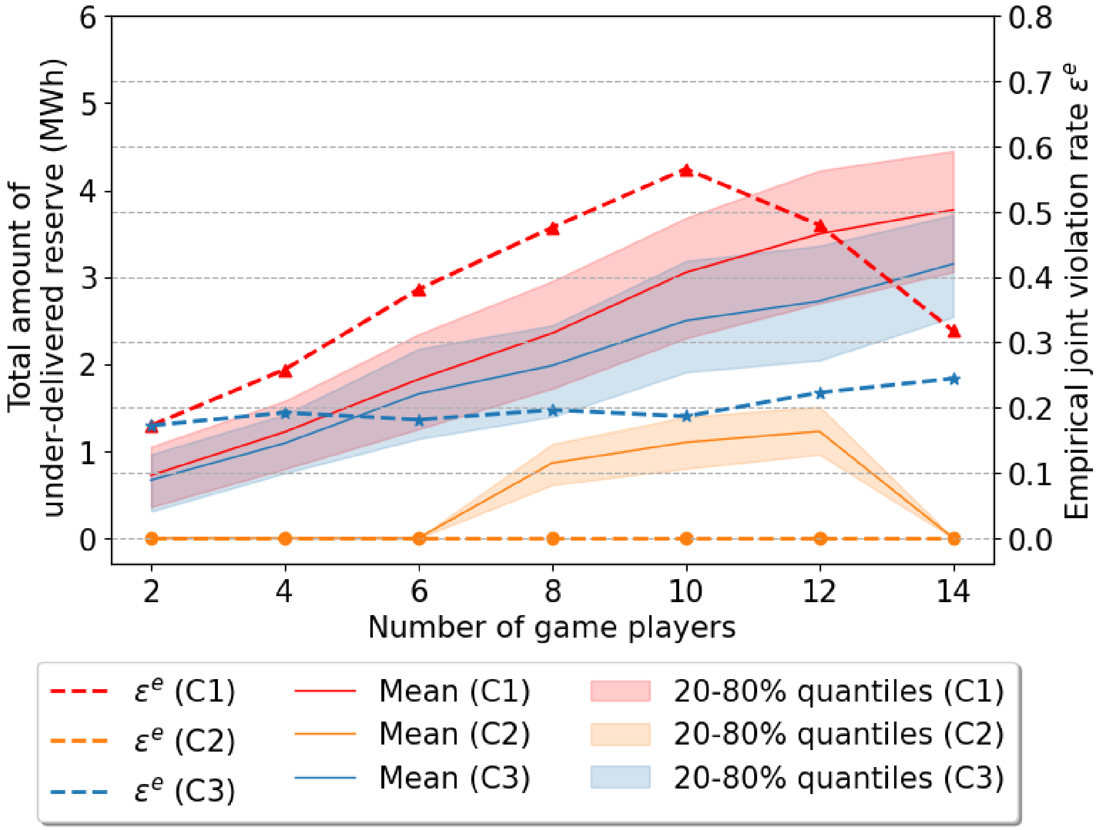

Figure 5.8: The total amount of under-delivered reserve and empirical joint violation rates in three scenarios

Fig. 5.8 shows how the under-delivered reserve and joint violation rate change in three scenarios. Only the proposed method can regulate the under-delivered reserve and secure the joint violation rate of all microgrids in a game. Using the Bayesian optimization C3 (blue), the total under-delivered reserve increases steadily, and the empirical joint violation rate is regulated at 0.2 ± 0.05. In scenario C1, the total amount of under-delivered reserve increases rapidly, and the joint violation rate exceeds 0.2 even for the 4-player game and quickly increases to 0.6 in the 10-player game. In scenario C2, on the contrary, the joint violation rate stays close to zero. Neglecting the correlation between players leads to an overly-conservative reserve regulation and almost no reserve bids.

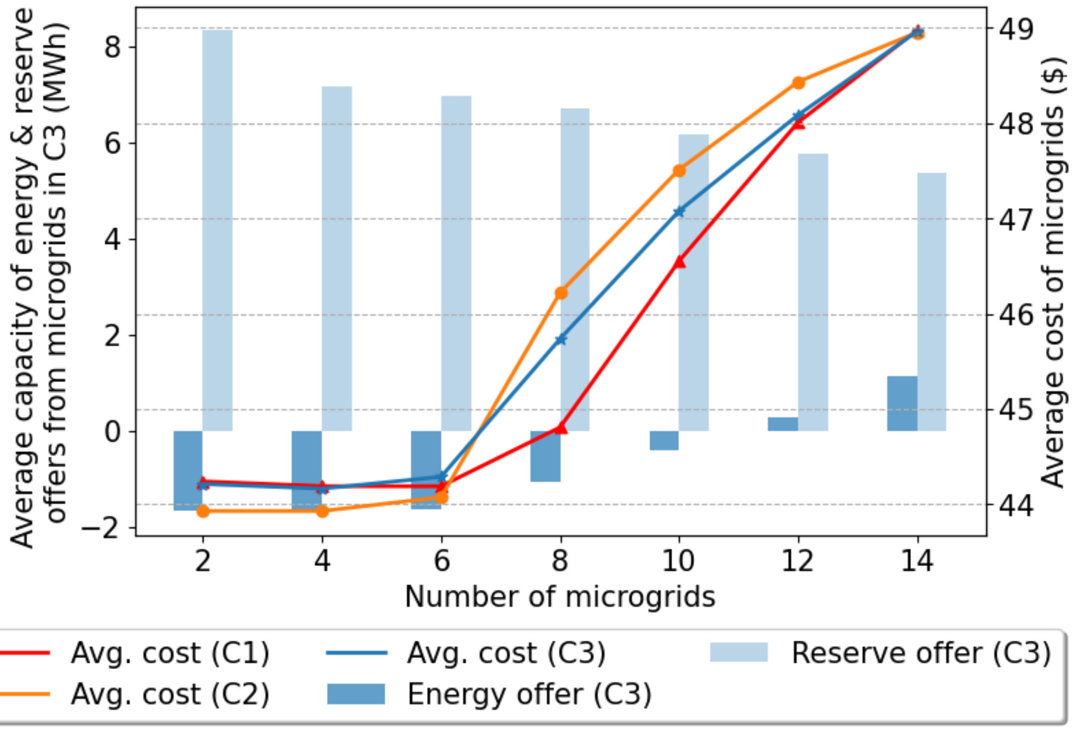

Figure 5.9: Average costs of microgrids in three scenarios and their average capacity of energy and reserve offers in C3

Fig. 5.9 summarizes the average cost of microgrids and the average energy and reserve offerings. The positive value represents that microgrids sell energy or reserve to the grid and vice versa. There is a trade-off between offering energy and reserve services.

The game starts with two microgrid players. At first, the average cost of microgrids in scenarios C1 and C3 is more than in C2. As in scenarios C1 and C3, microgrids prefer to invest in the reserve market for higher profits, even if they import power from the grid (as shown in light and dark blue bars). They gain a high load-shedding penalty in return. As the players increase, microgrids reduce their reserve offers due to market competition. The average cost of microgrids in scenarios C1 and C3 becomes lower than in C2. If the number increases to 10 or more, the reserve market becomes saturated, and microgrids need to sell energy for more profits.

Fig. 5.10 summarizes the cost of ISO and the total capacity of energy and reserve offers from microgrids in scenario C3. The total reserve capacity of microgrids increases,

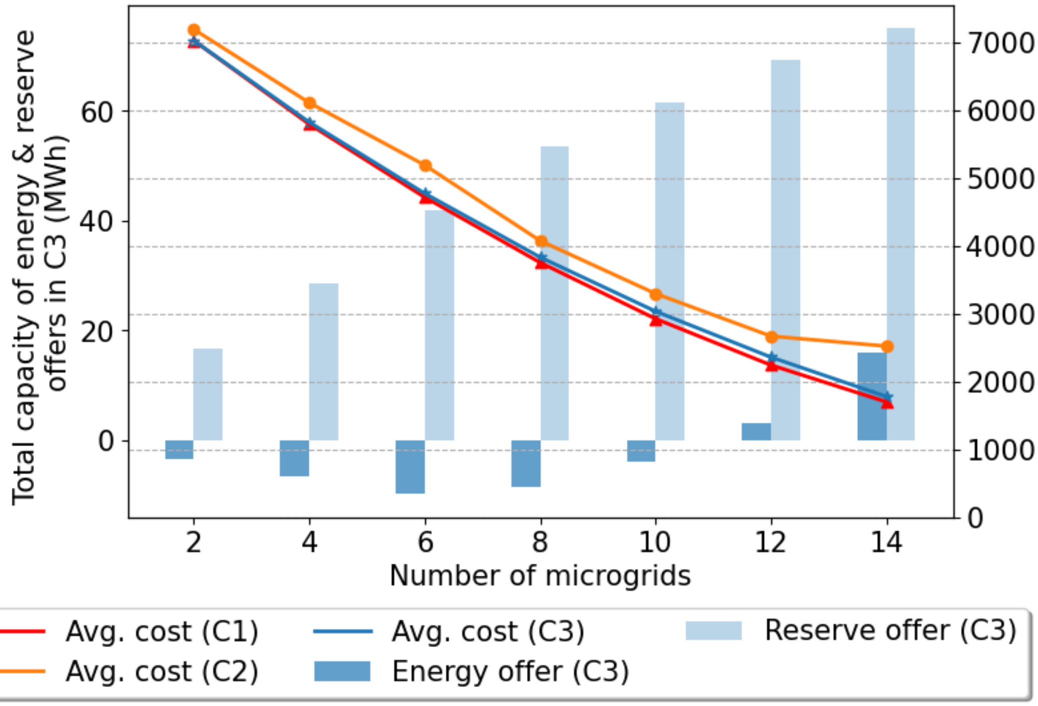

Figure 5.10: Average costs of ISO in three scenarios and the total capacity of energy and reserve offers in C3

replacing the expensive reserve services from thermal generators (as presented in the light blue bars). Therefore, the cost of ISO straightly decreases with the number of microgrid players increase in the game in all three scenarios. Three scenarios ranked by the ISO's cost are C2, C3, and then C1, which reflect the conservativeness of different system reserve regulations.

5.8 Concluding remarks

This paper designs a DRJCC game to coordinate renewable micro-grids in energy and reserve markets, which consists of risk-aware bids and system-wise reserve policy. Due to solar power uncertainty, microgrids might breach contracts and under-deliver reserve services in the real-time market. The differences result in penalties modeled as load shedding in the worst-case scenario. A two-stage DRO using the Wasserstein-metric ambiguity set models the worst-case reserve performance of each microgrid. A JCC regulates all reserve contracts as a system-wide policy. In summary, ISO and microgrids are modeled in a one-leader-multiple-follower game using MPEC.

This DRJCC game is intractable with unknown individual violation rates. Considering the unspecified correlation in the market game, the optimal individual violation rates should regulate the under-delivered reserve and ensure profits for each player. Instead of enforcing conservative assumptions as in Chapter 3, a Bayesian optimization method is proposed to approximate the optimal solutions and solve the market equilibrium iteratively. It can converge to the optimal individual violation rates in all experiments, and the joint violation rates obtained only deviate from the predefined requirement by 0.05 in 20 iterations. For comparison, simulations include two scenarios that assume players have the same performances (C1) or no correlations in their reserve provision performances (C2). Only the proposed Bayesian optimization (C3) can yield the desired outcome.

Two case studies are performed on a 30-bus network using the CAISO data. In the first case study, the optimal radius of the ambiguity set is identified as $\nu = 0.035$ for a joint violation rate of 0.2. In the second case study, the game is simulated with an increasing number of players, up to 14. The Bayesian optimization approach can regulate reserve performances effectively in all simulations and ensure the profit of microgrids in energy and reserve markets.

Although this method further reduces solution conservativeness built upon the evolutionary algorithm in Chapter 3 by identifying the correlation between CCs, improvements can still be made. For example, Bayesian optimization could be tested in a higher dimensional problem where all the violation rates of CCs are different and need to be optimized. Additionally, this JCC game framework and Bayesian optimization method are not restricted to power system operations - They could be applied in many operational and control scenarios, such as action planning and resource allocation in multi-robot systems [204].

6
Conclusions and Future Work

Contents

6.1 Concluding discussion

This thesis investigates how to apply DRJCC optimization in uncertainty-aware power system operations leveraging probabilistic energy information and approximate the solution in a novel and non-conservative way. The main findings of each chapter are summarized below.

Chapter 2 classifies different types of ML probabilistic forecasting and uncertainty-aware formulations and compares their impacts on power system decision-making. Research gaps are also identified based on their application, for example, the approximation methods to the NP-hard JCC problem in power system operations. Then, chapter 3 develops and compares two ML models for weather-informed probabilistic net load predictions. They are additive GPR and deep learning NNs. These net load probabilistic forecasting results are used for simulating dynamic reserve requirements in Chapter 5.

