

PROBABILISTIC RENEWABLE ENERGY FORECASTING BY CONSIDERING  
SPATIAL-TEMPORAL CORRELATION

by

Mucun Sun



APPROVED BY SUPERVISORY COMMITTEE:

---

Jie Zhang, Chair

---

Giacomo Valerio Iungo

---

Stefano Leonardi

---

Yulia Gel

Copyright © 2020

Mucun Sun

All rights reserved

*To my family.*

PROBABILISTIC RENEWABLE ENERGY FORECASTING BY CONSIDERING  
SPATIAL-TEMPORAL CORRELATION

by

MUCUN SUN, MS

DISSERTATION

Presented to the Faculty of  
The University of Texas at Dallas  
in Partial Fulfillment  
of the Requirements  
for the Degree of

DOCTOR OF PHILOSOPHY IN  
MECHANICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

August 2020

## ACKNOWLEDGMENTS

Many people offered help during my PhD. study. I am thankful for all of them.

I would like to thank my parents for their love and financial support. They are my inspiration to pursue my doctoral degree.

It is so lucky to have Dr. Jie Zhang as my PhD advisor. I would like to express my sincere gratitude to him for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my PhD study.

Besides my advisor, I would like to thank the rest of my dissertation committee: Dr. Giacomo Valerio Iungo, Dr. Stefano Leonardo, and Dr. Yulia Gel, for their insightful comments and encouragement, but also for the hard questions which incited me to widen my research from various perspectives.

My sincere thanks also go to Dr. Kate Anderson at the National Renewable Energy Laboratory, and Dr. Brian McCabe at the United Technologies Research Center, who provided me an opportunity to join their team as an intern, and who gave me access to the laboratory and research facilities. Without their precious support it would not be possible to conduct my research.

I thank my fellow lab-mates for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years.

Last but not least, I would like to thank Dr. Yong Kang who introduced me to the field of power systems and enlighten me the first glance of research. I am thankful for the guidance, responsiveness, and effort he put forth for me to build my fundamental background of power system analysis.

June 2020

PROBABILISTIC RENEWABLE ENERGY FORECASTING BY CONSIDERING  
SPATIAL-TEMPORAL CORRELATION

Mucun Sun, PhD  
The University of Texas at Dallas, 2020

Supervising Professor: Jie Zhang, Chair

Renewable energy forecasts can help reduce the amount of operating reserves needed for the system, reducing costs of balancing the system, and improving the reliability of the system. Conventional deterministic forecasts might not be sufficient to characterize the inherent uncertainty of renewable energy. Probabilistic forecasts that provide quantitative uncertainty information associated with renewable energy are therefore expected to better assist power system operations. Meanwhile, studies have shown that the integration of geographically dispersed and correlated wind/solar farms could reduce extreme power output, which is referred to as smoothing effect. In addition, power produced from one wind/solar farm at different times is typically temporally correlated. The impacts of spatial-temporal correlation between wind/solar farms on renewable energy forecasting are not well studied in the literature.

This dissertation aims at mitigating power system uncertainty by improving probabilistic renewable energy forecasting accuracy utilizing spatial-temporal correlation modeling. In this dissertation, a variety of methods, such as predictive distribution optimization, ensemble learning, deep learning, and scenario generation-based methods, are developed to assist spatial-temporal correlation modeling, and improve probabilistic forecasting accuracy. Computational experiments indicate that the presented study can provide scholars and engineers with critical insights to the usage

of spatial-temporal modeling in probabilistic renewable energy forecasting and anomaly detection, and also serve as a valuable reference for practical industry forecasting systems.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . . .	v
ABSTRACT . . . . .	vi
LIST OF FIGURES . . . . .	xi
LIST OF TABLES . . . . .	xiii
CHAPTER 1 INTRODUCTION . . . . .	1
CHAPTER 2 LITERATURE REVIEW . . . . .	5
2.1 Probabilistic Renewable Energy Forecasting . . . . .	5
2.1.1 Review of Probabilistic Renewable Energy Forecasting Approaches . . . . .	5
2.1.2 Evaluation Metrics . . . . .	11
2.2 Spatial-Temporal Modeling in Renewable Energy . . . . .	19
2.3 Cyberattacks and Anomaly Detection Approaches in Renewable Energy Forecasting . . . . .	21
CHAPTER 3 IMPROVING PROBABILISTIC FORECASTING VIA PREDICTIVE DISTRIBUTION OPTIMIZATION . . . . .	26
3.1 Motivation and Objective . . . . .	26
3.2 Methodology . . . . .	27
3.2.1 Machine Learning-Based Multimodel (M3) Deterministic Forecasting . . . . .	28
3.2.2 Multi-Distribution Database . . . . .	29
3.2.3 Pinball Loss-Based Optimization . . . . .	29
3.2.4 Surrogate Model . . . . .	30
3.3 Case Studies and Results . . . . .	30
3.4 Conclusion . . . . .	41
CHAPTER 4 IMPROVING PROBABILISTIC FORECASTING VIA ENSEMBLE LEARNING . . . . .	42
4.1 Motivation and Objective . . . . .	42
4.2 Methodology . . . . .	44
4.2.1 Q-learning enhanced deterministic forecasting . . . . .	46
4.2.2 Probabilistic wind power forecasting model . . . . .	47
4.2.3 Competitive MDE method . . . . .	50

4.2.4	Cooperative MDE method . . . . .	51
4.2.5	Surrogate Model . . . . .	52
4.3	Case Studies and Results . . . . .	53
4.4	Conclusion . . . . .	60
<b>CHAPTER 5 IMPROVING PROBABILISTIC RENEWABLE FORECASTING VIA SPATIAL-TEMPORAL CORRELATION MODELING . . . . .</b>		<b>73</b>
5.1	Motivation and Objective . . . . .	74
5.2	Methodology . . . . .	74
5.2.1	Q-learning Enhanced Deterministic Forecasting . . . . .	76
5.2.2	Wind Power Distribution Modeling . . . . .	77
5.2.3	Clustering Analysis . . . . .	78
5.2.4	Unsupervised Clustering based on Spatial Correlation . . . . .	78
5.2.5	Wind Farms spatial-temporal Correlation Modeling . . . . .	79
5.3	Case Studies and Results . . . . .	82
5.4	Conclusion . . . . .	92
<b>CHAPTER 6 IMPROVING PROBABILISTIC RENEWABLE FORECASTING VIA SCENARIO GENERATION . . . . .</b>		<b>95</b>
6.1	Motivation and Objective . . . . .	95
6.2	Methodology . . . . .	97
6.2.1	Machine Learning-based Multi-Model (M3) Deterministic Forecasting . . . . .	98
6.2.2	Weather Distribution Modeling . . . . .	99
6.2.3	Weather Scenario Generation . . . . .	100
6.2.4	Probabilistic Solar Power Forecasting . . . . .	103
6.3	Case Studies and Results . . . . .	103
6.4	Conclusion . . . . .	112
<b>CHAPTER 7 UTILIZING PROBABILISTIC FORECASTING FOR ANOMALY DETECTION IN RENEWABLE ENERGY . . . . .</b>		<b>115</b>
7.1	Motivation and Objective . . . . .	115
7.2	Methodology . . . . .	117
7.2.1	CNN-based Feature Extraction . . . . .	119

7.2.2	LSTM Network . . . . .	120
7.3	Case Studies and Results . . . . .	125
7.3.1	Data Summary . . . . .	125
7.3.2	Data Integrity Attack Templates . . . . .	126
7.3.3	CNN-LSTM Deterministic Forecasting Results . . . . .	127
7.3.4	Benchmarks and Comparison Settings . . . . .	128
7.3.5	Probabilistic Forecasting Results . . . . .	129
7.3.6	Probabilistic Anomaly Detection Results . . . . .	129
7.4	Conclusion . . . . .	132
CHAPTER 8	CONCLUSION . . . . .	135
REFERENCES	. . . . .	139
BIOGRAPHICAL SKETCH	. . . . .	150
CURRICULUM VITAE		

## LIST OF FIGURES

1.1	Wind power installation capacity by country (by 2018). . . . .	2
2.1	An example reliability diagram of probabilistic wind power forecasts . . . . .	13
2.2	Example PIT histogram of probabilistic wind power forecasts . . . . .	14
2.3	An example boxplot of probabilistic wind power prediction interval . . . . .	15
2.4	An example $\delta$ diagram of probabilistic wind power forecasts . . . . .	16
3.1	Overall framework of the pinball loss optimization-based probabilistic wind forecasting methodology . . . . .	28
3.2	M3-Laplace and quantile regression forecasts at the C2 site . . . . .	34
3.3	Reliability of probabilistic forecasts at selected sites . . . . .	36
3.3	(continued) Reliability of probabilistic forecasts at selected sites . . . . .	37
3.4	Sharpness of probabilistic forecasts at selected sites . . . . .	38
3.4	(continued) Sharpness of probabilistic forecasts at selected sites . . . . .	39
3.5	The relationship between NMAE and NPL . . . . .	39
4.1	Wind forecasting methods categorization . . . . .	43
4.2	Overall framework of the MDE competitive probabilistic wind power forecasting method	45
4.3	Overall framework of the MDE cooperative probabilistic wind power forecasting method	46
4.4	Surrogate model selection . . . . .	53
4.5	Normalized pinball loss of different ensemble methods and single methods . . . . .	58
4.6	Normalized pinball loss of different models with different look-ahead hours at the C7 site . . . . .	59
4.7	Day-ahead MDE-competitive and MDE-cooperative forecasts at the C3 site . . . . .	60
4.8	Reliability and sharpness comparison of different models at the C7 site with different look-ahead times (blue lines: reliability; orange lines: sharpness) . . . . .	61
4.9	1HA reliability and sharpness comparison of different models at the 7 sites . . . . .	63
4.9	(continued) 1HA reliability and sharpness comparison of different models at the 7 sites	64
4.10	2HA reliability and sharpness comparison of different models at the 7 sites . . . . .	64
4.10	(continued) 2HA reliability and sharpness comparison of different models at the 7 sites	65
4.11	3HA reliability and sharpness comparison of different models at the 7 sites . . . . .	65
4.11	(continued) 3HA reliability and sharpness comparison of different models at the 7 sites	66