Chapter 4 develops a DRJCC framework to control solar-powered networked microgrids to aid grid contingencies due to extreme weather or natural disasters. This microgrid can operate in two operational modes, the grid-connected and island modes. Uncertain solar forecast error distributions are modeled using three ambiguity sets based on empirical data. The network constraints within the microgrid are formulated based on DC-OPF, and three types of network violation events are controlled jointly using DRJCC formulation. They are voltage violations, energy shortages, or surplus in the storage. The optimal allocation of individual violation rates in the JCC problem is known as the NP-hard *optimal Bonferroni Approximation*. A novel evolutionary algorithm is proposed to approximate solutions. The solution obtained from the proposed algorithm is verified by out-of-sample tests. The proposed algorithm can secure the joint violation rates at the required operational reliability requirement and reduces solution conservativeness (i.e., the system operation cost) by around 50%, compared to the benchmark case using the equally-divided violation rates, *Bonferroni Approximation*.

Chapter 5 develops a DRJCC market game to coordinate solar-powered microgrids for reliable energy and reserve services. In the two-settlement power market, renewable microgrids might breach day-ahead contracts and under-deliver the reserve. Microgrids should compensate for the differences and be penalized by the under-delivered reserve. Additionally, a severe grid contingency might occur if multiple microgrids under-deliver reserve simultaneously. A DRJCC game framework is designed to model risk-aware bidding and system-wise reserve regulations. Given the changing correlation among players, the efficacy of the *Bonferroni Approximation* diminishes, and a novel Bayesian optimization is introduced to solve this problem in a non-conservative fashion. Using the California power market data, this game framework and the Bayesian method are tested with the number of players ranging from 2 to 14. The empirical joint violation rates only deviate from the predefined requirement by 0.05 in 20 iterations.

In brief, this thesis applies the DRJCC formulation in two grid operation scenarios. Distinguished from the existing literature, innovative global optimization methods are employed to approximate their solutions in a non-conservative manner for the first time. These methods include the evolutionary algorithm and Bayesian optimization, and their efficacy is verified by out-of-sample tests using empirical forecast error samples. The results are more robust and less conservative than the existing methods and SCC formulation. These findings show that the DRJCC formulation is applicable for reliable power system operations and has greater potential in broader safe-critical control and optimization applications.

6.2 Recommendations and future work

Motivated by the results and discussions in this thesis, future work could be advanced through the following recommendations.

Reducing solution conservativeness of the high-dimension JCC problem

Solving the JCC problem is an ongoing research challenge in operational research and control engineering. The evolutionary algorithm used in Chapter 4 effectively approximates the JCC solution. However, as the dimension of the problem increases (i.e., the number of individual CCs), the *Bonferroni Approximation* can no longer stand due to the increasing correlation. The evolutionary algorithm is also not scalable to higher-dimension problems. In Chapter 5, another option to approximate the high-dimensional DRJCC problem, the Bayesian method, has been investigated. Bayesian optimization method has the advantage of enforcing no conservative assumption and theoretically can support the problem up to 100 dimensions [221]). Chapter 4, however, only experiments with a limited dimension of the problem. Research has yet to explore the scalability and solution conservativeness, as well as the design choice of the acquisition function and other hyperparameters. As the benchmark, the scenario-based approach can be implemented to compare solution conservativeness and robustness.

Leveraging probabilistic forecasting to improve the ambiguity set

This thesis uses the solar power forecast error samples to construct the ambiguity set in two case studies. This could raise an issue as the forecast error distribution is conditional to solar power generation and weather conditions (For example, there should be no overestimation of solar power if the solar panel is operated at the maximum power capacity. Although this thesis suggests in Chapter 5 by updating the forecast errors in the ambiguity set, the ambiguity set can consider the instantaneous solar power generation and weather conditions. A more effective way could be introducing the weather-informed probabilistic forecast to improve the ambiguity set.

Extending the application to the large-scale power system optimization and planning

Current power system planning models usually employ a 'representative day' method [222]. This means the representative day is selected using the clustering method to represent all the scenarios for an entire year or longer. Theoretically, the ambiguity set can capture the daily load or weather distribution more comprehensively than the representative day, thus offering a more reliable solution. In addition, the power planning models could include a higher number of scenarios, such as multiple weather years. Specifically, the model could leverage the long-term climate data, quantify the probability distribution of weather variables or renewable power outputs (e.g., air temperature and wind power), and construct the power system planning model with the appropriate uncertainty set. However, uncertainty-aware formulation, such as DRJCC, has a relatively higher computation cost than the deterministic formulation, and this could make the application to large-scale optimization problems difficult.

Appendices

A
Additional Proofs

A.1 Appendix 1: Proofs for Chapter 4

A.1.1 SOCP equivalent formulation based on the symmetric & unimodal distribution

We have a single CC based on symmetric & unimodal distribution as,

$$\inf_{\mathbb{P}\in\mathscr{D}_\zeta^2} \mathbb{P}(\tilde{A}(x)\zeta \leq \tilde{b}(x)) \geq 1-\varepsilon \tag{A.1}$$

which is equivalent to,

$$\sup_{\mathbb{P}\in\mathscr{D}_\zeta^2} \mathbb{P}(\tilde{A}(x)\zeta \leq \tilde{b}(x)) \leq \varepsilon \tag{A.2}$$

Based on *Gauss inequality* for symmetric distributions (see [223], Proposition 7.1). Given the mean M and variance σ^2 of a random variable x, the following bounds on $\mathbb{P}(x \geq a)$ are optimal,

$$\sup_{\mathbb{P}\in\mathscr{D}_\zeta^2} \mathbb{P}(x \geq a) = \begin{cases} \frac{1}{2}\min\{1, \frac{4\cdot\sigma^2}{9\cdot(M-a)^2}\} & \text{if } a < M \\ 1 & \text{otherwise} \end{cases} \tag{A.3}$$

Hence, we have the following inequalities hold,

$$\sup_{\mathbb{P}\in\mathscr{D}_\zeta^2} \mathbb{P}(x \geq 0) = \frac{1}{2}\cdot\frac{4\cdot\sigma^2}{9\cdot(M-a)^2} \leq \varepsilon \tag{A.4}$$

$$M > 0 \tag{A.5}$$

Applying (A.4) and (A.5) in our case, we have the following inequalities hold,

$$\sup_{\mathbb{P}\in\mathscr{D}_\zeta^2} \mathbb{P}(b(x) - A(x)\zeta \geq 0) = \frac{1}{2} \cdot \frac{4\cdot\Sigma\cdot A(x)^2}{9\cdot(b(x)-\mu^T A(x))^2} \leq \varepsilon \tag{A.6}$$

$$b(x) - \mu^T A(x) > 0 \tag{A.7}$$

where μ and Σ are the mean and variance of distributions $\mathbb{P}$ in the ambiguity set $\mathscr{D}_\zeta^2$ of uncertain forecast error ζ respectively. By solving (A.6) and (A.7) together, we have the SOCP formulation as,

$$\lambda(\varepsilon)||\Sigma^{1/2}A(x)||_2 \leq b_i(x) - \mu^T A(x) \tag{A.8}$$

where

$$\lambda(\varepsilon) := (\frac{2}{9\varepsilon})^{\frac{1}{2}} \quad \forall\varepsilon \in (0, \frac{1}{6}) \tag{A.9}$$

A.1.2 SOCP formulation of voltage-limit CCs

We here derive the SOCP formulation of voltage-limit CCs in detail. As (4.1p) is a symmetric two-side constraint, we provide proof of one side for conciseness.

Re-arranging (4.1n) and (4.1p)

$$\mathbf{v} = \frac{\mathbf{P}_t^{inj} + \mathbf{diag}(\mathbf{I}_o)\mathbf{V}_o}{\mathbf{diag}(\mathbf{V}_o)\mathbf{Y} + \mathbf{diag}(\mathbf{I}_o)} \leq \overline{\mathbf{v}} \tag{A.10}$$

$$\mathbf{P}_t^{inj} = \mathbf{Q} + \mathbf{H}\mathbf{P}_t^{\tilde{pv}} \tag{A.11}$$

where

$$\mathbf{Q} := \mathbf{C}^g\mathbf{P}_t^g + \mathbf{C}^s(\mathbf{P}_t^d - \mathbf{P}_t^c) + \mathbf{C}^{pv}\mathbf{P}_t^{pv'} - \mathbf{C}_t^l\mathbf{P}_t^{l'} \tag{A.12}$$

$$\mathbf{H} := -\mathbf{C}^s\mathbf{d}_t^g + \mathbf{C}^{pv} \tag{A.13}$$

Constraint (A.10) can be re-written as,

$$\mathbf{G}^{-1}\{\mathbf{Q}+\mathbf{H}\mathbf{P}_t^{\tilde{p}v}+\mathbf{diag}(\mathbf{I}_o)\mathbf{V}_o\} \leq \overline{\mathbf{v}} \tag{A.14}$$

$$\mathbf{G} := \mathbf{diag}(\mathbf{V}_o)\mathbf{Y}+\mathbf{diag}(\mathbf{I}_o) \tag{A.15}$$

Re-arranging (A.14)

$$\mathbf{G}^{-1}\mathbf{H}\mathbf{P}_t^{\tilde{p}v} \leq \overline{\mathbf{v}}-\mathbf{G}^{-1}\{\mathbf{diag}(\mathbf{I}_o)\mathbf{V}_o+\mathbf{Q}\} \tag{A.16}$$

Based on (4.13) and (A.11), we can obtain.

$$\lambda(\varepsilon)||\mathbf{G}^{-1}\mathbf{H}\Sigma_{pv,t}^{1/2}||_2 \leq \overline{\mathbf{v}}-\mathbf{G}^{-1}\{\mathbf{diag}(\mathbf{I}_o)\mathbf{V}_o+\mathbf{Q}+\mu_{pv,t}^T\mathbf{1}\mathbf{H}\} \tag{A.17}$$

A.2 Appendix 2: Proofs for Chapter 5

MPEC models, also known as Complementarity models, can represent the simultaneous optimization problems of one or several interacting decision-makers [210]. In terms of solving techniques, it uses Karush-Kuhn-Tucker (KKT) condition to reformulate one optimization problem into several constraints at the optimal point. However, it involves non-linear constraints requiring techniques such as the Big-M method. This section details methodologies.

A.2.1 KKT conditions for both linear and non-linear optimization

Given a general problem which can be a linear or non-linear optimization,

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) \tag{A.18a}$$

$$\text{subject to} \quad \mathbf{h}_i(\mathbf{x}) = 0, \quad (\mathbf{u}_i) \quad i = 1, ..., m \tag{A.18b}$$

$$\mathbf{g}_j(\mathbf{x}) \leq 0, \quad (\mathbf{v}_j) \quad j = 1, ..., r \tag{A.18c}$$

where $\mathbf{u}_i$ and $\mathbf{v}_j$ are dual variables for constraints (A.18b) and (A.18c) respectively. The KKT conditions include the stationarity, complementary slackness, primal, and dual feasibility conditions as below.

$$\partial f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{u}_i \partial \mathbf{h}_i(\mathbf{x}) + \sum_{j=1}^{r} \mathbf{v}_j \partial \mathbf{g}_j(\mathbf{x}) = 0 \quad \text{(stationarity)} \tag{A.19a}$$

$$\mathbf{v}_j \partial \mathbf{g}_j(\mathbf{x}) = 0 \quad \text{for all } i \quad \text{(complementary slackness)} \tag{A.19b}$$

$$\mathbf{h}_i(\mathbf{x}) = 0, \mathbf{g}_j(\mathbf{x}) = 0 \quad \text{for all } i, j \quad \text{(primal feasibility)} \tag{A.19c}$$

$$\mathbf{v}_j \geq 0 \quad \text{for all } i \quad \text{(dual feasibility)} \tag{A.19d}$$

An optimization model can be integrated into another as a constraint under the KKT conditions. However, constraint (A.19b) is non-linear, requiring additional reformulation. One of the feasible approaches, the Big-M method, is introduced in the later section.

A.2.2 The linearization of the objective function

The objective function (5.3) of the complementarity model in Chapter 5 is non-linear because it has a multiplication of two variables (i.e., $\pi_i^r R_{t,i}^{up} - \pi_i^e P_{t,i}^{ex}$). We linearize this objective function based on the strong duality theory. The objective function of the dual problem of the lower level problem M2 (5.18) is given by,

$$\max \mathscr{J}^{LL}(\mathbf{x}') = \mu_{i,t}^{pb} P_{i,t}^{nl} - \mu_{i,t}^{uppl} \overline{P^s} - \mu_{i,t}^{dnpl} \underline{P^s} \tag{A.20}$$

Based on the strong duality theory [224], the objective function of dual and primal problems are equal at the optimal point.

$$\max \mathscr{J}^{LL}(\mathbf{x}') = \min \mathscr{J}^{LL}(\mathbf{x}) \tag{A.21}$$

This is equivalent to,

$$\begin{aligned} \mu_{i,t}^{pb} P_{i,t}^{nl} - \mu_{i,t}^{uppl} \overline{P^s} - \mu_{i,t}^{dnpl} \underline{P^s} \\ = (m_i^s (\Delta t P_{i,t}^s)^2 + \Delta t m_i^{mgr} R_{i,t}^{up} + \Delta t \pi_i^e P_{i,t}^{ex} - \Delta t \pi_i^r R_{i,t}^{up}) \\ + (\nu l_i^l + \frac{1^\top \mathbf{W}_i^l}{N_S}) \end{aligned} \tag{A.22}$$

Then the non-linear term in the objective function can be written as below,

$$\begin{aligned} \sum_{i \in \mathscr{I}} (\Delta t \pi_i^r R_{i,t}^{up} - \Delta t \pi_i^e P_{i,t}^{ex}) \\ = \sum_{i \in \mathscr{I}} (m_i^s (\Delta t P_{i,t}^s)^2 + \Delta t m_i^{mgr} R_{i,t}^{up} + (\nu l_i^l + \frac{1^\top \mathbf{W}_i^l}{N_S}) \\ - \mu_{i,t}^{pb} P_{i,t}^{nl} + \mu_{i,t}^{uppl} \overline{P^s} + \mu_{i,t}^{dnpl} \underline{P^s}) \end{aligned} \tag{A.23}$$

A.2.3 Big-M transformation

The complementarity slackness under the KKT condition results in a set of non-linear constraints in the optimization problem. In Chapter 5, the corresponding non-linear constraints are constraints (5.33) - (5.39). This thesis employs the big-M method to solve these constraints. This section shows the detailed procedures for one constraint (5.33), while the other constraints use the same approach.

The original constraint (5.33) is given below,

$$0 \leq (\mathbf{W} - \Delta t\mathbf{s}(\mathbf{R}^{up})^\top V + \eta^l) \perp \mu^{dro1} \geq 0 \tag{A.24}$$

This is equivalent to the following constraints.

$$0 \leq (\mathbf{W} - \Delta t\mathbf{s}(\mathbf{R}^{up})^\top V + \eta^l) \tag{A.25}$$

$$0 \leq \mu^{dro1} \tag{A.26}$$

$$(\mathbf{W} - \Delta t\mathbf{s}(\mathbf{R}^{up})^\top V + \eta^l)\mu^{dro1} = 0 \tag{A.27}$$

Notably, constraint (A.27) is non-linear, and I introduce the Big-M approach to reformulate it into two linear constraints.

$$(\mathbf{W} - \Delta t\mathbf{s}(\mathbf{R}^{up})^\top V + \eta^l) \leq M\mathbf{U}_1 \tag{A.28}$$

$$\mu^{dro1} \leq M(1 - \mathbf{U}_1) \tag{A.29}$$

Where M is a sufficiently large positive number (i.e., significantly larger than $\mathbf{W} - \Delta t\mathbf{s}(\mathbf{R}^{up})^\top V + \eta^l$ and μ^{dro1}. $\mathbf{U}_1 \in \mathbb{R}^{N_s \times T}$ is a binary variable matrix whose dimension is the same as μ^{dro1}. This approach is then applied to the other constraints (5.34)- (5.39).

B
Case Study Inputs and Experiment Data

B.1 Appendix 1: Experiment Data in Chapter 3

B.1.1 CPRS for the optimal kernel selection

The CPRS for eight combinations in experiments for two locations are listed in Table B.1 and B.2. The best model with its lowest score is highlighted in **bold**.

Table B.1: The kernel combinations for weather features and their CPRS (YMCA)

Features	Temp.	Humidity	Solar	CPRS
Kernels	Linear	Linear	Linear	0.767
	Linear	SE	Linear	0.758
	Linear	Linear	SE	0.774
	Linear	SE	SE	0.773
	SE	Linear	Linear	0.720
	SE	**SE**	**Linear**	**0.716**
	SE	Linear	SE	0.731
	SE	SE	SE	0.731

B.1.2 Probabilistic evaluation scores of all experiments in three cases

Evaluation scores of all experiments of three cases in Chapter 3 are presented in Table B.3.

Table B.2: The kernel combinations for weather features and their CPRS (EC)

Features	Temp.	Humidity	Solar	CPRS
Kernels	Linear	Linear	Linear	0.961
	Linear	SE	Linear	0.919
	Linear	Linear	SE	0.924
	Linear	SE	SE	0.912
	SE	Linear	Linear	0.933
	SE	SE	Linear	0.915
	SE	Linear	SE	0.924
	SE	**SE**	**SE**	**0.909**

Table B.3: Probabilistic evaluation scores for experiments in three cases

	Models	$\mathbf{F}_t$			$\mathbf{F}_t + \mathbf{F}_w$ (Case 2)			$\mathbf{F}_t + \mathbf{F}_w$ (Case 3)		
		nCPRS	nWS	nMAE	nCPRS	nWS	nMAE	nCPRS	nWS	nMAE
EC	GPR	0.08	0.95	0.1	0.06	0.66	0.08	0.08	0.97	0.09
	Q-NN	–	0.38	0.09	–	0.41	0.09	–	0.41	0.09
	E-NN	0.06	0.37	0.09	0.06	0.37	0.09	0.06	0.37	0.09
YMCA	GPR	0.06	0.62	0.08	0.05	0.46	0.07	0.07	0.88	0.09
	Q-NN	–	0.6	0.12	-	0.52	0.09	-	0.52	0.09
	E-NN	0.06	0.56	0.09	0.06	0.57	0.08	0.06	0.57	0.08

B.2 Appendix 2: Case study data in Chapter 4

The configuration of the solar-powered network microgrids is designed by the Robust Extra-Low Cost Solar Nanogrid (RELCON) project. The proposed optimization and control scheme is not location-specific and could be replicated using the same types of input data as listed.

1. **Solar irradiance and air temperature data**: The weather dataset is two-year 15-min measurements in Gitaru dam, Kenya (Lat: -0.789, Lon: 37.73, Elevation: 969 m a.s.l.) by SolarGIS. The datasets include the global and diffuse horizontal irradiance, PV output, air temperature at 2m, and wind speed at 10 m.

2. **Appliance-level load profiles**: Given the fixed amount of daily electricity consumption, the household chooses how to consume the energy on different appliances according to the utility u_t^a. This decision-making process can be formulated as an optimization model. The objective function (B.1) describes how each household maximizes their total utility.

Each appliance has a set of constraints for time use window, energy, and power capacity.

$$U = \sum_{a=0}^{A} \sum_{t=0}^{T} u_t^a \tag{B.1}$$

Three basic electrical appliances are considered, including lights, fans, and phone chargers. Their power, energy constraints, and utility functions are as follows.

- Lights: The household can have three light bulbs with a 3W power rate. It could be switched on during $T = [4:00, 24:00]$. The utility is related to the solar irradiance and formulated as $u_t^{light} = -2\frac{(G_{stc}-G_t)}{G_{stc}} p_t^{light}(p_t^{light} - 400)$.
- Fans: The fan has a 5W power rating. It can be turned on when the ambient temperature exceeds the threshold $T_{th} = 23^o$. The utility is related to the ambient temperature and formulated as $u_t^{fan} = -\frac{T_t}{T_{th}} p_t^{fan}(p_t^{fan} - 400)$.
- Phone charger: The phone can be charged anytime. The utility is formulated as $u_t^{phone} = 0.5 p_t^{phone}(p_t^{phone} - 400)$

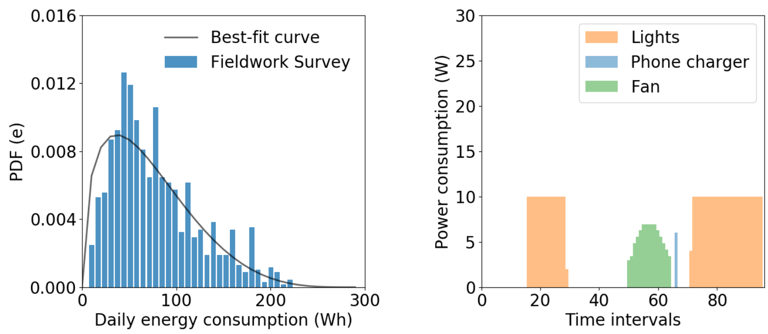

Figure B.1: (a) Electricity consumption distribution [225] (b)Appliances-level load profile

Using the data of 51 households using off-grid energy systems [225], their daily electricity consumption fits the beta distribution as shown in Fig. B.1 (a), which can represent the situation in a typical energy-poor community. The daily consumption values are sampled from a predefined beta distribution and used to construct the load profiles. The appliance-level load profile for one or multiple households are presented in Fig. B.1 (b)) and Fig. B.2 respectively.

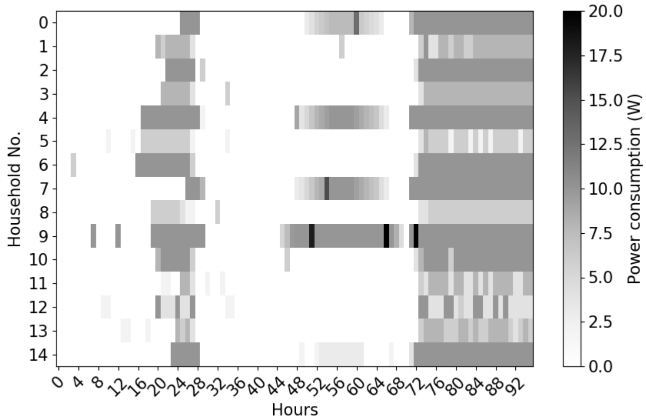

Figure B.2: The 15-min load profiles of 15 households

Bibliography

[1] IPCC, Global Warming of 1.5°C: IPCC Special Report on impacts of global warming of 1.5°C above pre-industrial levels in context of strengthening response to climate change, sustainable development, and efforts to eradicate poverty, 1st Edition, Cambridge University Press, 2022. doi:10.1017/9781009157940.
URL https://www.cambridge.org/core/product/identifier/9781009157940/type/book

[2] Energy and Climate Intelligence Unit, Oxford Net Zero, Net Zero Tracker.
URL https://zerotracker.net/

[3] Y. Wang, C. Chen, J. Wang, R. Baldick, Research on Resilience of Power Systems Under Natural Disasters—A Review, IEEE Transactions on Power Systems 31 (2) (2016) 1604–1613. doi:10.1109/TPWRS.2015.2429656.
URL http://ieeexplore.ieee.org/document/7105972/

[4] Zhongda Chu, Guoxuan Cui, Fei Teng, Simultaneous Provision of Virtual Inertia and Damping from Inverter-Based Resources for Optimal Microgrid Frequency Regulation.
URL https://arxiv.org/pdf/2212.14250.pdf

[5] A. Tuohy, Enhancing energy system reliability and resiliency in a net-zero economy (Nov. 2020).
URL http://mydocs.epri.com/docs/public/EPRI-Report-EnhancingEnergySystemReliability-20210804.pdf

[6] California ISO, OASIS - OASIS Prod - PUBLIC - 0.
URL http://oasis.caiso.com/mrioasis/logon.do

[7] F. Fahiman, S. Disano, S. M. Erfani, P. Mancarella, C. Leckie, Data-Driven Dynamic Probabilistic Reserve Sizing Based on Dynamic Bayesian Belief Networks, IEEE Transactions on Power Systems 34 (3) (2019) 2281–2291. doi:10.1109/TPWRS.2018.2884711.
URL https://ieeexplore.ieee.org/document/8556483/

[8] J. Fox, E. Ela, B. Hobbs, J. Sharp, J. Novacheck, A. Motley, R. J. Bessa, P. Pinson, G. Kariniotakis, Forecasting and Market Design Advances: Supporting an Increasing Share of Renewable Energy, IEEE Power and Energy Magazine 19 (6) (2021) 77–85. doi:10.1109/MPE.2021.3104132.
URL https://ieeexplore.ieee.org/document/9586385/

[9] B. Li, C. Feng, C. Siebenschuh, R. Zhang, E. Spyrou, V. Krishnan, B. F. Hobbs, J. Zhang, Sizing ramping reserve using probabilistic solar forecasts: A data-driven method, Applied Energy 313 (2022) 118812. doi:10.1016/j.apenergy.2022.118812.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261922002574

[10] S. E. Haupt, M. Garcia Casado, M. Davidson, J. Dobschinski, P. Du, M. Lange, T. Miller, C. Mohrlen, A. Motley, R. Pestana, J. Zack, The Use of Probabilistic Forecasts: Applying Them in Theory and Practice, IEEE Power and Energy Magazine 17 (6) (2019) 46–57. doi:10.1109/MPE.2019.2932639.
URL https://ieeexplore.ieee.org/document/8878053/

[11] A. Hirsch, Y. Parag, J. Guerrero, Microgrids: A review of technologies, key drivers, and outstanding issues, Renewable and Sustainable Energy Reviews 90 (2018) 402–411. doi:10.1016/j.rser.2018.03.040.
URL https://linkinghub.elsevier.com/retrieve/pii/S136403211830128X

[12] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, M. Zugno, Integrating Renewables in Electricity Markets: Operational Problems, 1st Edition, no. 205 in International Series in Operations Research & Management Science, Springer US : Imprint: Springer, New York, NY, 2014. doi:10.1007/978-1-4614-9411-9.