4.12	4HA reliability and sharpness comparison of different models at the 7 sites . . . . .	67
4.12	(continued) 4HA reliability and sharpness comparison of different models at the 7 sites	68
4.13	5HA reliability and sharpness comparison of different models at the 7 sites . . . . .	68
4.13	(continued) 5HA reliability and sharpness comparison of different models at the 7 sites	69
4.14	6HA reliability and sharpness comparison of different models at the 7 sites . . . . .	69
4.14	(continued) 6HA reliability and sharpness comparison of different models at the 7 sites	70
4.15	24HA reliability and sharpness comparison of different models at the 7 sites . . . . .	71
4.15	(continued) 24HA reliability and sharpness comparison of different models at the 7 sites	72
5.1	Overall framework of the conditional probabilistic aggregated wind power forecasting .	75
5.2	Distributions of wind power forecasts at the S4 site . . . . .	85
5.3	Locations of clustered wind farms . . . . .	87
5.4	Clustering analysis . . . . .	88
5.5	Day-ahead aggregated probabilistic wind power forecasts . . . . .	88
5.5	(continued) Day-ahead aggregated probabilistic wind power forecasts . . . . .	89
5.6	Reliability comparison of different models with different look-ahead times . . . . .	91
5.6	(continued) Reliability comparison of different models with different look-ahead times	92
5.7	Sharpness comparison of different models with different look-ahead times . . . . .	93
5.7	(continued) Sharpness comparison of different models with different look-ahead times	94
6.1	Overall framework of weather scenario generation-based probabilistic solar power forecasting . . . . .	97
6.2	Data splitting for model training and testing . . . . .	105
6.3	1HA probabilistic solar power forecasts . . . . .	108
6.3	(continued) 1HA probabilistic solar power forecasts . . . . .	109
6.4	Reliability comparison of different models at different sites . . . . .	111
6.4	(continued) Reliability comparison of different models at different sites . . . . .	112
6.5	Sharpness comparison of different models at different sites . . . . .	113
6.5	(continued) Sharpness comparison of different models at different sites . . . . .	114
7.1	Overview of data integrity attacks in solar power forecasting . . . . .	117
7.2	Spatial solar power sequence . . . . .	118
7.3	Spatial-temporal solar power sequence . . . . .	119
7.4	Overall framework of CNN-LSTM-based probabilistic anomaly detection . . . . .	122
7.5	Probabilistic anomaly detection on different attack templates . . . . .	133

## LIST OF TABLES

3.1	Data duration at selected sites . . . . .	31
3.2	Deterministic forecasting results using M3 and PS . . . . .	32
3.3	NMAE and NRMSE of the SVR surrogate model . . . . .	32
3.4	Normalized optimal averaged sum of pinball loss . . . . .	33
3.5	One-hour-ahead forecasting NMAE and NPL of single-algorithm models with different kernels . . . . .	40
3.6	R-square of the least square fit . . . . .	40
4.1	Data summary of the selected 7 WIND Toolkit sites . . . . .	54
4.2	Deterministic forecasting results by using Q-learning and NWP . . . . .	56
5.1	Data summary of the selected 9 WIND Toolkit sites . . . . .	83
5.2	Deterministic forecasting results by using Q-learning and NWP . . . . .	84
5.3	Information criteria of the estimated distribution . . . . .	85
5.4	Normalized pinball loss of different models with different look-ahead hours . . . . .	90
6.1	Data summary of the selected 7 solar sites . . . . .	104
6.2	1HA deterministic weather forecasting results by using M3 and PS . . . . .	106
6.3	Normalized pinball loss and relative improvement of different models . . . . .	110
7.1	1HA deterministic solar power forecasting results by using CNN-LSTM and PS . . . . .	128
7.2	Normalized NPL-A under different attack templates . . . . .	130
7.3	Contingency table of attack detection . . . . .	130
7.4	Probabilistic Anomaly Detection Results . . . . .	131

# CHAPTER 1

## INTRODUCTION

The penetration of renewable energy, especially wind and solar energy, has been rising dramatically over the past years. With 60 GW annual installations and 591 GW total installations worldwide, wind power contributed nearly a half of renewable power capacity (not including hydro power) by 2018. Figure 1.1 shows the wind power installation capacity around the world. Several countries, such as China and the US, set ambitious policy targets for the wind power development. The global weighted average cost of electricity falls to 0.049 USD/kWh for onshore wind, which makes it a competitive energy source in the electricity markets (IRENA, 2019a). According to REN21, at least 12 countries in the world met at least 10% of their annual electricity consumption with wind energy in 2018 (IRENA, 2019b). For solar energy, at the end of 2018, the total installed capacity of solar power in the US amounts to 32,239 MW (including both solar PV and solar thermal) and it supplies over 2% of total electricity consumption in the US. In addition, U.S. EIA (2018) estimates that there is 19,547 MW of small scale solar PV installed. As the two fastest growing renewable energy sources, the increasing penetration of wind and solar power has profound impact on the grid reliability and security.

The uncertain and variable nature of renewable energy makes it challenging to be integrated into power systems, particularly at ever-increasing level of renewable energy penetration. Variability in generation sources can require additional actions to balance the system. Greater flexibility in the system will be needed to accommodate supply-side variability and the relationship between generation levels and loads. To balance the electricity supply and demand, accurate renewable energy forecasting has been studied in unit commitment and economic dispatch decision-making to provide dynamic operating reserves, which could provide benefits to system operators and electricity traders. For example, most of the independent system operators (ISO) in the U.S. have adopted wind and solar power forecasting to assist their system operations (Cibulka et al., 2012). ISO New England utilizes day-ahead forecasts for the dispatch scheduling of generating capacity, reliability

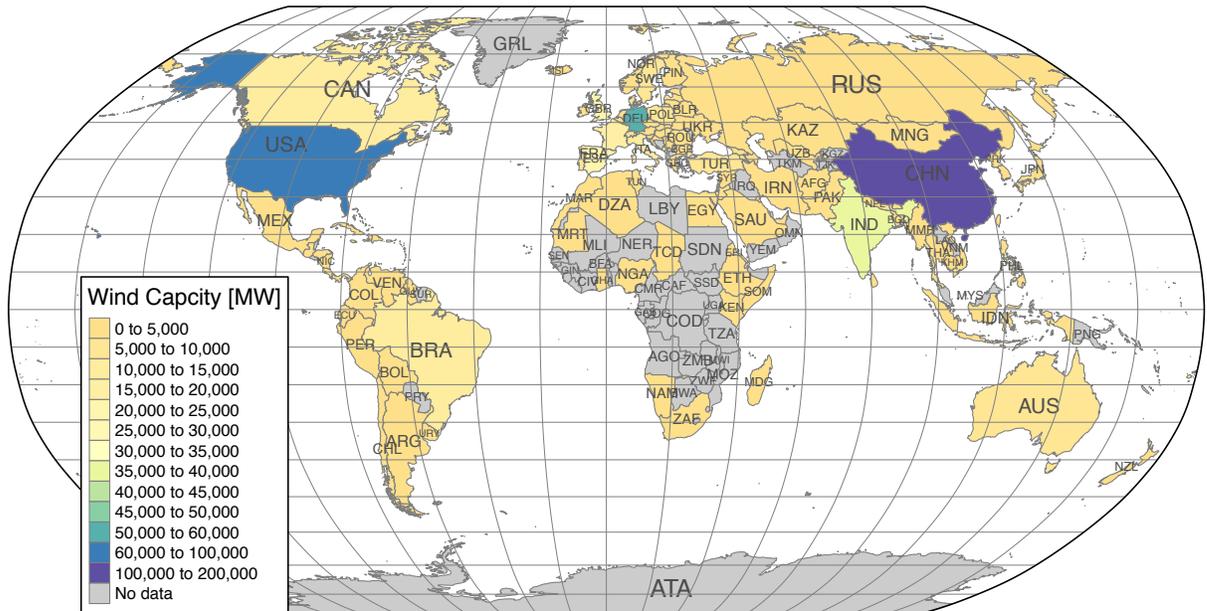


Figure 1.1: Wind power installation capacity by country (by 2018).

analysis, and maintenance planning for the generators (Henderson et al., 2011). A number of forecasting projects have been or is being conducted to promote renewable energy forecasting, such as wind forecast improvement project (WFIP) (Wilczak et al., 2015), WindView (Gamble et al., 2018), Watt-Sun (Hamann, 2017), and SUMMER-GO (Jascourt et al., 2019).

Renewable energy forecasts can help reduce the amount of operating reserves needed for the system, reducing costs of balancing the system. However, conventional deterministic forecasts might not be sufficient to characterize the inherent uncertainty of renewable energy. Probabilistic forecasts that provide quantitative uncertainty information associated with renewable energy are therefore expected to better assist power system operations. Studies have shown that the geographically dispersed renewable energy farms are spatially correlated, and the integration of geographically dispersed renewable energy farms could reduce extreme power output, which is referred to as smoothing effect (Tastu et al., 2011). In addition, power produced from one renewable energy farm at different times is typically temporally correlated (Malvaldi et al., 2017). It would be in-

teresting to explore the impacts of spatial-temporal correlation on the performance of renewable energy forecasting.