[13] Y. Parag, B. K. Sovacool, Electricity market design for the prosumer era, Nature Energy 1 (4) (2016) 16032. doi:10.1038/nenergy.2016.32.
URL http://www.nature.com/articles/nenergy201632

[14] J. T. Lee, D. S. Callaway, The cost of reliability in decentralized solar power systems in sub-Saharan Africa, Nature Energy 3 (11) (2018) 960–968. doi:10.1038/s41560-018-0240-y.
URL http://www.nature.com/articles/s41560-018-0240-y

[15] S. Cephas, High-level control algorithms for maximising energy efficiency of solar nano-grids designed for energy access, Ph.D. thesis, Univerisity of Oxford (2020).
URL https://ora.ox.ac.uk/objects/uuid:c16f98b5-80fd-4e92-82ed-f88fc7571029

[16] C. Chen, J. Wang, F. Qiu, D. Zhao, Resilient Distribution System by Microgrids Formation After Natural Disasters, IEEE Transactions on Smart Grid 7 (2) (2016) 958–966. doi:10.1109/TSG.2015.2429653.
URL http://ieeexplore.ieee.org/document/7127029/

[17] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi, C. A. Canizares, R. Iravani, M. Kazerani, A. H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke, G. A. Jimenez-Estevez, N. D. Hatziargyriou, Trends in Microgrid Control, IEEE Transactions on Smart Grid 5 (4) (2014) 1905–1919. doi:10.1109/TSG.2013.2295514.
URL http://ieeexplore.ieee.org/document/6818494/

[18] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, K. P. Wong, Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine, IEEE Transactions on Power Systems 29 (3) (2014) 1033–1044. doi:10.1109/TPWRS.2013.2287871.
URL http://ieeexplore.ieee.org/document/6665108/

[19] K. Li, R. Wang, H. Lei, T. Zhang, Y. Liu, X. Zheng, Interval prediction of solar power using an Improved Bootstrap method, Solar Energy 159 (2018) 97–112. doi:10.1016/j.solener.2017.10.051.
URL https://linkinghub.elsevier.com/retrieve/pii/S0038092X17309313

[20] D. Yang, J.-e. Guo, S. Sun, J. Han, S. Wang, An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting, Applied Energy 306 (2022) 117992. doi:10.1016/j.apenergy.2021.117992.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921012952

[21] D. Saez, F. Avila, D. Olivares, C. Canizares, L. Marin, Fuzzy Prediction Interval Models for Forecasting Renewable Resources and Loads in Microgrids, IEEE Transactions on Smart Grid 6 (2) (2015) 548–556. doi:10.1109/TSG.2014.2377178.
URL http://ieeexplore.ieee.org/document/6994295/

[22] Yong-Ming Zhang, Wei-Gui Qi, Interval forecasting for heating load using support vector regression and error correcting Markov chains, in: 2009 International Conference on Machine Learning and Cybernetics, IEEE, Baoding, China, 2009, pp. 1106–1110. doi:10.1109/ICMLC.2009.5212405.
URL http://ieeexplore.ieee.org/document/5212405/

[23] Qun Zhou, L. Tesfatsion, Chen-Ching Liu, Short-Term Congestion Forecasting in Wholesale Power Markets, IEEE Transactions on Power Systems 26 (4) (2011) 2185–2196. doi:10.1109/TPWRS.2011.2123118.
URL http://ieeexplore.ieee.org/document/5741753/

[24] R. Yuan, B. Wang, Z. Mao, J. Watada, Multi-objective wind power scenario forecasting based on PG-GAN, Energy 226 (2021) 120379. doi:10.1016/j.energy.2021.120379.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544221006289

[25] M. Sun, C. Feng, J. Zhang, Probabilistic solar power forecasting based on weather scenario generation, Applied Energy 266 (2020) 114823. doi:10.1016/j.apenergy.2020.114823.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920303354

[26] J. Taylor, R. Buizza, Neural network load forecasting with weather ensemble predictions, IEEE Transactions on Power Systems 17 (3) (2002) 626–632. doi:10.1109/TPWRS.2002.800906.
URL http://ieeexplore.ieee.org/document/1033703/

[27] Z. Xia, H. Ma, T. K. Saha, R. Zhang, Consumption Scenario-Based Probabilistic Load Forecasting of Single Household, IEEE Transactions on Smart Grid 13 (2) (2022) 1075–1087. doi:10.1109/TSG.2021.3132039.
URL https://ieeexplore.ieee.org/document/9632812/

[28] J. Yan, J. Zhang, Y. Liu, G. Lv, S. Han, I. E. G. Alfonzo, EV charging load simulation and forecasting considering traffic jam and weather to support the integration of renewables and EVs, Renewable Energy 159 (2020) 623–641. doi:10.1016/j.renene.2020.03.175.
URL https://linkinghub.elsevier.com/retrieve/pii/S0960148120305085

[29] M. Dahl, A. Brun, G. B. Andresen, Using ensemble weather predictions in district heating operation and load forecasting, Applied Energy 193 (2017) 455–465. doi:10.1016/j.apenergy.2017.02.066.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261917302143

[30] X. Lu, J. Qiu, G. Lei, J. Zhu, Scenarios modelling for forecasting day-ahead electricity prices: Case studies in Australia, Applied Energy 308 (2022) 118296. doi:10.1016/j.apenergy.2021.118296.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921015555

[31] A. U. Haque, M. H. Nehrir, P. Mandal, A Hybrid Intelligent Model for Deterministic and Quantile Regression Approach for Probabilistic Wind Power Forecasting, IEEE Transactions on Power Systems 29 (4) (2014) 1663–1672. doi:10.1109/TPWRS.2014.2299801.
URL http://ieeexplore.ieee.org/document/6727578/

[32] X. G. Agoua, R. Girard, G. Kariniotakis, Probabilistic Models for Spatio-Temporal Photovoltaic Power Forecasting, IEEE Transactions on Sustainable Energy 10 (2) (2019) 780–789. doi:10.1109/TSTE.2018.2847558.
URL https://ieeexplore.ieee.org/document/8385146/

[33] A. Mashlakov, T. Kuronen, L. Lensu, A. Kaarna, S. Honkapuro, Assessing the performance of deep learning models for multivariate probabilistic energy forecasting, Applied Energy 285 (2021) 116405. doi:10.1016/j.apenergy.2020.116405.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920317748

[34] S. Ben Taieb, R. Huser, R. J. Hyndman, M. G. Genton, Forecasting Uncertainty in Electricity Smart Meter Data by Boosting Additive Quantile Regression, IEEE Transactions on Smart Grid 7 (5) (2016) 2448–2455. doi:10.1109/TSG.2016.2527820.
URL http://ieeexplore.ieee.org/document/7423794/

[35] T. Hu, K. Liu, H. Ma, Probabilistic Electric Vehicle Charging Demand Forecast Based on Deep Learning and Machine Theory of Mind, in: 2021 IEEE Transportation Electrification Conference & Expo (ITEC), IEEE, Chicago, IL, USA, 2021, pp. 795–799. doi:10.1109/ITEC51675.2021.9490147.
URL https://ieeexplore.ieee.org/document/9490147/

[36] R. Niemierko, J. Töppel, T. Tränkler, A D-vine copula quantile regression approach for the prediction of residential heating energy consumption based on historical data, Applied Energy 233-234 (2019) 691–708. doi:10.1016/j.apenergy.2018.10.025.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261918315812

[37] A. Bello, D. W. Bunn, J. Reneses, A. Munoz, Medium-Term Probabilistic Forecasting of Electricity Prices: A Hybrid Approach, IEEE Transactions on Power Systems 32 (1) (2017) 334–343. doi:10.1109/TPWRS.2016.2552983.
URL http://ieeexplore.ieee.org/document/7450641/

[38] A. Michiorri, H.-M. Nguyen, S. Alessandrini, J. B. Bremnes, S. Dierer, E. Ferrero, B.-E. Nygaard, P. Pinson, N. Thomaidis, S. Uski, Forecasting for dynamic line rating, Renewable and Sustainable Energy Reviews 52 (2015) 1713–1730. doi:10.1016/j.rser.2015.07.134.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032115007819

[39] T. Zufferey, S. Renggli, G. Hug, Probabilistic State Forecasting and Optimal Voltage Control in Distribution Grids under Uncertainty, Electric Power Systems Research 188 (2020) 106562. doi:10.1016/j.epsr.2020.106562.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378779620303667

[40] H. Zhang, Y. Liu, J. Yan, S. Han, L. Li, Q. Long, Improved Deep Mixture Density Network for Regional Wind Power Probabilistic Forecasting, IEEE Transactions on Power Systems 35 (4) (2020) 2549–2560. doi:10.1109/TPWRS.2020.2971607.
URL https://ieeexplore.ieee.org/document/8982039/

[41] K. Doubleday, S. Jascourt, W. Kleiber, B.-M. Hodge, Probabilistic Solar Power Forecasting Using Bayesian Model Averaging, IEEE Transactions on Sustainable Energy 12 (1) (2021) 325–337. doi:10.1109/TSTE.2020.2993524.
URL https://ieeexplore.ieee.org/document/9091068/

[42] C. Feng, M. Sun, J. Zhang, Reinforced Deterministic and Probabilistic Load Forecasting via Q -Learning Dynamic Model Selection, IEEE Transactions on Smart Grid 11 (2) (2020) 1377–1386. doi:10.1109/TSG.2019.2937338.
URL https://ieeexplore.ieee.org/document/8813103/

[43] M. Sun, T. Zhang, Y. Wang, G. Strbac, C. Kang, Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting, IEEE Transactions on Power Systems 35 (1) (2020) 188–201. doi:10.1109/TPWRS.2019.2924294.
URL https://ieeexplore.ieee.org/document/8743433/

[44] M. S. Islam, N. Mithulananthan, D. Q. Hung, A Day-Ahead Forecasting Model for Probabilistic EV Charging Loads at Business Premises, IEEE Transactions on Sustainable Energy 9 (2) (2018) 741–753. doi:10.1109/TSTE.2017.2759781.
URL http://ieeexplore.ieee.org/document/8061012/

[45] J. Lange, M. Kaltschmitt, Probabilistic day-ahead forecast of available thermal storage capacities in residential households, Applied Energy 306 (2022) 117957. doi:10.1016/j.apenergy.2021.117957.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921012642

[46] E. Heylen, J. Browell, F. Teng, Probabilistic Day-Ahead Inertia Forecasting, IEEE Transactions on Power Systems 37 (5) (2022) 3738–3746. doi:10.1109/TPWRS.2021.3134811.
URL https://ieeexplore.ieee.org/document/9648007/

[47] Y. Ji, R. J. Thomas, L. Tong, Probabilistic Forecasting of Real-Time LMP and Network Congestion, IEEE Transactions on Power Systems (2016) 1–1doi:10.1109/TPWRS.2016.2592380.
URL http://ieeexplore.ieee.org/document/7514918/

[48] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, Probabilistic forecasting of dynamic line rating for over-head transmission lines, in: 2015 IEEE Power & Energy Society General Meeting, IEEE, Denver, CO, USA, 2015, pp. 1–5. doi:10.1109/PESGM.2015.7286604.
URL http://ieeexplore.ieee.org/document/7286604/

[49] B. Li, J. Zhang, A review on the integration of probabilistic solar forecasting in power systems, Solar Energy 210 (2020) 68–86. doi:10.1016/j.solener.2020.07.066.
URL https://linkinghub.elsevier.com/retrieve/pii/S0038092X20307982

[50] J. Nowotarski, R. Weron, Recent advances in electricity price forecasting: A review of probabilistic forecasting, Renewable and Sustainable Energy Reviews 81 (2018) 1548–1568. doi:10.1016/j.rser.2017.05.234.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032117308808

[51] R. Tawn, J. Browell, A review of very short-term wind and solar power forecasting, Renewable and Sustainable Energy Reviews 153 (2022) 111758. doi:10.1016/j.rser.2021.111758.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032121010285

[52] S. Yu, F. Fang, Y. Liu, J. Liu, Uncertainties of virtual power plant: Problems and countermeasures, Applied Energy 239 (2019) 454–470. doi:10.1016/j.apenergy.2019.01.224.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261919302442

[53] A. Zakaria, F. B. Ismail, M. H. Lipu, M. Hannan, Uncertainty models for stochastic optimization in renewable energy applications, Renewable Energy 145 (2020) 1543–1571. doi:10.1016/j.renene.2019.07.081.
URL https://linkinghub.elsevier.com/retrieve/pii/S0960148119311012