Forecasting target variables in renewable energy systems include wind speed, wind generation, solar generation, geothermal heat generation, and tide generation. In this study, we focus on improving probabilistic wind and solar forecasting, by exploring the spatial-temporal correlation modeling among renewable energy plants, and utilizing probabilistic forecasting for anomaly detection in renewable energy farms. The hypothesis is that the spatial-temporal modeling among renewable energy farms could improve probabilistic renewable energy forecasting accuracy, and also help detect anomaly events in renewable forecasts

The goal of this research is to mitigate power system uncertainty by improving probabilistic renewable energy forecasting accuracy utilizing spatial-temporal correlation modeling, and to reduce potential malicious cyber attacks in power system through probabilistic anomaly detection. This dissertation is composed of three major research thrusts:

1. Thrust I: Improving probabilistic renewable energy forecasting via predictive distribution optimization and ensemble learning
  - (a) Chapter 3: Improving probabilistic forecasting via predictive distribution optimization
  - (b) Chapter 4: Improving probabilistic forecasting via ensemble learning
2. Thrust II: Improving probabilistic renewable energy forecasting via spatial-temporal correlation and scenario generation;
  - (a) Chapter 5: Improving probabilistic renewable energy forecasting via spatial-temporal correlation modeling
  - (b) Chapter 6: Improving probabilistic forecasting via scenario generation
3. Thrust III: Utilizing probabilistic forecasting for anomaly detection in renewable energy systems (Chapter 7)

There are a variety of ways to categorize probabilistic renewable energy forecasting. For example, based on forecasting time horizon, probabilistic renewable energy forecasting can be grouped into very short-term, short-term, mid-term, and long-term forecasting. Based on the algorithm principles, probabilistic forecasting techniques can be divided into parametric methods, non-parametric methods, and hybrid methods. For all the different categories mentioned above, spatial-temporal modeling could be integrated to improve the probabilistic forecasting accuracy. In addition, spatial-temporal analysis and probabilistic forecasting methods could be used for anomaly detection in renewable energy forecasting.

## **CHAPTER 2**

### **LITERATURE REVIEW**

Different from deterministic forecasting, probabilistic wind forecasting could quantitatively describe the uncertainty information lies within the renewable energy, which helps reliable and economic power system operations and planning. The probabilistic renewable energy forecasts are only used in a primitive way in most systems, and there is not a systematic way to integrate them into system operation and scheduling routines. Methods of probabilistic forecasting could be divided to different categories in terms of forecasting horizon or mathematical methodology. Based on forecasting horizon, probabilistic wind forecasting can be classified to very short-term, short-term, medium-term, and long-term forecasting. Based on mathematical methodology, probabilistic wind forecasting can be classified to parametric methods and non-parametric methods. In this chapter, we first give a comprehensive review on the state-of-the-art probabilistic renewable energy forecasting methods in the literature. To further analyze the probabilistic renewable energy forecasting techniques, state-of-the-art spatial-temporal modeling approaches in renewable energy are reviewed. Finally, we give a comprehensive review on the application of probabilistic renewable energy forecasting on cyberattacks and anomaly detection.

#### **2.1 Probabilistic Renewable Energy Forecasting**

##### **2.1.1 Review of Probabilistic Renewable Energy Forecasting Approaches**

Probabilistic renewable energy forecasts usually take the form of probability distributions associated with point forecasts, namely, the expectation. Existing methods of constructing predictive distributions can be mainly classified into parametric and non-parametric approaches in terms of distribution shape assumptions (Lefèvre et al., 2014). A prior assumption of the predictive distribution shape is made in parametric methods, and unknown distribution parameters are estimated based on historical data. Parametric approaches generally require low computational cost. There

are limited number of papers on probabilistic renewable energy forecasting based on parametric method in the past 5 years, and most of them adopt mixture distributions (i.e., GMM, GGMM, and bi-variate distribution) due to their capability of modeling irregular residuals. However mixture distributions are always constrained by the unity integration of each member model and the summation of weight must strictly equal unity. These constrains may limit the robustness of mixture distributions in practical use. The heyday of parametric methods for probabilistic renewable energy forecasting is around early 2010s when the probabilistic forecasting methods start attracting more researchers' attention. In the early stage, Gaussian (Lange, 2005) and beta (Bludszuweit et al., 2008) distributions are two commonly used predictive distributions in probabilistic renewable energy forecasting. However, Gaussian and beta distributions could not capture the fat tails and double bounded properties of wind power distribution. To better account for the nonlinear and double-bounded properties of renewable energy generation in short-term probabilistic forecasting, Pinson et al. (Pinson, 2012) proposed a generalized logit-normal (GL-normal) distribution, which could be used for minutes- to days- ahead probabilistic forecasting. However, the characteristics of the GL-normal distribution ignores the stochastic process. Furthermore, when estimating the GL-normal distribution parameters, noise will be unavoidably introduced to both the deterministic forecasts and the predictor. Some other earlier parametric distributions for probabilistic renewable energy forecasting are the combined Beta distribution and Dirac delta function (Tewari et al., 2011), Levy-stable distribution (Bruninx and Delarue, 2014), versatile distribution with adjustable parameters (Zhang et al., 2013), and time-varying Gamma-like distribution (Menemenlis et al., 2012), etc.

Once an analytical form of the predictive distribution is defined, distribution parameters can be estimated by using different methods. Generally, a predictive distribution could be characterized by location parameter and scale parameter, namely mean  $\mu$  and standard deviation  $\sigma$  (Sun et al., 2018). The deterministic renewable energy forecasts are usually treated as the estimation of location parameters (Sun et al., 2019). For estimators of location parameters, non-linear time

series is one of the most popular categories of methods. For example, Pinson et al. (Pinson, 2012) developed a conditional parametric auto-regression model to estimate the parameters of a GL-normal distribution, which is a discrete-continuous mixture of the GL-normal distribution and two probability masses. For estimators of scale parameter, autoregression-generalized autoregressive with conditional heteroscedasticity models are one of the most popular used methods (Meitz and Saikkonen, 2011). Other methods to estimate distribution parameters in probabilistic renewable energy forecasts include maximum likelihood (Delignette-Muller et al., 2015), least squares (Kantar, 2015), method of moments (Hall, 2005), and the fast Bayesian approach (Jin, 2008). Overall, none of the distribution parameter estimation methods mentioned above aim at optimizing the probabilistic forecasting metrics (e.g., pinball loss). In summary, parametric method is more competitive for very short-term renewable energy forecasting. Moreover, it has low computational costs and requires less samples. However, finding a suitable predictive distribution shape for probabilistic renewable energy forecasting is still an open topic.

In addition to parametric approach, non-parametric approach is another way to provide probabilistic renewable energy forecasts. Instead of assuming a predictive distribution, the quantiles are estimated through a finite number of observations. In the past five years, there are more than hundred probabilistic wind forecasting papers based on non-parametric methods, which dominates the probabilistic wind forecasting literature. From the statistic summary of GEFCOM2014, the methods used by the top five teams in the probabilistic wind forecasting track are all non-parametric methods (Hong and Fan, 2016), which show the effectiveness of non-parametric methods. Three representative categories of non-parametric method: kernel density estimation (KDE)-based method, quantile regression (QR)-based method, and artificial intelligence (AI)-based ensemble method are further reviewed.

KDE is one of the most popular used methods in probabilistic renewable energy forecasting since it can provide the full probability density function of different shapes. Based on finite number of samples from an unknown density function  $f$ , the univariate KDE is given by (Sheather and

Jones, 1991):

$$\hat{f}(x) = \frac{1}{N \cdot h} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right) \quad (2.1)$$

where  $h$  is the bandwidth,  $K(\cdot)$  is kernel function,  $X_i$  is the sample point, and  $N$  is the sample size. Some of the KDE-based methods are focused on tuning and optimizing the parameters of KDE. It is worthwhile to mention that the renewable energy distribution is heavily skewed and heavily tailed due to the nonlinear conversion. Therefore, due to its non-Gaussian property, when modeling renewable energy distributions, KDE with a small bandwidth may underestimate the tail while the KDE with a larger bandwidth may overestimate the mode (Zhang et al., 2015). To overcome this difficulty, a bandwidth selection process could be adopted to adaptively select the optimal bandwidth (El-Dakkak et al., 2019).

In addition to bandwidth, another important parameter in KDE is kernel. Based on the shape of random variable, a suitable kernel could be chosen. For wind speed and wind power, Gamma kernel and Beta kernel are two most popular used kernels in the literature, respectively (Bessa et al., 2012). In summary, KDE-based non-parametric method is competitive for wind forecasting at different horizons due to its simplicity. Moreover, KDE could be used together with multiple advanced deterministic wind forecasting methods to generate probabilistic forecasts, which further improve the forecast accuracy. However, finding optimal bandwidth and kernels is still a hot topic.

QR is one of the traditional non-parametric probabilistic forecasting methods (Zhang et al., 2014). Instead of estimating models for conditional mean functions like traditional linear regression methods through minimizing the sums of squared residuals, QR estimates models for both the conditional median function and the full range of other conditional quantile functions. In the literature, QR has been integrated with different machine learning and deep learning models to provide accurate probabilistic forecasts (Wang et al., 2019). For example, quantile regression neural network (QRNN) combines a nonlinear relation between input and output data and the pinball loss function to obtain the conditional quantiles through NN in order to incorporate nonlinearities. QR could be integrated with decision tree-based models as well. For example, random forest

(RF) builds several trees and combines their output by averaging each tree leaf in the forest, which helps to improve the generalization ability of the model. In quantile regression forests (QRF), all outcomes are stored, thus the quantiles from each tree leaf can be calculated to form probabilistic forecasts. In addition to decision trees, QR could be integrated with gradient boosting model (Wang et al., 2018), kernel functions (Takeuchi et al., 2009), and deep learning models. However, the widely used QR is a direct function of the point forecast and predictors, and it can only provide the range of the given percentage (Lee and Baldick, 2016). Haben et al. (Haben and Giasemidis, 2016) developed a non-parametric hybrid method that combines KDE and QR to generate probabilistic forecasts. Ordiano et al. (Ordiano et al., 2017) conducted probabilistic forecasting using a nearest-neighbor-based non-parametric method. Another non-parametric probabilistic wind forecasting method is non-parametric Dirichlet process mixture model (DPMM). DPMM has no prior knowledge on the distribution, and researches found that it can generate more suitable residual distribution than the classical Gaussian distribution (Wang et al., 2017). Most existing non-parametric probabilistic wind forecasting approaches focus on statistical methods. In addition to traditional statistical approaches, the performance of probabilistic wind forecasting can be further improved by machine learning techniques. For example, Wan et al. (Wan et al., 2014) used an extreme learning machine to predict the optimal prediction interval without using statistical inferences and distribution assumptions. In the Global Energy Forecasting Competition 2014 (GEFCom2014), Landry et al. (Landry et al., 2016) used gradient-boosted machines (GBM) for multiple quantile regression to fit each quantile and zone independently and generate probabilistic forecasts. Zhang et al. (Zhang and Wang, 2015) developed a probabilistic forecasting method based on k-nearest neighbor point forecasts through KDE. Wang et al. (Wang et al., 2017) used deep convolutional neural network and wavelet transform to quantify the renewable energy uncertainties with respect to model misspecification and data noise.

Ensemble methods have shown to be able to improve the performance of probabilistic renewable energy forecasting. Methods of constructing ensemble forecasts can be mainly classified into

two groups: competitive ensemble methods and cooperative ensemble methods (Opitz and Maclin, 1999). *Competitive ensemble methods* use induction algorithms with different parameters or initial conditions to build individual forecasting models. The final refined ensemble prediction is obtained from pruning and aggregating individual forecasts. *Competitive ensemble methods* generally require a large diversity of data and parameters to obtain different forecasts from individual predictors (Ren et al., 2015). Therefore, competitive ensemble methods usually require expensive computation cost, and they are usually used in medium-term and long-term forecasting. Bagging (Breiman, 1996) and boosting (Schapire, 1990) are two commonly used competitive ensemble methods. To better account for the performance of weak models, ensemble forecasting approaches based on adaptive boosting (i.e., assign large weights to the models with larger errors) are used in (Shrestha and Solomatine, 2006; Ren et al., 2014; Wu et al., 2012). For *cooperative ensemble methods*, the dataset is divided into several sub-datasets and each sub-dataset is forecasted separately, and the final forecasts are obtained by aggregating all the sub-forecasts. Cooperative ensemble methods usually have lower computation burden due to less parameter tuning work, which are normally used in very short-term or short-term forecasting (Ren et al., 2015). Artificial neural networks based autoregressive integrated moving average (Liu et al., 2012) and generalized autoregressive conditional heteroskedasticity based autoregressive integrated moving average (Lojowska et al., 2010) are two commonly used cooperative ensemble methods that combine suitable models for linear and non-linear time series.

A few recent studies have been done in the literature on ensemble probabilistic renewable energy forecasting. In (Zhang et al., 2014), Zhang et al. proposed an ensemble probabilistic wind power forecasting approach based on empirical mode decomposition, sample entropy techniques and extreme learning machine. Lin et al. (Lin et al., 2018) combined multiple probabilistic forecasting models based on sparse Bayesian learning, kernel density estimation, and beta distribution estimation. The weight parameters of the multi-model ensemble are solved by an expectation maximizing algorithm and continuous ranked probability score optimization. Kim et al. (Kim

and Hur, 2018) developed an enhanced ensemble method for probabilistic wind power forecasting. The wind speed spatial ensemble was built by using correlation based weight and kriging models, and the temporal ensemble was built through an average ensemble of three models (i.e., an exogenous variable model, a polynomial regression model, and an analog ensemble model). Wang et al. (Wang et al., 2017) used wavelet transform to distinguish the non-linear series, and an ensemble technique was used to cancel out the diverse errors of point forecasts. Then, the probabilistic forecasts were generated by using a convolutional neural network. Overall, AI-based ensemble methods have shown to be able to improve the performance of probabilistic renewable energy forecasting.

In summary, non-parametric method is distribution free, and it is more competitive for short- and medium- term wind forecasting. However, it requires high computational costs due to the complicated estimation and larger sample size. non-parametric forecasts, particularly AI-based ensemble methods, dominate the studies of probabilistic renewable energy forecasting. Therefore, there is a clear gap between parametric methods and AI-based ensemble methods. Particularly, future studies should bridge the gap by placing more emphasis on the combined use of parametric and non-parametric probabilistic renewable energy forecasting.

### **2.1.2 Evaluation Metrics**

Evaluation of forecasting performance is of great importance for model development and quality control. It is possible your forecasts perform well with the measurement from one evaluation metric, but may perform poorly with the measurement from other evaluation metrics. Unfortunately, there is no universal forecast evaluation metrics ensuring a global optimal for both the preference of end users and other objectives. Therefore, it is of great importance to use varied evaluation metrics to evaluate your model to achieve particular challenges. Different from deterministic forecasts, probabilistic forecasts usually take the form of prediction intervals or scenarios. To some extent,

the evaluation of probabilistic forecasts will be more challenging than that of deterministic forecasts since you are evaluating some specific or all the quantiles. Generally, probabilistic forecasts can be evaluated from three single aspects: reliability, sharpness, and resolution (Pinson et al., 2007). However, in probabilistic wind forecasting, the dependencies of wind power on explanatory variables are captured through two or more aspects together (e.g., sharpness and resolution). To this end, in addition to the aforementioned three single probabilistic forecasting evaluation aspects, a comprehensive evaluation index, skill score is proposed. In this section, we list a number of popular used probabilistic forecasting evaluation metrics.

### **Reliability**

Reliability (RE) stands for the correctness of a probabilistic forecast that matches the observation frequencies (Juban et al., 2007):

$$RE = \left[ \frac{\xi^{(1-\alpha)}}{N} - (1 - \alpha) \right] \times 100\% \quad (2.2)$$

where  $N$  is the number of test samples, and  $\xi^{(1-\alpha)}$  is the number of times that the actual test samples lie within the  $\alpha$ th prediction interval. The reliability of probabilistic forecasts could be measured by the reliability diagram, verification rank histogram, probability integral transform (PIT) histogram, and quantile-quantile (Q-Q) diagram. Since reliability of probabilistic forecast is an agreement between forecast probability and mean observed frequency, reliability is the most important property of probabilistic forecast, which should be checked first. Unreliable probabilistic forecasts could lead to erroneous power system operation, which may cause power system collapse and black out.

### **Reliability diagram**

With measured empirical coverages, a reliability diagram can be plotted to describe the quantile forecast series with different nominal proportions. A reliability diagram shows whether a given

method tends to systematically underestimate or overestimate the uncertainty. The nominal coverage rate ranges from 10% to 90% with a 10% increment. Fig. 2.1 shows an example reliability diagram of probabilistic wind power forecast. In this example, the reliability diagram is close to the diagonal. Generally, a forecast presents better reliability when the reliability curve is closer to the diagonal. To be more specific, if the curve lies below the diagonal, the probabilities are overestimated. Similarly, if the curve lies above the diagonal, the probabilities are underestimated.

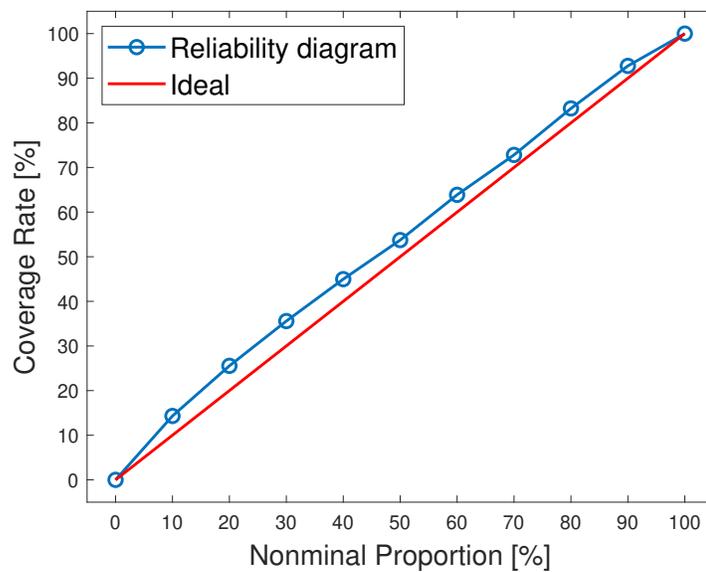


Figure 2.1: An example reliability diagram of probabilistic wind power forecasts

### Verification rank histogram and probability integral transform (PIT) histogram

The verification rank histogram is used to examine the frequencies the observations falls within different prediction intervals constructed by the empirical quantiles of probabilistic forecasts. Different from verification rank histogram, PIT histograms are continuous form of verification histogram, which measure the observations over the value of the predictive CDF. Fig. 2.2 shows an example PIT histogram of probabilistic wind power forecast. In this example, the red dash line indicates the ideal reliability. Overall, the PIT histogram is close to the red dash line, and nearly

uniform distributed. Generally, a forecast presents better reliability when the verification rank histogram or PIT histogram is more flat or uniform distributed.