[54] P. McSharry, S. Bouwman, G. Bloemhof, Probabilistic Forecasts of the Magnitude and Timing of Peak Electricity Demand, IEEE Transactions on Power Systems 20 (2) (2005) 1166–1172. doi:10.1109/TPWRS.2005.846071.
URL http://ieeexplore.ieee.org/document/1425617/

[55] J. W. Taylor, Density forecasting for the efficient balancing of the generation and consumption of electricity, International Journal of Forecasting 22 (4) (2006) 707–724. doi:10.1016/j.ijforecast.2006.02.001.
URL https://linkinghub.elsevier.com/retrieve/pii/S0169207006000227

[56] J. Browell, M. Fasiolo, Probabilistic Forecasting of Regional Net-Load With Conditional Extremes and Gridded NWP, IEEE Transactions on Smart Grid 12 (6) (2021) 5011–5019. doi:10.1109/TSG.2021.3107159.
URL https://ieeexplore.ieee.org/document/9520817/

[57] D. S. G. Team, Data Study Group Final Report: TelenorPublisher: Zenodo (Jun. 2020). doi:10.5281/ZENODO.3786852.
URL https://zenodo.org/record/3786852

[58] I. Ghalehkhondabi, E. Ardjmand, G. R. Weckman, W. A. Young, An overview of energy demand forecasting methods published in 2005–2015, Energy Systems 8 (2) (2017) 411–447. doi:10.1007/s12667-016-0203-y.
URL http://link.springer.com/10.1007/s12667-016-0203-y

[59] T.-Y. Kim, S.-B. Cho, Predicting residential energy consumption using CNN-LSTM neural networks, Energy 182 (2019) 72–81. doi:10.1016/j.energy.2019.05.230.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544219311223

[60] F. vom Scheidt, X. Dong, A. Bartos, P. Staudt, C. Weinhardt, Probabilistic Forecasting of Household Loads: Effects of Distributed Energy Technologies on Forecast Quality, in: Proceedings of the Twelfth ACM International Conference on Future Energy Systems, ACM, Virtual Event Italy, 2021, pp. 231–238. doi:10.1145/3447555.3464861.
URL https://dl.acm.org/doi/10.1145/3447555.3464861

[61] Z. Wang, T. Hong, M. A. Piette, Building thermal load prediction through shallow machine learning and deep learning, Applied Energy 263 (2020) 114683. doi:10.1016/j.apenergy.2020.114683.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920301951

[62] O. Sprangers, S. Schelter, M. de Rijke, Parameter-efficient deep probabilistic forecasting, International Journal of Forecasting (2022) S0169207021001850doi:10.1016/j.ijforecast.2021.11.011.

URL https://linkinghub.elsevier.com/retrieve/pii/S0169207021001850

[63] J. Peng, A. Kimmig, J. Wang, X. Liu, Z. Niu, J. Ovtcharova, Dual-stage attention-based long-short-term memory neural networks for energy demand prediction, Energy and Buildings 249 (2021) 111211. doi:10.1016/j.enbuild.2021.111211.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378778821004953

[64] J. Lin, J. Ma, J. Zhu, Y. Cui, Short-term load forecasting based on LSTM networks considering attention mechanism, International Journal of Electrical Power & Energy Systems 137 (2022) 107818. doi:10.1016/j.ijepes.2021.107818.
URL https://linkinghub.elsevier.com/retrieve/pii/S0142061521010346

[65] Y. Yang, W. Li, T. A. Gulliver, S. Li, Bayesian Deep Learning-Based Probabilistic Load Forecasting in Smart Grids, IEEE Transactions on Industrial Informatics 16 (7) (2020) 4703–4713. doi:10.1109/TII.2019.2942353.
URL https://ieeexplore.ieee.org/document/8844831/

[66] Y. Wang, N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, C. Kang, Combining Probabilistic Load Forecasts, IEEE Transactions on Smart Grid 10 (4) (2019) 3664–3674. doi:10.1109/TSG.2018.2833869.
URL https://ieeexplore.ieee.org/document/8356127/

[67] Y. Yang, W. Hong, S. Li, Deep ensemble learning based probabilistic load forecasting in smart grids, Energy 189 (2019) 116324. doi:10.1016/j.energy.2019.116324.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544219320195

[68] T. Li, Y. Wang, N. Zhang, Combining Probability Density Forecasts for Power Electrical Loads, IEEE Transactions on Smart Grid 11 (2) (2020) 1679–1690. doi:10.1109/TSG.2019.2942024.

[69] S. Zhang, Y. Wang, Y. Zhang, D. Wang, N. Zhang, Load probability density forecasting by transforming and combining quantile forecasts, Applied Energy 277 (2020) 115600. doi:10.1016/j.apenergy.2020.115600.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920311065

[70] T. Xiong, Y. Bao, Z. Hu, Interval forecasting of electricity demand: A novel bivariate EMD-based support vector regression modeling framework, International Journal of Electrical Power & Energy Systems 63 (2014) 353–362. doi:10.1016/j.ijepes.2014.06.010.
URL https://linkinghub.elsevier.com/retrieve/pii/S0142061514003573

[71] M. Rafiei, T. Niknam, J. Aghaei, M. Shafie-Khah, J. P. S. Catalao, Probabilistic Load Forecasting Using an Improved Wavelet Neural Network Trained by Generalized Extreme Learning Machine, IEEE Transactions on Smart Grid 9 (6) (2018) 6961–6971. doi:10.1109/TSG.2018.2807845.
URL https://ieeexplore.ieee.org/document/8298533/

[72] L. Alfieri, P. De Falco, Wavelet-Based Decompositions in Probabilistic Load Forecasting, IEEE Transactions on Smart Grid 11 (2) (2020) 1367–1376. doi:10.1109/TSG.2019.2937072.
URL https://ieeexplore.ieee.org/document/8809919/

[73] F. He, J. Zhou, L. Mo, K. Feng, G. Liu, Z. He, Day-ahead short-term load probability density forecasting method with a decomposition-based quantile regression forest, Applied Energy 262 (2020) 114396. doi:10.1016/j.apenergy.2019.114396.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261919320835

[74] H. Quan, D. Srinivasan, A. Khosravi, S. Nahavandi, D. Creighton, Construction of neural network-based prediction intervals for short-term electrical load forecasting, in: 2013 IEEE Computational Intelligence Applications in Smart Grid (CIASG), IEEE, Singapore, Singapore, 2013, pp. 66–72. doi:10.1109/CIASG.2013.6611500.
URL http://ieeexplore.ieee.org/document/6611500/

[75] H. Quan, D. Srinivasan, A. Khosravi, Uncertainty handling using neural network-based prediction intervals for electrical load forecasting, Energy 73 (2014) 916–925. doi:10.1016/j.energy.2014.06.104.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544214008032

[76] C. Zhao, C. Wan, Y. Song, Z. Cao, Optimal Nonparametric Prediction Intervals of Electricity Load, IEEE Transactions on Power Systems 35 (3) (2020) 2467–2470. doi:10.1109/TPWRS.2020.2965799.
URL https://ieeexplore.ieee.org/document/8955954/

[77] X. Serrano-Guerrero, M. Briceño-León, J.-M. Clairand, G. Escrivá-Escrivá, A new interval prediction methodology for short-term electric load forecasting based on pattern recognition, Applied Energy 297 (2021) 117173. doi:10.1016/j.apenergy.2021.117173.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921006048

[78] Y. He, R. Liu, H. Li, S. Wang, X. Lu, Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory, Applied Energy 185 (2017) 254–266. doi:10.1016/j.apenergy.2016.10.079.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261916315239

[79] W. Zhang, H. Quan, D. Srinivasan, Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination, Energy 160 (2018) 810–819. doi:10.1016/j.energy.2018.07.019.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544218313148

[80] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, C. Kang, Probabilistic individual load forecasting using pinball loss guided LSTM, Applied Energy 235 (2019) 10–20. doi:10.1016/j.apenergy.2018.10.078.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261918316465

[81] W. Zhang, H. Quan, O. Gandhi, R. Rajagopal, C.-W. Tan, D. Srinivasan, Improving Probabilistic Load Forecasting Using Quantile Regression NN With Skip Connections, IEEE Transactions on Smart Grid 11 (6) (2020) 5442–5450. doi:10.1109/TSG.2020.2995777.
URL https://ieeexplore.ieee.org/document/9096383/

[82] Y. Yang, S. Li, W. Li, M. Qu, Power load probability density forecasting using Gaussian process quantile regression, Applied Energy 213 (2018) 499–509. doi:10.1016/j.apenergy.2017.11.035.

URL https://linkinghub.elsevier.com/retrieve/pii/S0306261917316100

[83] S. Lu, Q. Xu, C. Jiang, Y. Liu, A. Kusiak, Probabilistic load forecasting with a non-crossing sparse-group Lasso-quantile regression deep neural network, Energy 242 (2022) 122955. doi:10.1016/j.energy.2021.122955.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544221032047

[84] Y. Wang, G. Hug, Z. Liu, N. Zhang, Modeling load forecast uncertainty using generative adversarial networks, Electric Power Systems Research 189 (2020) 106732. doi:10.1016/j.epsr.2020.106732.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378779620305356

[85] L. Ge, W. Liao, S. Wang, B. Bak-Jensen, J. R. Pillai, Modeling Daily Load Profiles of Distribution Network for Scenario Generation Using Flow-Based Generative Network, IEEE Access 8 (2020) 77587–77597. doi:10.1109/ACCESS.2020.2989350.
URL https://ieeexplore.ieee.org/document/9075210/

[86] J. Dumas, A. Wehenkel, D. Lanaspeze, B. Cornélusse, A. Sutera, A deep generative model for probabilistic energy forecasting in power systems: normalizing flows, Applied Energy 305 (2022) 117871. doi:10.1016/j.apenergy.2021.117871.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921011909

[87] J. Munkhammar, D. van der Meer, J. Widén, Very short term load forecasting of residential electricity consumption using the Markov-chain mixture distribution (MCM) model, Applied Energy 282 (2021) 116180. doi:10.1016/j.apenergy.2020.116180.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261920315816

[88] B. Wang, M. Mazhari, C. Y. Chung, A Novel Hybrid Method for Short-Term Probabilistic Load Forecasting in Distribution Networks, IEEE Transactions on Smart Grid 13 (5) (2022) 3650–3661. doi:10.1109/TSG.2022.3171499.
URL https://ieeexplore.ieee.org/document/9765577/

[89] D. van der Meer, M. Shepero, A. Svensson, J. Widén, J. Munkhammar, Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes, Applied Energy 213 (2018) 195–207. doi:10.1016/j.apenergy.2017.12.104.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261917318275

[90] P. Kou, F. Gao, A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises, International Journal of Electrical Power & Energy Systems 55 (2014) 144–154. doi:10.1016/j.ijepes.2013.09.002.
URL https://linkinghub.elsevier.com/retrieve/pii/S014206151300375X

[91] D. Cao, J. Zhao, W. Hu, Y. Zhang, Q. Liao, Z. Chen, F. Blaabjerg, Robust Deep Gaussian Process-Based Probabilistic Electrical Load Forecasting Against Anomalous Events, IEEE Transactions on Industrial Informatics 18 (2) (2022) 1142–1153. doi:10.1109/TII.2021.3081531.
URL https://ieeexplore.ieee.org/document/9435025/

[92] M. Blum, M. Riedmiller, Electricity demand forecasting using gaussian processes, in: AAAI Workshop - Technical Report, 2013.