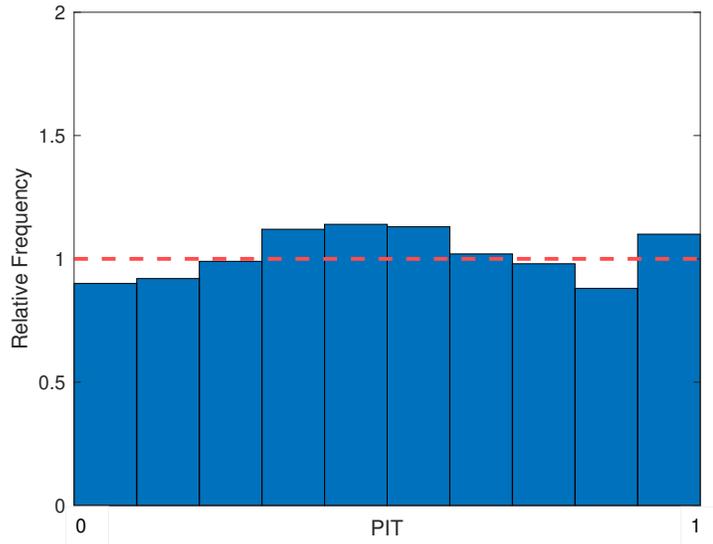


Figure 2.2: Example PIT histogram of probabilistic wind power forecasts

### Quantile-quantile (Q-Q) plot

In statistics, Q-Q plot is a probability plot, which compares two probability distributions by plotting their quantiles against each other. A forecast is considered reliable if it is statistically consistent with the observed uncertainty. The x-axis and y-axis of Q-Q plot are predictive quantiles and empirical quantiles, respectively. In general, probabilistic forecast presents better reliability when the Q-Q plot is closer to diagonal.

### Sharpness

Sharpness indicates the capacity of a forecasting system to forecast wind power with extreme probability (Gallego-Castillo et al., 2016). To be more specific, it measures the dispersion of the

predictive distributions, and this criterion evaluates the predictions independently of the observations, which gives an indication of the level of usefulness of the predictions. For example, a system that provides only uniformly distributed predictions is less useful for decision-making under uncertainty. Predictions with perfect sharpness are discrete predictions with a probability of one (i.e., deterministic predictions). The sharpness of probabilistic forecasts could be measured by the box plot of the quantiles of the prediction interval width distribution (Shin et al., 2020) and  $\delta$  diagram.

### Box plot of prediction interval

If probabilistic wind power forecasting possesses high sharpness, quantiles of the prediction interval width distribution should be as concentrated as possible. A box plot of the prediction intervals could be adopted to illustrate the sharpness. Fig. 2.3 shows an example box plot of the quantiles of the wind power prediction interval. From the figure, the wind power predictive distributions of each hour are characterized by the 5, 25, 50, 75 and 95 quantiles (from bottom to top).

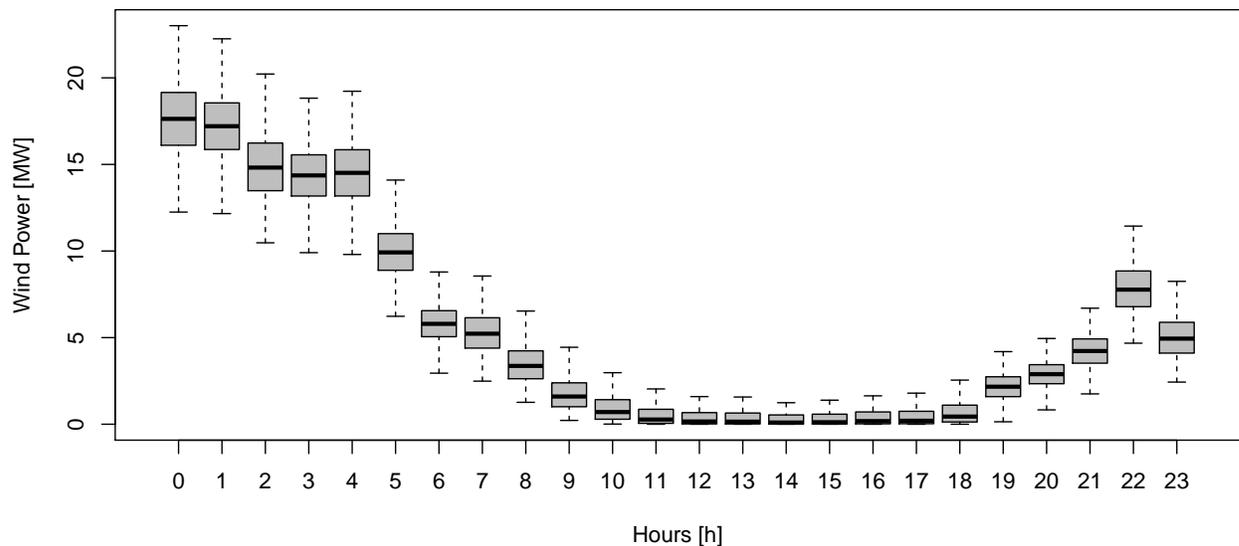


Figure 2.3: An example boxplot of probabilistic wind power prediction interval

## $\delta$ diagram

In addition to the box plot, the sharpness could be measured through the average size of the predictive intervals through  $\delta$  diagram as well. An example of the  $\delta$  diagram is shown in Fig. 2.4. Generally, the expected interval size increases with increasing nominal coverage rate, and the lower value indicates better sharpness.

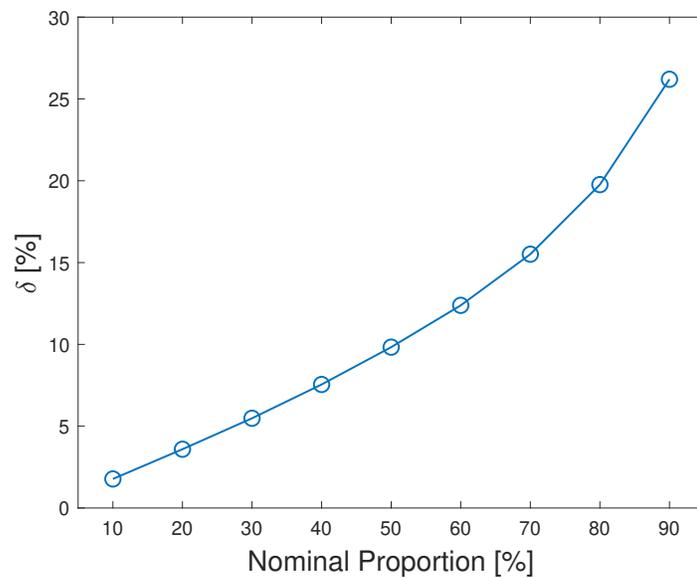


Figure 2.4: An example  $\delta$  diagram of probabilistic wind power forecasts

## Resolution

Resolution measures how well the probabilistic forecasts distinguish situations with different frequencies of occurrence. A probabilistic forecast with good resolution should have the ability to resolve the set of sample events into subsets with characteristically different outcomes, i.e., a larger resolution.

## **Skill scores**

It is possible some aspects of probabilistic forecasts are contradictory to each other. For example, a forecast with very large prediction intervals may have a bad sharpness and good reliability. Moreover, in decision makings, different quantiles may be evaluated together or separately, i.e., evaluate the sharpness and resolution together. Therefore, a comprehensive skill score is needed based on different scoring rules. Brier score (BS), continuous ranked probability score (CRPS), pinball loss, ignorance score, and Winkler score are popular used skill scores in the literature.

### **Brier score**

Proposed by Glenn W. Brier, Brier score measures the mean squared difference between the predicted probability and the actual outcome for binary or categorical events (Brier, 1950), and could be expressed as:

$$BS = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2 \quad (2.3)$$

where  $f_t$  denotes the forecast probability at time  $t$ ,  $o_t$  denotes the actual outcome at time  $t$ , and  $N$  is the sample size. Generally, BS takes values between 0 and 1. A smaller BS indicates better probabilistic forecasts. BS could also be decomposed into three components that represent important properties of probabilistic forecasts (Murphy, 1973):

$$BS = RE - RES + UNC \quad (2.4)$$

where RE, RES, and UNC denote reliability, resolution, and uncertainty, respectively.

### **Continuous ranked probability score**

Based on different needs in power system, probabilistic wind power forecasts may be required in discrete form (e.g., quantiles) or continuous form (e.g., probability distribution). For the measurement of continuous variables, CRPS is one of the most popular used comprehensive evaluation

score, which evaluates the quality of predictive CDF and expressed as:

$$CRPS = \frac{1}{N} \sum_{t=1}^N \int_{-\infty}^{+\infty} (F_t(x) - 1(y_t \leq x))^2 dx \quad (2.5)$$

where  $F_t$  is the CDF of the predictive distribution at  $t$ ,  $y_t$  denotes observation, 1 is Heaviside function, and  $N$  is sample size. CRPS is a generalized case of mean absolute error (MAE) in the probabilistic fashion. Therefore, a lower CRPS indicates better probabilistic forecasts.

### Pinball loss

Pinball loss is a widely used metrics to evaluate probabilistic forecasts, which is defined by

$$L_{i,t}(q_{i,t}, x_t) = \begin{cases} (1 - \frac{i}{100}) \times (q_{i,t} - x_t), & x_t < q_{i,t} \\ \frac{i}{100} \times (x_t - q_{i,t}), & x_t \geq q_{i,t} \end{cases} \quad (2.6)$$

where  $x_t$  represents the observation at time  $t$ . For a given  $i$  percentage, the quantile  $q_{i,t}$  represents the value of a random variable whose CDF is  $i$  percentage at time  $t$ . Different from CRPS which focus on the evaluation of predictive distribution as a whole, pinball loss focuses on the quantiles of the predictive distribution. It measures the difference between quantiles and observation. Therefore, a lower pinball loss score indicates better probabilistic forecasts.

### Winkler score

Winkler score is another comprehensive evaluation metric for probabilistic forecasts, it allows a joint measurement of the unconditional coverage and interval width (Liu et al., 2015). The averaged prediction interval (PI) size  $\delta^\beta$  at nominal coverage rate  $(1-\beta)$  can be expresses as:

$$\delta^\beta = \frac{1}{T} \sum_{t=1}^T (q_t^{\alpha_U} - q_t^{\alpha_L}) \times 100\% \quad (2.7)$$

The Winkler score  $W^\beta(x_t)$  is defined to reward narrow PIs and penalize the targets which are out of PIs range. For the  $\beta$ th prediction interval, the Winkler score is expressed as:

$$W^\beta(x_t) = \begin{cases} \delta^\beta, & q_t^{\alpha_L} < x_t < q_t^{\alpha_U} \\ \delta^\beta + 2(q_t^{\alpha_L} - x_t), & x_t < q_t^{\alpha_L} \\ \delta^\beta + 2(x_t - q_t^{\alpha_U}), & x_t > q_t^{\alpha_U} \end{cases} \quad (2.8)$$

where  $q_t^{\alpha_L}$  and  $q_t^{\alpha_U}$  are the lower and upper bound of the  $\beta$ th PI,  $x_t$  is the observation at  $t$ . Note that the penalize coefficient could be modified based on the interest of end-users for different forecasting scenarios. Overall, a lower Winkler score indicates a better prediction interval.

*In Research Thrust I, we seek to (i) improve short-term wind forecasting accuracy by predictive distribution optimization, and (ii) adaptively ensemble multiple possible predictive distributions through competitive and cooperative ensemble strategies.*

## 2.2 Spatial-Temporal Modeling in Renewable Energy

Methods of spatial-temporal correlation based renewable energy forecasting proposed in the recent literature can be generally classified into three groups: (1) Physical models, which are usually developed based on sophisticated meteorological information. For example, Pelikan et al. (Pelikan et al., 2010) developed a robust wind power forecasting model based on a combination of sigmoidal power functions with wind speed forecasts obtained from a mesoscale spatial-temporal refined numerical weather prediction (NWP) model. Physical models usually require high computation cost, and have better performance than purely statistical time series approaches in longer prediction time horizon (e.g., day-ahead and week-ahead) (Feng et al., 2017). (2) Statistical models, which quantify the relationship between geographically dispersed time series wind power. For example, Xie et al. (Xie et al., 2014) developed a short-term wind power forecasting model by leveraging the spatial-temporal correlation in wind speed and direction among different wind farms in west Texas. Statistical models are cost-saving since they do not require any data beyond historical wind power generation. However, the accuracy of statistical models drops with the increment of prediction time horizon. (3) Hybrid models, which combine the advantages of physical and

statistical approaches to obtain globally optimal forecasts. For example, Qin et al. (Qin et al., 2019) proposed a hybrid wind power forecasting model based on long short-term memory network and deep learning neural network. Convolutional network and long short-term memory are used to exploit the spatial-temporal properties of wind farms. More research about spatial-temporal model based wind power forecasting has been studied in (Zhao et al., 2018; Baxevani and Lenzi, 2018; Fang et al., 2018).

In contrast to traditional probabilistic renewable energy forecasting models, probabilistic renewable energy forecasting technologies based on spatial-temporal effects have also been developed in the literature (Zhang et al., 2013). For example, Zhang et al. (Zhang and Wang, 2018) used off-site information of geographically dispersed wind farms to capture spatial-temporal correlation and generated quantile forecasts. Then an Alternating Direction Method of Multipliers (ADMM)-based method was used to generate distributed probabilistic forecasts. Dowell et al. (Dowell and Pinson, 2016) proposed a probabilistic wind power forecasting method and the spatial-temporal correlation was captured through Sparse Vector Autoregression. A study conducted by Tang et al. (Tang et al., 2018) focused on multiple plants' probabilistic power forecasting, and a simple aggregation strategy was used. The marginal distribution of wind power in (Tang et al., 2018) is modeled by Gaussian distribution. Another study conducted by Li et al. (Li et al., 2015) modeled the wind power uncertainties by using a particle filter algorithm. The aggregated wind power predictive distribution is obtained by mesoscale numerical weather prediction model and particle filter. Both (Tang et al., 2018) and (Li et al., 2015) have aggregated all the wind farms in one step without clustering the farms. However, it is important to note that both (Tang et al., 2018) and (Li et al., 2015) have found that considering spatial-temporal correlation among wind farms could improve the performance of probabilistic forecasts. Nevertheless, several challenges present in existing methods: (i) high dimensional matrices are involved in the spatial-temporal modeling, which adds additional computational burden; (ii) communication channels are needed for information transmission in (Zhang and Wang, 2018), which might not be widely applicable at various spatial and temporal

scales; (iii) the single logit-normal and Gaussian distribution in (Dowell and Pinson, 2016) and (Tang et al., 2018) may not be reliable for wind farm data with varying characteristics.

One of the most intuitive ways of modeling spatial-temporal correlation is to use a multivariate joint distribution. For example, Pinson et al. (Pinson et al., 2009) has used a multivariate Gaussian distribution to describe the spatial-temporal relationship between forecasts of different wind farms. To address the challenge of modeling high dimensional multivariate non-Gaussian distributions, the Copula theory can be used. Based on the Sklar’s theorem, the joint distribution can be modeled through univariate marginal-distribution functions and a Copula (Rüschendorf, 2009). Copula theory has been widely used to characterize the dependence between different variables. For example, Tang et al. (Tang et al., 2018) used Copula to model correlation between wind farms and solar farms. Cui et al. (Cui et al., 2018) utilized high-dimensional Copula to couple the wind power ramp features among wind farms.

In the literature, the marginal distribution of wind power is commonly modeled by unimodal distributions such as Beta and Gamma (Louie, 2010) or non-parametric distributions such as KDE (Wang et al., 2018). However, the unimodal distributions may not accurately quantify the variability of wind power and the non-parametric distributions are challenging to be solved analytically (Cui et al., 2018).

*To bridge these gaps, in Research Thrust II, we develop a short-term aggregated probabilistic wind power forecasting methodologies that (i) take spatial-temporal correlation into consideration, (ii) are aware of different correlated weather scenarios, and (iii) provide cost-efficient and widely applicable solutions to distributed energy systems.*

### **2.3 Cyberattacks and Anomaly Detection Approaches in Renewable Energy Forecasting**

Accurate renewable energy forecasting plays an important role in power system planning and renewable energy integration. However, with the rapid development of information and communication technologies in smart grid, utility-scale renewable energy plants are exposed to a larger

cyber attack surface, which may compromise the power devices, leading to local outages, equipment damages, and grid instabilities. For example, the black out at the Ukrainian electricity grid in December 2015 caused by malicious cyberattacks has led to severe economic losses (Liang et al., 2016).

The accuracy of renewable energy forecasting largely depends on the quality of dataset. Incorrect readings or abnormal values may directly lead to defective training models or forecasting results. In recent years, renewable energy forecasting under data integrity attacks has attracted great concerns for power system operators and renewable energy farm owners. Though data quality control has been performed by many renewable energy farm operators, the cyberattacks could be designed sophisticatedly enough to bypass the bad data detection algorithms (Liu et al., 2011). One of the most cited cyber attacks in the literature is the false data injection attack (FDIA). FDIA can be applied to various data interfaces of the renewable energy farm, such as field sensors, data servers in control centers, or remote communication channels. The FDIA requires two assumptions: (1) the attacker has knowledge of the target renewable energy farm, and (2) the attacker can compromise a large number of measurements (Wolf and Serpanos, 2020). For FDIA on renewable energy forecasting, it could be applied to renewable energy measurement itself or measured meteorological data that serves as explanatory variables (Sobhani et al., 2020). The meteorologic data of most renewable energy farms are obtained from external services, which is more vulnerable to malicious cyberattacks (Wang and Govindarasu, 2019). Based on the prospective of the attackers' capability, FDIA could be classified into four types, i.e., scaling attack (Cui et al., 2019), correlated attack (Lore et al., 2018), replay/swap attack, and random attack (Yue, 2017).