[93] M. Shepero, D. van der Meer, J. Munkhammar, J. Widén, Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data, Applied Energy 218 (2018) 159–172. doi:10.1016/j.apenergy.2018.02.165.
URL https://linkinghub.elsevier.com/retrieve/pii/S030626191830299X

[94] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.
URL https://proceedings.neurips.cc/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf

[95] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, M. Khazaei, Deep-Based Conditional Probability Density Function Forecasting of Residential Loads, IEEE Transactions on Smart Grid 11 (4) (2020) 3646–3657. doi:10.1109/TSG.2020.2972513.
URL https://ieeexplore.ieee.org/document/8988175/

[96] A. Brusaferri, M. Matteucci, S. Spinelli, A. Vitali, Probabilistic electric load forecasting through Bayesian Mixture Density Networks, Applied Energy 309 (2022) 118341. doi:10.1016/j.apenergy.2021.118341.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921015907

[97] R. Koenker, Quantile Regression, Cambridge University Press, Cambridge, 2005. doi:10.1017/CBO9780511754098.
URL https://www.cambridge.org/core/books/quantile-regression/C18AE7BCF3EC43C16937390D44A328B1

[98] J. D. Lara, O. Dowson, K. Doubleday, B.-M. Hodge, D. S. Callaway, A Multi-Stage Stochastic Risk Assessment With Markovian Representation of Renewable Power, IEEE Transactions on Sustainable Energy 13 (1) (2022) 414–426. doi:10.1109/TSTE.2021.3114615.
URL https://ieeexplore.ieee.org/document/9546644/

[99] G. Papaefthymiou, B. Klockl, MCMC for Wind Power Simulation, IEEE Transactions on Energy Conversion 23 (1) (2008) 234–240. doi:10.1109/TEC.2007.914174.
URL http://ieeexplore.ieee.org/document/4453993/

[100] D. Apostolopoulou, Z. De Greve, M. McCulloch, Robust Optimization for Hydroelectric System Operation Under Uncertainty, IEEE Transactions on Power Systems 33 (3) (2018) 3337–3348. doi:10.1109/TPWRS.2018.2807794.
URL http://ieeexplore.ieee.org/document/8295134/

[101] K. Dab, K. Agbossou, N. Henao, Y. Dubé, S. Kelouwani, S. S. Hosseini, A compositional kernel based gaussian process approach to day-ahead residential load forecasting, Energy and Buildings 254 (2022) 111459. doi:10.1016/j.enbuild.2021.111459.
URL https://linkinghub.elsevier.com/retrieve/pii/S037877882100743X

[102] Y. Gal, Z. Ghahramani, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, in: Proceedings of The 33rd International Conference on Machine Learning, PMLR, 2016, pp. 1050–1059, iSSN: 1938-7228.
URL https://proceedings.mlr.press/v48/gal16.html

[103] T. Vanderschueren, T. Verdonck, B. Baesens, W. Verbeke, Predict-then-optimize or predict-and-optimize? An empirical evaluation of cost-sensitive learning strategies, Information Sciences 594 (2022) 400–415. doi:10.1016/j.ins.2022.02.021.
URL https://linkinghub.elsevier.com/retrieve/pii/S0020025522001542

[104] B. Hobbs, S. Jitprapaikulsarn, S. Konda, V. Chankong, K. Loparo, D. Maratukulam, Analysis of the value for unit commitment of improved load forecasts, IEEE Transactions on Power Systems 14 (4) (1999) 1342–1348. doi:10.1109/59.801894.
URL http://ieeexplore.ieee.org/document/801894/

[105] S. Luo, Y. Weng, A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources, Applied Energy 242 (2019) 1497–1512. doi:10.1016/j.apenergy.2019.03.129.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261919305380

[106] L. Sang, Y. Xu, H. Long, Q. Hu, H. Sun, Electricity Price Prediction for Energy Storage System Arbitrage: A Decision-Focused Approach, IEEE Transactions on Smart Grid 13 (4) (2022) 2822–2832. doi:10.1109/TSG.2022.3166791.
URL https://ieeexplore.ieee.org/document/9755891/

[107] C. Zhao, C. Wan, Y. Song, Cost-Oriented Prediction Intervals: On Bridging the Gap Between Forecasting and Decision, IEEE Transactions on Power Systems 37 (4) (2022) 3048–3062. doi:10.1109/TPWRS.2021.3128567.
URL https://ieeexplore.ieee.org/document/9616458/

[108] Z. Liang, R. Mieth, Y. Dvorkin, Operation-adversarial scenario generation, Electric Power Systems Research 212 (2022) 108451. doi:10.1016/j.epsr.2022.108451.
URL https://linkinghub.elsevier.com/retrieve/pii/S037877962200582X

[109] P. Donti, B. Amos, J. Z. Kolter, Task-based End-to-end Model Learning in Stochastic Optimization, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.
URL https://proceedings.neurips.cc/paper/2017/file/3fc2c60b5782f641f76bcefc39fb2392-Paper.pdf

[110] J. Zhang, Y. Wang, G. Hug, Cost-oriented load forecasting, Electric Power Systems Research 205 (2022) 107723. doi:10.1016/j.epsr.2021.107723.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378779621007045

[111] J. Wang, A. AlShelahi, M. You, E. Byon, R. Saigal, Integrative Density Forecast and Uncertainty Quantification of Wind Power Generation, IEEE Transactions on Sustainable Energy 12 (4) (2021) 1864–1875. doi:10.1109/TSTE.2021.3069111.
URL https://ieeexplore.ieee.org/document/9387546/

[112] P. Fortenbacher, J. L. Mathieu, G. Andersson, Modeling and Optimal Operation of Distributed Battery Storage in Low Voltage Grids, IEEE Transactions on Power Systems 32 (6) (2017) 4340–4350. doi:10.1109/TPWRS.2017.2682339.
URL http://ieeexplore.ieee.org/document/7879188/

[113] T. Hong, S. Fan, Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting 32 (3) (2016) 914–938.

doi:10.1016/j.ijforecast.2015.11.011.
URL https://linkinghub.elsevier.com/retrieve/pii/S0169207015001508

[114] J. M. Morales, J. Perez-Ruiz, Point Estimate Schemes to Solve the Probabilistic Power Flow, IEEE Transactions on Power Systems 22 (4) (2007) 1594–1601. doi:10.1109/TPWRS.2007.907515.
URL http://ieeexplore.ieee.org/document/4349110/

[115] Rui Bo, Fangxing Li, Probabilistic LMP Forecasting Considering Load Uncertainty, IEEE Transactions on Power Systems 24 (3) (2009) 1279–1289. doi:10.1109/TPWRS.2009.2023268.
URL http://ieeexplore.ieee.org/document/5161350/

[116] H. Zhang, P. Li, Chance Constrained Programming for Optimal Power Flow Under Uncertainty, IEEE Transactions on Power Systems 26 (4) (2011) 2417–2424. doi:10.1109/TPWRS.2011.2154367.
URL http://ieeexplore.ieee.org/document/5871715/

[117] T. Adefarati, R. Bansal, Reliability, economic and environmental analysis of a microgrid system in the presence of renewable energy resources, Applied Energy 236 (2019) 1089–1114. doi:10.1016/j.apenergy.2018.12.050.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261918318671

[118] V. Dvorkin, P. Van Hentenryck, J. Kazempour, P. Pinson, Differentially Private Distributed Optimal Power Flow, in: 2020 59th IEEE Conference on Decision and Control (CDC), IEEE, Jeju, Korea (South), 2020, pp. 2092–2097. doi:10.1109/CDC42340.2020.9303768.
URL https://ieeexplore.ieee.org/document/9303768/

[119] A. Maghami, E. Ursavas, A. Cherukuri, A two-step approach to Wasserstein distributionally robust chance- and security-constrained dispatch, IEEE Transactions on Power Systems (2023) 1–13doi:10.1109/TPWRS.2023.3242468.
URL https://ieeexplore.ieee.org/document/10037220/

[120] Q. Zhang, F. Bu, Y. Guo, Z. Wang, Tractable Data Enriched Distributionally Robust Chance-Constrained Conservation Voltage Reduction, IEEE Transactions on Power Systems (2023) 1–15doi:10.1109/TPWRS.2023.3244895.
URL https://ieeexplore.ieee.org/document/10044219/

[121] R. Rockafellar, S. Uryasev, Conditional value-at-risk for general loss distributions, Journal of Banking & Finance 26 (7) (2002) 1443–1471. doi:10.1016/S0378-4266(02)00271-6.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378426602002716

[122] R. Moreno, M. Panteli, P. Mancarella, H. Rudnick, T. Lagos, A. Navarro, F. Ordonez, J. C. Araneda, From Reliability to Resilience: Planning the Grid Against the Extremes, IEEE Power and Energy Magazine 18 (4) (2020) 41–53. doi:10.1109/MPE.2020.2985439.
URL https://ieeexplore.ieee.org/document/9120304/

[123] F. Pourahmadi, J. Kazempour, Distributionally Robust Generation Expansion Planning With Unimodality and Risk Constraints, IEEE Transactions on Power Systems 36 (5) (2021) 4281–4295. doi:10.1109/TPWRS.2021.3057265.
URL https://ieeexplore.ieee.org/document/9347813/

[124] K. Baker, A. Bernstein, Joint Chance Constraints in AC Optimal Power Flow: Improving Bounds Through Learning, IEEE Transactions on Smart Grid 10 (6) (2019) 6376–6385. doi:10.1109/TSG.2019.2903767.
URL https://ieeexplore.ieee.org/document/8662704/

[125] M. Jia, G. Hug, C. Shen, Iterative Decomposition of Joint Chance Constraints in OPF, IEEE Transactions on Power Systems 36 (5) (2021) 4836–4839, conference Name: IEEE Transactions on Power Systems. doi:10.1109/TPWRS.2021.3072541.

[126] X. Li, C. Li, G. Chen, Z. Y. Dong, A Data-driven Joint Chance-constrained Game for Renewable Energy Aggregators in the Local Market, IEEE Transactions on Smart Grid (2022) 1–1doi:10.1109/TSG.2022.3163131.
URL https://ieeexplore.ieee.org/document/9744121/

[127] M. Vrakopoulou, K. Margellos, J. Lygeros, G. Andersson, Probabilistic Guarantees for the N-1 Security of Systems with Wind Power Generation, in: R. Billinton, R. Karki, A. K. Verma (Eds.), Reliability and Risk Evaluation of Wind Integrated Power Systems, Springer India, India, 2013, pp. 59–73. doi:10.1007/978-81-322-0987-4_5.
URL http://link.springer.com/10.1007/978-81-322-0987-4_5

[128] L. Roald, G. Andersson, Chance-Constrained AC Optimal Power Flow: Reformulations and Efficient Algorithms, IEEE Transactions on Power Systems 33 (3) (2018) 2906–2918. doi:10.1109/TPWRS.2017.2745410.
URL http://ieeexplore.ieee.org/document/8017474/

[129] K. Margellos, P. Goulart, J. Lygeros, On the Road Between Robust Optimization and the Scenario Approach for Chance Constrained Optimization Problems, IEEE Transactions on Automatic Control 59 (8) (2014) 2258–2263. doi:10.1109/TAC.2014.2303232.
URL http://ieeexplore.ieee.org/document/6727399/

[130] A. Pena-Ordieres, D. K. Molzahn, L. A. Roald, A. Wachter, DC Optimal Power Flow With Joint Chance Constraints, IEEE Transactions on Power Systems 36 (1) (2021) 147–158. doi:10.1109/TPWRS.2020.3004023.
URL https://ieeexplore.ieee.org/document/9122389/

[131] A. Nemirovski, A. Shapiro, Convex Approximations of Chance Constrained Programs, SIAM Journal on Optimization 17 (4) (2007) 969–996. doi:10.1137/050622328.
URL http://epubs.siam.org/doi/10.1137/050622328

[132] W. Chen, M. Sim, J. Sun, C.-P. Teo, From CVaR to Uncertainty Set: Implications in Joint Chance-Constrained Optimization, Operations Research 58 (2) (2010) 470–485. doi:10.1287/opre.1090.0712.
URL http://pubsonline.informs.org/doi/10.1287/opre.1090.0712

[133] H. Sun, Z. Gao, W. Y. Szeto, J. Long, F. Zhao, A Distributionally Robust Joint Chance Constrained Optimization Model for the Dynamic Network Design Problem under Demand Uncertainty, Networks and Spatial Economics 14 (3-4) (2014) 409–433. doi:10.1007/s11067-014-9236-8.
URL http://link.springer.com/10.1007/s11067-014-9236-8

[134] W. Xie, S. Ahmed, R. Jiang, Optimized Bonferroni approximations of distributionally robust joint chance constraints, Mathematical Programming 191 (1) (2019) 79–112. doi:10.1007/s10107-019-01442-8.
URL https://link.springer.com/10.1007/s10107-019-01442-8

[135] P. Zhao, C. Gu, D. Huo, Y. Shen, I. Hernando-Gil, Two-Stage Distributionally Robust Optimization for Energy Hub Systems, IEEE Transactions on Industrial Informatics 16 (5) (2020) 3460–3469. doi:10.1109/TII.2019.2938444.
URL https://ieeexplore.ieee.org/document/8820049/

[136] A. Arrigo, C. Ordoudis, J. Kazempour, Z. De Grève, J.-F. Toubeau, F. Vallée, Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation, European Journal of Operational Research 296 (1) (2022) 304–322. doi:10.1016/j.ejor.2021.04.015.
URL https://linkinghub.elsevier.com/retrieve/pii/S0377221721003271

[137] X. Fang, B.-M. Hodge, E. Du, C. Kang, F. Li, Introducing Uncertainty Components in Locational Marginal Prices for Pricing Wind Power and Load Uncertainties, IEEE Transactions on Power Systems 34 (3) (2019) 2013–2024. doi:10.1109/TPWRS.2018.2881131.
URL https://ieeexplore.ieee.org/document/8601342/

[138] B. Zhao, T. Qian, W. Tang, Q. Liang, A data-enhanced distributionally robust optimization method for economic dispatch of integrated electricity and natural gas systems with wind uncertainty, Energy 243 (2022) 123113. doi:10.1016/j.energy.2022.123113.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544222000160

[139] Z. Shi, H. Liang, S. Huang, V. Dinavahi, Distributionally Robust Chance-Constrained Energy Management for Islanded Microgrids, IEEE Transactions on Smart Grid 10 (2) (2019) 2234–2244. doi:10.1109/TSG.2018.2792322.
URL https://ieeexplore.ieee.org/document/8254387/

[140] B. Li, R. Jiang, J. L. Mathieu, Distributionally Robust Chance-Constrained Optimal Power Flow Assuming Unimodal Distributions With Misspecified Modes, IEEE Transactions on Control of Network Systems 6 (3) (2019) 1223–1234. doi:10.1109/TCNS.2019.2930872.
URL https://ieeexplore.ieee.org/document/8771187/

[141] Y. Chen, Q. Guo, H. Sun, Z. Li, W. Wu, Z. Li, A Distributionally Robust Optimization Model for Unit Commitment Based on Kullback–Leibler Divergence, IEEE Transactions on Power Systems 33 (5) (2018) 5147–5160. doi:10.1109/TPWRS.2018.2797069.
URL https://ieeexplore.ieee.org/document/8267333/

[142] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, A. Cherukuri, Wasserstein Distributionally Robust Look-Ahead Economic Dispatch, IEEE Transactions on Power Systems 36 (3) (2021) 2010–2022. doi:10.1109/TPWRS.2020.3034488.
URL https://ieeexplore.ieee.org/document/9242289/

[143] S. Wang, C. Zhao, L. Fan, R. Bo, Distributionally Robust Unit Commitment With Flexible Generation Resources Considering Renewable Energy Uncertainty, IEEE Transactions on Power Systems 37 (6) (2022) 4179–4190. doi:10.1109/TPWRS.2022.3149506.
URL https://ieeexplore.ieee.org/document/9707673/

[144] C. Ordoudis, V. A. Nguyen, D. Kuhn, P. Pinson, Energy and reserve dispatch with distributionally robust joint chance constraints, Operations Research Letters 49 (3) (2021) 291–299. doi:10.1016/j.orl.2021.01.012.
URL https://linkinghub.elsevier.com/retrieve/pii/S0167637721000213

[145] L. Cipcigan, P. Taylor, Investigation of the reverse power flow requirements of high penetrations of small-scale embedded generation, IET Renewable Power Generation 1 (3) (2007) 160. doi:10.1049/iet-rpg:20070011.
URL https://digital-library.theiet.org/content/journals/10.1049/iet-rpg_20070011

[146] J. P. Lopes, N. Hatziargyriou, J. Mutale, P. Djapic, N. Jenkins, Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities, Electric Power Systems Research 77 (9) (2007) 1189–1203. doi:10.1016/j.epsr.2006.08.016.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378779606001908

[147] Q. Hou, N. Zhang, E. Du, M. Miao, F. Peng, C. Kang, Probabilistic duck curve in high PV penetration power system: Concept, modeling, and empirical analysis in China, Applied Energy 242 (2019) 205–215. doi:10.1016/j.apenergy.2019.03.067.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261919304763

[148] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, Q. Xia, Data-Driven Probabilistic Net Load Forecasting With High Penetration of Behind-the-Meter PV, IEEE Transactions on Power Systems 33 (3) (2018) 3255–3264. doi:10.1109/TPWRS.2017.2762599.
URL http://ieeexplore.ieee.org/document/8066371/

[149] R. Saeedi, S. K. Sadanandan, A. K. Srivastava, K. L. Davies, A. H. Gebremedhin, An Adaptive Machine Learning Framework for Behind-the-Meter Load/PV Disaggregation, IEEE Transactions on Industrial Informatics 17 (10) (2021) 7060–7069. doi:10.1109/TII.2021.3060898.
URL https://ieeexplore.ieee.org/document/9360481/

[150] L. Cheng, H. Zang, Y. Xu, Z. Wei, G. Sun, Probabilistic Residential Load Forecasting Based on Micrometeorological Data and Customer Consumption Pattern, IEEE Transactions on Power Systems 36 (4) (2021) 3762–3775. doi:10.1109/TPWRS.2021.3051684.
URL https://ieeexplore.ieee.org/document/9324967/

[151] D. K. Duvenaud, H. Nickisch, C. Rasmussen, Additive Gaussian Processes, in: J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 24, Curran Associates, Inc., 2011.
URL https://proceedings.neurips.cc/paper/2011/file/4c5bde74a8f110656874902f07378009-Paper.pdf

[152] M. Seeger, GAUSSIAN PROCESSES FOR MACHINE LEARNING, International Journal of Neural Systems 14 (02) (2004) 69–106. doi:10.1142/S0129065704001899.
URL https://www.worldscientific.com/doi/abs/10.1142/S0129065704001899

[153] UK Power Networks, Validation of Photovoltaic (PV) Connection Assessment Tool, Tech. rep. (2015).
URL https://www.ofgem.gov.uk/ofgem-publications/93938/pvtoolcdrfinal-pdf

[154] T. Gneiting, A. E. Raftery, A. H. Westveld, T. Goldman, Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation, Monthly Weather Review 133 (5) (2005) 1098–1118. doi:

10.1175/MWR2904.1.
URL http://journals.ametsoc.org/doi/10.1175/MWR2904.1

[155] S. Elsworth, S. Güttel, Time series forecasting using lstm networks: A symbolic approach (2020). arXiv:2003.05672.

[156] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, G. Sun, Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations, Renewable Energy 160 (2020) 26–41. doi:10.1016/j.renene.2020.05.150.
URL https://linkinghub.elsevier.com/retrieve/pii/S0960148120308557

[157] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow.org (2015).
URL https://www.tensorflow.org/

[158] S. C. Johnson, D. J. Papageorgiou, D. S. Mallapragada, T. A. Deetjen, J. D. Rhodes, M. E. Webber, Evaluating rotational inertia as a component of grid reliability with high penetrations of variable renewable energy, Energy 180 (2019) 258–271. doi:10.1016/j.energy.2019.04.216.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544219308564

[159] I. Penn, Climate Change and Poor Planning Are Blamed for California Blackouts (oct 2020).
URL https://www.nytimes.com/2020/10/06/business/energy-environment/california-blackout-cause-report.html

[160] T. Levin, A. Botterud, W. N. Mann, J. Kwon, Z. Zhou, Extreme weather and electricity markets: Key lessons from the February 2021 Texas crisis, Joule 6 (1) (2022) 1–7. doi:10.1016/j.joule.2021.12.015.
URL https://linkinghub.elsevier.com/retrieve/pii/S2542435121005845

[161] N. Pogaku, M. Prodanovic, T. C. Green, Modeling, Analysis and Testing of Autonomous Operation of an Inverter-Based Microgrid, IEEE Transactions on Power Electronics 22 (2) (2007) 613–625. doi:10.1109/TPEL.2006.890003.
URL http://ieeexplore.ieee.org/document/4118327/

[162] F. Teng, V. Trovato, G. Strbac, Stochastic Scheduling With Inertia-Dependent Fast Frequency Response Requirements, IEEE Transactions on Power Systems 31 (2) (2016) 1557–1566. doi:10.1109/TPWRS.2015.2434837.
URL http://ieeexplore.ieee.org/document/7115982/

[163] A. Khodaei, Microgrid Optimal Scheduling With Multi-Period Islanding Constraints, IEEE Transactions on Power Systems 29 (3) (2014) 1383–1392. doi:10.1109/TPWRS.2013.2290006.
URL http://ieeexplore.ieee.org/document/6671442/

[164] G. Liu, M. Starke, B. Xiao, X. Zhang, K. Tomsovic, Microgrid optimal scheduling with chance-constrained islanding capability, Electric Power Systems Research 145 (2017) 197–206. doi:10.1016/j.epsr.2017.01.014.

URL https://linkinghub.elsevier.com/retrieve/pii/S0378779617300238

[165] N. Li, L. Chen, S. H. Low, Optimal demand response based on utility maximization in power networks, in: 2011 IEEE Power and Energy Society General Meeting, IEEE, San Diego, CA, 2011, pp. 1–8. doi:10.1109/PES.2011.6039082.
URL https://ieeexplore.ieee.org/document/6039082/

[166] S. Bahrami, M. H. Amini, M. Shafie-Khah, J. P. S. Catalao, A Decentralized Renewable Generation Management and Demand Response in Power Distribution Networks, IEEE Transactions on Sustainable Energy 9 (4) (2018) 1783–1797. doi:10.1109/TSTE.2018.2815502.
URL https://ieeexplore.ieee.org/document/8314720/

[167] G. Calafiore, M. Campi, The Scenario Approach to Robust Control Design, IEEE Transactions on Automatic Control 51 (5) (2006) 742–753. doi:10.1109/TAC.2006.875041.
URL http://ieeexplore.ieee.org/document/1632303/

[168] H. Rahimian, S. Mehrotra, Distributionally Robust Optimization: A Review 60208 (1) (2019). arXiv:1908.05659.
URL http://arxiv.org/abs/1908.05659

[169] Y. Zhang, S. Shen, J. Mathieu, Distributionally Robust Chance-Constrained Optimal Power Flow with Uncertain Renewables and Uncertain Reserves Provided by Loads, IEEE Transactions on Power Systems (2016) 1–1doi:10.1109/TPWRS.2016.2572104.
URL http://ieeexplore.ieee.org/document/7478165/

[170] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, T. H. Summers, Data-Based Distributionally Robust Stochastic Optimal Power Flow—Part I: Methodologies, IEEE Transactions on Power Systems 34 (2) (2019) 1483–1492. doi:10.1109/TPWRS.2018.2878385.
URL https://ieeexplore.ieee.org/document/8513852/

[171] C. Duan, W. Fang, L. Jiang, L. Yao, J. Liu, Distributionally Robust Chance-Constrained Approximate AC-OPF With Wasserstein Metric, IEEE Transactions on Power Systems 33 (5) (2018) 4924–4936. doi:10.1109/TPWRS.2018.2807623.
URL https://ieeexplore.ieee.org/document/8294298/

[172] H. Qiu, W. Gu, Y. Xu, B. Zhao, Multi-Time-Scale Rolling Optimal Dispatch for AC/DC Hybrid Microgrids With Day-Ahead Distributionally Robust Scheduling, IEEE Transactions on Sustainable Energy 10 (4) (2019) 1653–1663. doi:10.1109/TSTE.2018.2868548.
URL https://ieeexplore.ieee.org/document/8454320/

[173] G. A. Hanasusanto, V. Roitch, D. Kuhn, W. Wiesemann, Ambiguous Joint Chance Constraints Under Mean and Dispersion Information, Operations Research 65 (3) (2017) 751–767. doi:10.1287/opre.2016.1583.
URL http://pubsonline.informs.org/doi/10.1287/opre.2016.1583

[174] M. Vrakopoulou, K. Margellos, J. Lygeros, G. Andersson, A probabilistic framework for security constrained reserve scheduling of networks with wind power generation, in: 2012 IEEE International Energy Conference and Exhibition (ENERGYCON), IEEE, Florence, Italy, 2012, pp. 452–457. doi:10.1109/EnergyCon.2012.6348195.
URL http://ieeexplore.ieee.org/document/6348195/

[175] J. Warrington, P. J. Goulart, S. Mariethoz, M. Morari, Robust reserve operation in power systems using affine policies, in: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, Maui, HI, USA, 2012, pp. 1111–1117. doi:10.1109/CDC.2012.6425913.
URL http://ieeexplore.ieee.org/document/6425913/

[176] M. Vrakopoulou, B. Li, J. L. Mathieu, Chance Constrained Reserve Scheduling Using Uncertain Controllable Loads Part I: Formulation and Scenario-Based Analysis, IEEE Transactions on Smart Grid 10 (2) (2019) 1608–1617. doi:10.1109/TSG.2017.2773627.
URL https://ieeexplore.ieee.org/document/8110715/

[177] S. Zymler, D. Kuhn, B. Rustem, Distributionally robust joint chance constraints with second-order moment information, Mathematical Programming 137 (1-2) (2013) 167–198. doi:10.1007/s10107-011-0494-7.
URL http://link.springer.com/10.1007/s10107-011-0494-7

[178] T. Morstyn, B. Hredzak, V. G. Agelidis, Dynamic optimal power flow for DC microgrids with distributed battery energy storage systems, in: 2016 IEEE Energy Conversion Congress and Exposition (ECCE), IEEE, Milwaukee, WI, USA, 2016, pp. 1–6. doi:10.1109/ECCE.2016.7855059.
URL http://ieeexplore.ieee.org/document/7855059/

[179] E. Delage, Y. Ye, Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems, Operations Research 58 (3) (2010) 595–612. doi:10.1287/opre.1090.0741.
URL http://pubsonline.informs.org/doi/10.1287/opre.1090.0741

[180] G. C. Calafiore, L. E. Ghaoui, On distributionally robust chance-constrained linear programs, Journal of Optimization Theory and Applications 130 (1) (2006) 1–22. doi:10.1007/s10957-006-9084-x.