Most of the current methods to detect cyberattacks on energy forecasting (e.g., renewable energy, load, and electricity price) are focused on data quality control (Sobhani et al., 2020). Generally, anomaly detection methods in energy forecasting can be classified into three categories: statistical-based methods, clustering-based methods, and classification-based methods (Zheng et al., 2019). For statistical-based methods, statistical inferences are conducted based on fitted error distributions or statistical models to determine if a data point is abnormal or not. A statistical-based

method is straight forward and unsupervised, which can be trained without a full knowledge about the data. However it requires relatively larger samples for distribution estimation and invariably involves high dimensional computation. For example, a descriptive analytics-based anomaly detection method was proposed by Meng et al. (Yue et al., 2019) for load forecasting, which focused on the characteristics of data. For clustering-based methods, we assume the normal data are belong to one group while the anomalies don't belong to any group. Clustering-based methods are unsupervised and all the data are unlabelled. However, clustering-based methods only work when the anomalies don't form groups themselves (Chandola et al., 2009). Zheng et al. (Zheng et al., 2018) used a clustering technique by fast search and find of density peaks (CFSFDP) for power consumption anomaly detection. Different from clustering-based methods, classification-based methods are supervised, and a set of labelled data is used to train the classification model, which is then adopted for anomaly detection. However, the labelled historical scenarios are not always available in real cases, and the length of historical scenarios with attacks are limited as well. Cui et al. (Cui et al., 2019) used a naive Bayes classification method for anomaly detection in load forecasting based on the cumulative distribution function and statistical features of the scaling load data. In addition to the aforementioned three types of methods, some researchers seek to combine different types of anomaly detection methods together to get a better detection outcome. For example, Krishna et al. (Krishna et al., 2015) adopted Principal Component Analysis (PCA) and density-based spatial clustering on noise pattern to detect the anomalies which are deviations from the normal electricity consumption behavior.

Machine learning and deep learning based deterministic forecasting models have been widely adopted in the literature to assist the aforementioned three types of anomaly detection methods due to their strong abilities of modeling the non-linearity within data (Xin et al., 2018). Trained forecasting models are tested on input data and to compare the predicted values with the measurements. A model trained through health data will be able to predict the expected behavior based on history observations. However, there will be a mismatched behavior between the prediction and

observation if the data is under cyberattacks. For example, Wang et al. (Wang et al., 2019) presented a power consumption anomaly detection method based on long short-term memory (LSTM) point forecasts and error patterns. In (Wang and Ahn, 2020), a real-time anomaly detection method was proposed for electric load based on an integration of support vector machine (SVM), k-nearest neighbors (KNN), and a cross-entropy loss function.

Probabilistic renewable energy forecasting is getting more popular in recent years due to its ability to quantitatively characterize the uncertainty in renewable energy generation. However, most of the existing anomaly detection methods mentioned above are deterministic and thus insufficient to characterize the uncertainties in cyberattacks. Probabilistic approaches that provide quantitative uncertainty information associated with cyberattacks are therefore expected to better assist power system operations and renewable energy integration. Probabilistic forecasting usually takes the form of prediction intervals (PIs) or scenarios (Sun et al., 2020a). The PIs of different confidence levels are generally used to assist decision making, which could naturally be used as thresholds to label anomaly events. If a prediction falls outside a desired PI, such prediction may be labeled as anomaly. Probabilistic forecasting methods can be classified into parametric methods and non-parametric methods (Sun et al., 2020b). However, none of these two types of probabilistic forecasting methods has been applied to data integrity anomaly detection in solar power forecasting.

Overall, several challenges present in existing methods for anomaly detection in renewable energy forecasting: (i) most of the existing anomaly detection methods focus on load forecasting and electricity price forecasting, and little work has been done on solar power forecasting and wind power forecasting; (ii) most of the anomaly detection models are deterministic, which could not well describe the uncertainties in solar power; (iii) the spatial-temporal correlation between different targets (e.g., solar farms, wind farms, or load centers) is not considered, which may not be reliable towards sophisticated designed cyberattacks; (iv) most anomaly detection methods could not detect multiple solar farms simultaneously.

To address the aforementioned limitations, *in Research Thrust III, a deep learning-based probabilistic anomaly detection methodology is developed to provide reliable defense strategies against various FDIA scenarios in solar power forecasting.*

## **CHAPTER 3**

### **IMPROVING PROBABILISTIC FORECASTING VIA PREDICTIVE DISTRIBUTION OPTIMIZATION**

Probabilistic wind forecasting usually takes the form of predictive distributions, where the distribution parameters are normally estimated through finite number of observations. As mentioned in Chapter 2.1, the parametric methods in probabilistic renewable energy forecasting is not well studied in the recent literature. Therefore, in this chapter, we present a two-step probabilistic wind speed forecasting methods based on predictive distribution optimization. First, a multi-model machine learning-based ensemble deterministic forecasting framework is adopted to generate deterministic forecasts. The deterministic forecast is assumed to be the mean value of the predictive distribution at each forecasting time stamp. Then, the optimal unknown parameter (i.e., standard deviation) of the predictive distribution is estimated by a support vector regression surrogate model based on the deterministic forecasting value. Finally, probabilistic forecasts are generated from the predictive distribution. Numerical results of case studies at eight locations show that the developed two-step probabilistic forecasting methodology has improved the pinball loss metric score by up to 35% compared to a baseline quantile regression forecasting model.

#### **3.1 Motivation and Objective**

Wind energy is a sustainable alternative to the conventional energy in relieving global warming and fuel energy shortage. Notable progress has been made in increasing the wind energy capacity. However, the uncertain and variable characteristics of the wind resource present challenges to wind integration especially at large penetrations. Accurately forecasting of the wind power generation and the extreme wind power changes would greatly help power systems make better operation schedules, thereby improving the system economic and reliability performance. The one advantage of a deterministic forecast is its simplicity. However, point forecast usually has only a small

chance of being correct. Probabilistic forecasts that provide quantitative uncertainty information associated with desired forecasts become extremely important for reliable and economic power system operations and planning with an ever-increasing renewable energy penetration.

Probabilistic wind forecasts usually take the form of probability distributions associated with point forecasts, namely, the expectation. Parametric method is one of the most common used ways to construct predictive distributions (Lefèvre et al., 2014). A prior assumption of the predictive distribution shape is made in parametric methods, and unknown distribution parameters are estimated based on historical data. Parametric approaches generally require low computational cost. Gaussian (Lange, 2005) and beta (Bludszuweit et al., 2008) distributions are two commonly used predictive distributions in probabilistic wind forecasting. However, Gaussian and beta distributions could not capture the fat tails and double bounded properties of wind energy perfectly. To adaptively optimize the predictive distribution shape and combine the advantages of statistical and machine learning approaches, in this chapter, a two-step probabilistic forecasting method based on pinball loss optimization is developed.

## **3.2 Methodology**

The overall framework of the developed methodology is illustrated in Fig. 3.1. This is a two-step probabilistic forecasting method, consisting of deterministic forecast generation and predictive distribution (type and parameters) determination. In the first step, the machine learning-based multi-model (M3) forecasting framework is adopted to generate short-term deterministic wind forecasts (i.e., 1-hour-ahead), which are considered as a means of predictive distributions. In the second step, a set of optimal standard deviation values are determined through pinball loss optimization at the training stage. The relationship between the deterministic forecasts and the corresponding optimal standard deviations is quantified through a SVR surrogate model. At the forecasting stage, we generate deterministic forecasts first. Then, we feed the deterministic forecasts to the surrogate model built at the training stage to estimate a new set of standard deviation values (pseudo-optimal

standard deviation). Finally, these estimated pseudo-optimal standard deviation values and the deterministic forecasts are used together to generate probabilistic forecasts.

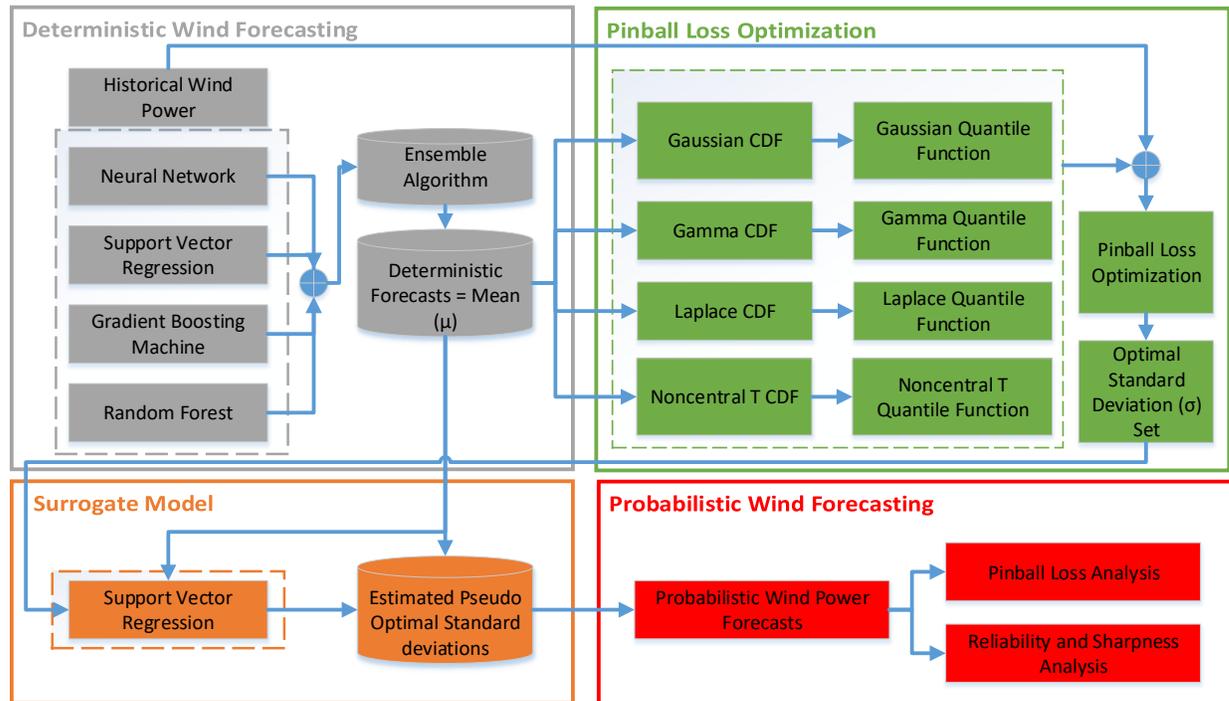


Figure 3.1: Overall framework of the pinball loss optimization-based probabilistic wind forecasting methodology

### 3.2.1 Machine Learning-Based Multimodel (M3) Deterministic Forecasting

M3, a two-layer, short-term forecasting method, is adopted to generate deterministic forecasts. Multiple sets of deterministic forecasts are generated by using different machine learning algorithms with different kernels in the first layer. Then, the forecasts are blended by another machine learning algorithm in the second layer to generate the final forecasts. Machine learning algorithms used in the M3 method include artificial neural networks (ANNs), support vector regression (SVR), GBMs, and random forests. Details about the M3 method can be found in (Feng et al., 2017).