[181] I. Popescu, A Semidefinite Programming Approach to Optimal-Moment Bounds for Convex Classes of Distributions, Mathematics of Operations Research 30 (3) (2005) 632–657. doi:10.1287/moor.1040.0137.
URL http://pubsonline.informs.org/doi/10.1287/moor.1040.0137

[182] B. P. G. Van Parys, P. J. Goulart, D. Kuhn, Generalized Gauss inequalities via semidefinite programming, Mathematical Programming 156 (1-2) (2016) 271–302. doi:10.1007/s10107-015-0878-1.
URL http://link.springer.com/10.1007/s10107-015-0878-1

[183] D. Floreano, C. Mattiussi, Bio-inspired artificial intelligence: theories, methods, and technologies, Choice Reviews Online 46 (08) (2009) 46–4490–46–4490. doi:10.5860/CHOICE.46-4490.
URL http://choicereviews.org/review/10.5860/CHOICE.46-4490

[184] A. E. Eiben, J. Smith, From evolutionary computation to the evolution of things, Nature 521 (7553) (2015) 476–482. doi:10.1038/nature14544.
URL https://www.nature.com/articles/nature14544

[185] B. Palmintier, J. Giraldez, P. Gotseff, A. Nagarajan, T. Harris, B. Bugbee, M. Baggu, Palmintier, Feeder Voltage Regulation with High-Penetration PV Using Advanced Inverters and a Distribution Management System A Duke Energy Case Study (2016).

[186] M. Thakur, A new genetic algorithm for global optimization of multimodal continuous functions, Journal of Computational Science 5 (2) (2014) 298–311. doi:10.1016/j.jocs.2013.05.005.
URL https://linkinghub.elsevier.com/retrieve/pii/S1877750313000677

[187] C. A. C. Coello, N. C. Cortés, Hybridizing a genetic algorithm with an artificial immune system for global optimization, Engineering Optimization 36 (5) (2004) 607–634. doi:10.1080/03052150410001704845.
URL http://www.tandfonline.com/doi/abs/10.1080/03052150410001704845

[188] RS, Cables Selection Guide.
URL https://docs.rs-online.com/a4dc/0900766b815c9b61.pdf

[189] S. Diamond, S. Boyd, CVXPY: A Python-embedded modeling language for convex optimization, Journal of Machine Learning Research 17 (83) (2016) 1–5.

[190] Y. Dvorkin, A Chance-Constrained Stochastic Electricity Market, IEEE Transactions on Power Systems 35 (4) (2020) 2993–3003. doi:10.1109/TPWRS.2019.2961231.
URL https://ieeexplore.ieee.org/document/8937824/

[191] Department for Business, Energy & Industrial Strategy, Powering our Net Zero Future, Tech. rep.
URL https://www.gov.uk/government/publications/energy-white-paper-powering-our-net-zero-future

[192] L. Bai, J. Wang, C. Wang, C. Chen, F. Li, Distribution Locational Marginal Pricing (DLMP) for Congestion Management and Voltage Support, IEEE Transactions on Power Systems 33 (4) (2018) 4061–4073. doi:10.1109/TPWRS.2017.2767632.
URL https://ieeexplore.ieee.org/document/8089425/

[193] A. Papavasiliou, Analysis of Distribution Locational Marginal Prices, IEEE Transactions on Smart Grid 9 (5) (2018) 4872–4882. doi:10.1109/TSG.2017.2673860.
URL https://ieeexplore.ieee.org/document/7862921/

[194] Z. Zhang, R. Li, F. Li, A Novel Peer-to-Peer Local Electricity Market for Joint Trading of Energy and Uncertainty, IEEE Transactions on Smart Grid 11 (2) (2020) 1205–1215. doi:10.1109/TSG.2019.2933574.
URL https://ieeexplore.ieee.org/document/8789684/

[195] Paredes, J. A. Aguado, P. Rodrí-guez, Uncertainty-Aware Trading of Congestion and Imbalance Mitigation Services for Multi-DSO Local Flexibility Markets, IEEE Transactions on Sustainable Energy (2023) 1–13doi:10.1109/TSTE.2023.3257405.
URL https://ieeexplore.ieee.org/document/10070801/

[196] W. Saad, Z. Han, H. V. Poor, Coalitional Game Theory for Cooperative Micro-Grid Distribution Networks, in: 2011 IEEE International Conference on Communications Workshops (ICC), IEEE, Kyoto, Japan, 2011, pp. 1–5. doi:10.1109/iccw.2011.5963577.
URL http://ieeexplore.ieee.org/document/5963577/

[197] T. Jiang, C. Wu, R. Zhang, X. Li, H. Chen, G. Li, Flexibility Clearing in Joint Energy and Flexibility Markets Considering TSO-DSO Coordination, IEEE Transactions on Smart Grid (2022) 1–1doi:10.1109/TSG.2022.3153634.
URL https://ieeexplore.ieee.org/document/9722366/

[198] S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, C. Ruiz, Equilibria and Complementarity Problems, in: Complementarity Modeling in Energy Markets, Vol. 180, Springer New York, New York, NY, 2013, pp. 127–179, series Title: International Series in Operations Research & Management Science. doi:10.1007/978-1-4419-6123-5_4.
URL http://link.springer.com/10.1007/978-1-4419-6123-5_4

[199] S. delaTorre, J. Contreras, A. Conejo, Finding Multiperiod Nash Equilibria in Pool-Based Electricity Markets, IEEE Transactions on Power Systems 19 (1) (2004) 643–651. doi:10.1109/TPWRS.2003.820703.
URL http://ieeexplore.ieee.org/document/1266624/

[200] W. Saad, Z. Han, H. Poor, T. Basar, Game-Theoretic Methods for the Smart Grid: An Overview of Microgrid Systems, Demand-Side Management, and Smart Grid Communications, IEEE Signal Processing Magazine 29 (5) (2012) 86–105. doi:10.1109/MSP.2012.2186410.
URL http://ieeexplore.ieee.org/document/6279592/

[201] M. Saguan, N. Keseric, P. Dessante, J.-m. Glachant, Market Power in Power Markets: Game Theory vs. Agent-Based Approach, in: 2006 IEEE/PES Transmission & Distribution Conference and Exposition: Latin America, IEEE, Caracas, Venezuela, 2006, pp. 1–6. doi:10.1109/TDCLA.2006.311439.
URL http://ieeexplore.ieee.org/document/4104670/

[202] P. Sheikhahmadi, S. Bahramara, J. Moshtagh, M. Yazdani Damavandi, A risk-based approach for modeling the strategic behavior of a distribution company in wholesale energy market, Applied Energy 214 (2018) 24–38. doi:10.1016/j.apenergy.2018.01.051.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261918300631

[203] M. Mazadi, W. D. Rosehart, H. Zareipour, O. P. Malik, M. Oloomi, Impact of wind integration on electricity markets: a chance-constrained Nash Cournot model: CHANCE-CONSTRAINED NASH COURNOT MODEL, International Transactions on Electrical Energy Systems 23 (1) (2013) 83–96. doi:10.1002/etep.650.
URL https://onlinelibrary.wiley.com/doi/10.1002/etep.650

[204] A. Gautier, M. Rigter, B. Lacerda, N. Hawes, M. Wooldridge, Risk-Constrained Planning for Multi-Agent Systems with Shared Resources (Feb. 2023).
URL https://ora.ox.ac.uk/objects/uuid:3d4289eb-d8ee-49e1-8524-b7131fe7f170/download_file?file_format=application%2Fpdf&safe_filename=Gautier_et_al_2023_risk_constrained_planning.pdf&type_of_work=Conference+item

[205] Y. Ding, T. Morstyn, M. D. McCulloch, Distributionally Robust Joint Chance-Constrained Optimization for Networked Microgrids Considering Contingencies and Renewable Uncertainty, IEEE Transactions on Smart Grid 13 (3) (2022) 2467–2478. doi:10.1109/TSG.2022.3150397.
URL https://ieeexplore.ieee.org/document/9709590/

[206] Y. Matamala, F. Feijoo, A two-stage stochastic Stackelberg model for microgrid operation with chance constraints for renewable energy generation uncertainty, Applied Energy 303 (2021) 117608. doi:10.1016/j.apenergy.2021.117608.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261921009788

[207] W. Xie, S. Ahmed, R. Jiang, Optimized Bonferroni approximations of distributionally robust joint chance constraints, Mathematical Programming 191 (1) (2022) 79–112. doi:10.1007/s10107-019-01442-8.
URL https://link.springer.com/10.1007/s10107-019-01442-8

[208] A. Arrigo, J. Kazempour, Z. De Greve, J.-F. Toubeau, F. Vallee, Embedding Dependencies Between Wind Farms in Distributionally Robust Optimal Power Flow, IEEE Transactions on Power Systems (2022) 1–14doi:10.1109/TPWRS.2022.3221348.
URL https://ieeexplore.ieee.org/document/9944957/

[209] Y. Inatsu, S. Takeno, M. Karasuyama, I. Takeuchi, Bayesian optimization for distributionally robust chance-constrained problem, in: International Conference on Machine Learning, PMLR, 2022, pp. 9602–9621.

[210] C. Ruiz, A. J. Conejo, J. D. Fuller, S. A. Gabriel, B. F. Hobbs, A tutorial review of complementarity models for decision-making in energy markets, EURO Journal on Decision Processes 2 (1-2) (2014) 91–120. doi:10.1007/s40070-013-0019-0.
URL https://linkinghub.elsevier.com/retrieve/pii/S2193943821000285

[211] D. Kuhn, P. M. Esfahani, V. A. Nguyen, S. Shafieezadeh-Abadeh, Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning, in: S. Netessine, D. Shier, H. J. Greenberg (Eds.), Operations Research & Management Science in the Age of Analytics, INFORMS, 2019, pp. 130–166. doi:10.1287/educ.2019.0198.
URL http://pubsonline.informs.org/doi/10.1287/educ.2019.0198

[212] P. Mohajerin Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations, Mathematical Programming 171 (1-2) (2018) 115–166. doi:10.1007/s10107-017-1172-1.
URL http://link.springer.com/10.1007/s10107-017-1172-1

[213] Masaki Adachi, Satoshi Hayakawa, Saad Hamid, Martin Jørgensen, Harald Oberhauser, Micheal A. Osborne, SOBER: Scalable Batch Bayesian Optimization and Quadrature using Recombination Constraints.
URL https://arxiv.org/abs/2301.11832

[214] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, N. de Freitas, Taking the Human Out of the Loop: A Review of Bayesian Optimization, Proceedings of the IEEE 104 (1) (2016) 148–175. doi:10.1109/JPROC.2015.2494218.
URL https://ieeexplore.ieee.org/document/7352306/

[215] J. Wilson, F. Hutter, M. Deisenroth, Maximizing acquisition functions for Bayesian optimization, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 31, Curran Associates, Inc., 2018.
URL https://proceedings.neurips.cc/paper/2018/file/498f2c21688f6451d9f5fd09d53edda7-Paper.pdf

[216] Power Systems Test Case Archive, 30 Bus Power Flow Test Case.
URL http://labs.ece.uw.edu/pstca/pf30/pg_tca30bus.htm

[217] UK Power Networks, Validation of Photovoltaic (PV) Connection Assessment Tool, Tech. rep.
URL https://www.ofgem.gov.uk/ofgem-publications/93938/pvtoolcdrfinal-pdf

[218] Y. Ding, M. McCulloch, Additive Gaussian process prediction for electrical loads compared with deep learning models, in: Proceedings of the Twelfth ACM International Conference on Future Energy Systems, ACM, Virtual Event Italy, 2021, pp. 499–506. doi:10.1145/3447555.3466592.
URL https://dl.acm.org/doi/10.1145/3447555.3466592

[219] S. Diamond, S. Boyd, CVXPY: a python-embedded modeling language for convex optimization, The Journal of Machine Learning Research 17 (1) (2016) 2909–2913.

[220] scikit-optimize.
URL https://scikit-optimize.github.io/stable/

[221] P. I. Frazier, A Tutorial on Bayesian OptimizationPublisher: arXiv Version Number: 1 (2018). doi:10.48550/ARXIV.1807.02811.
URL https://arxiv.org/abs/1807.02811

[222] I. J. Scott, P. M. Carvalho, A. Botterud, C. A. Silva, Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage, Applied Energy 253 (2019) 113603. doi:10.1016/j.apenergy.2019.113603.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261919312772

[223] I. Popescu, A semidefinite programming approach to optimal-moment bounds for convex classes of distributions, Mathematics of Operations Research 30 (3) (2005) 632–657. doi:10.1287/moor.1040.0137.

[224] M. V. Ramana, L. Tunçel, H. Wolkowicz, Strong Duality for Semidefinite Programming, SIAM Journal on Optimization 7 (3) (1997) 641–662. doi:10.1137/S1052623495288350.
URL http://epubs.siam.org/doi/10.1137/S1052623495288350

[225] A. Clements, M. McCulloch, Time series lighting electricity data for rural households using Solar Nano-grids in Kenya (2019).
URL 10.17632/4yv37hngp6.1