### 3.2.2 Multi-Distribution Database

A multi-distribution database is formulated to model the possible shapes of the predictive distribution. Four widely used predictive distribution types are considered: Gaussian, Gamma, Laplace, and noncentral t distributions. Probability density functions (PDFs) of the four distribution types are summarized in (Forbes et al., 2011). PDFs of the four distributions can be represented by the mean  $\mu$  and standard deviation  $\sigma$  as  $f(x|\mu, \sigma)$ , and the corresponding cumulative distribution functions (CDFs) can be deduced and denoted as  $F(x|\mu, \sigma)$ .

### 3.2.3 Pinball Loss-Based Optimization

Pinball loss is one of the most popular metrics for evaluating probabilistic forecasts (Steinwart et al., 2011); it is a function of observations and quantiles of a forecast distribution. A smaller pinball loss value indicates better probabilistic forecasting. The pinball loss value of a certain quantile  $L_m$  is expressed as:

$$L_m(q_m, x_i) = \begin{cases} (1 - \frac{m}{100}) \times (q_m - x_i), & x_i < q_m \\ \frac{m}{100} \times (x_i - q_m), & x_i \geq q_m \end{cases} \quad (3.1)$$

where  $x_i$  represents the  $i$ th observation,  $m$  represents a quantile percentage from 1 to 99, and  $q_m$  represents the predicted quantile. For a given  $m$  percentage, the quantile  $q_m$  represents the value of a random variable whose CDF is  $m$  percentage. The quantiles of different distributions types are represented by a standard deviation  $\sigma$ , denoted as  $q_m(\sigma)$ . At the offline training stage where  $x$ 's are available, the optimal standard deviation  $\sigma$  is determined by minimizing the pinball loss summation of the 1st to 99th quantiles at each point, which is formulated as follows:

$$\begin{aligned} & \min_{\sigma} \sum_{m=1}^{99} L(q_m(\sigma), x_i) \\ & \text{subject to} \\ & \sigma_l \leq \sigma \leq \sigma_u \end{aligned} \quad (3.2)$$

where  $\sigma_l$  and  $\sigma_u$  represent the lower and upper bounds of the unknown standard deviation, which are selected based on the forecasting target (Hansen et al., 2017). The genetic algorithm (Scrucca et al., 2013) is adopted in this study to solve this optimization problem. In this study, the maximum number of iterations is set to be 100, and the iteration stops if the improvement is less than 0.001. The distribution with the minimum pinball loss is selected as the predictive distribution shape. The optimal standard deviation's estimated using the training data are used to construct a surrogate model to be used at the forecasting stage.

### 3.2.4 Surrogate Model

To generate probabilistic forecasts, a pseudo-optimal standard deviation value is needed at every forecasting time point, which is estimated by a surrogate model. Several possible surrogate model types can be used, such as SVR, radial basis function, kriging, and ANN. SVR is adopted in this study because it is more accurate than other surrogate models in the case studies. The surrogate model is constructed by fitting the optimal standard deviation as a function of the deterministic forecasting value, which is expressed by:

$$\hat{\sigma} = f(x_p) \quad (3.3)$$

where  $x_p$  is a point forecast, and  $f(\cdot)$  is the SVR surrogate model of the optimal standard deviation of the predictive distribution. This surrogate model is used to estimate the standard deviation of the predictive distribution at the online forecasting stage.

## 3.3 Case Studies and Results

The proposed pinball loss optimization-based probabilistic forecasting approach is applied to eight locations to generate wind speed forecasts. The wind speed data were collected near hub height with 1-hour resolution (Zhang et al., 2015). The duration and measurement height of the collected data at all locations are summarized in Table 3.1. For all locations, the first 2/3 of data are used

as training data, in which the first 11/12 is used to train M3 and the remaining 1/12 of the training data is used to build the SVR surrogate model of the optimal standard deviation. The accuracy of the forecasts is evaluated by the remaining 1/3 of data. Although the proposed method is capable of generating forecasts at multiple forecasting horizons, only 1-hour-ahead forecasts are explored in this study.

Table 3.1: Data duration at selected sites

Case No.	Site	Data duration	Height (m)
C1	Boulder_NWTC	2009-01-02 to 2012-12-31	80
C2	Megler	2010-11-03 to 2012-11-01	53.3
C3	CedarCreek_H06	2009-01-02 to 2012-12-31	69
C4	Goodnoe_Hills	2007-01-01 to 2009-12-31	59.4
C5	Bovina50	2010-10-10 to 2012-10-08	50
C6	Bovina100	2010-03-03 to 2012-03-01	100
C7	CapeMay	2007-09-26 to 2009-09-24	100
C8	Cochran	2008-06-30 to 2011-06-29	70

Since the proposed method is a two-step method, it is important to evaluate the accuracy at both steps. In order to evaluate the deterministic forecasting accuracy of the developed framework, four error criteria are utilized: mean absolute error (*MAE*), root mean square error (*RMSE*), and their corresponding normalized indices, i.e., the normalized mean absolute error (*NMAE*) and the normalized root mean square error (*NRMSE*). For these metrics, a smaller value indicates better performance. Deterministic forecasting errors using M3 at the selected locations are summarized in Table 3.2. The persistence (PS) method is used as a baseline, and the forecasting errors are also summarized in Table 3.2.

Overall, the accuracies of the M3 deterministic forecasts are better than those of persistence forecasts, except C3. The smallest *NMAE*, *NRMSE*, and *MAPE* at C3 are produced by the model using the persistence method. This is mainly because C3 has less wind speed variance.

Table 3.2: Deterministic forecasting results using M3 and PS

Method	Metric	Site							
		C1	C2	C3	C4	C5	C6	C7	C8
M3	MAE(m/s)	1.32	0.69	1.26	0.99	1.03	1.10	0.94	0.78
	NMAE(%)	4.77	2.89	3.86	3.72	4.89	4.47	3.45	3.98
	RMSE(m/s)	1.93	0.96	1.78	1.35	1.37	1.53	1.31	1.06
	NRMSE(%)	6.95	3.99	5.48	5.10	6.54	6.22	4.81	5.41
	MAPE	0.40	0.21	0.18	0.19	0.24	0.18	0.27	0.19
PS	MAE(m/s)	1.38	0.84	1.20	1.02	1.10	1.15	0.98	0.85
	NMAE(%)	4.97	3.51	3.69	3.85	5.27	4.69	3.59	4.37
	RMSE(m/s)	2.00	1.17	1.65	1.40	1.48	1.59	1.36	1.16
	NRMSE(%)	7.23	4.87	5.08	5.27	7.04	6.46	4.98	5.92
	MAPE	0.41	0.24	0.17	0.20	0.25	0.20	0.29	0.23

SVR is adopted in this study to build the surrogate model of the optimal predictive distribution parameter (i.e., standard deviation). The metrics of NMAE and NRMSE of surrogate modeling are summarized in Table 3.3. A smaller NMAE/NRMSE value indicates better performance. It is observed that the NMAE and NRMSE are in the range of 5%–9% and 8%–14%, respectively. Overall, the accuracy of the SVR surrogate model is reasonable.

Table 3.3: NMAE and NRMSE of the SVR surrogate model

Metrics	Site							
	C1	C2	C3	C4	C5	C6	C7	C8
NMAE(%)	8.57	5.26	8.64	7.82	6.89	7.50	6.87	7.84
NRMSE(%)	13.48	8.16	13.27	11.51	10.13	11.15	10.64	11.96

Pinball loss values with different predictive distributions are summarized in Table 3.4. The sum of pinball loss is averaged over all quantiles from 1% to 99% and normalized by the maximum wind speed at each site. A lower loss score indicates a better probabilistic forecast. Table 3.4 shows that the M3-Laplace with pinball loss optimization (M3-Laplace) has the smallest pinball loss value at all locations except C3. The smallest pinball loss at C3 is produced by the model using Laplace distribution with the persistence method (PS-Laplace). This is mainly because the persis-

tence deterministic forecasts perform better than the M3 forecasts at this location. The models of quantile regression, PS-Laplace with pinball loss optimization, and M3-Laplace without pinball loss optimization (M3-Laplace-w) are used as benchmark models in case studies. The reasons for choosing these three baselines are: (i) quantile regression is a widely used method in probabilistic forecasts; (ii) the PS-Laplace method allows us to explore the impacts of point forecasts on this two-step probabilistic forecasting framework; (iii) the M3-Laplace-w method allows us to explore the effectiveness of pinball loss optimization. Results show that the M3-Laplace model has improved the pinball loss by up to 35% compared to the three benchmark models and M3 forecasts with other predictive distributions (i.e., M3-Gaussian, M3-Gamma, and M3-Noncentral T (M3-ncT)). Therefore, the Laplace distribution is finally chosen to generate probabilistic wind speed forecasts. Note also that the models of M3-Gaussian, M3-Gamma, and M3-Laplace perform similarly, which indicates that the optimization can help achieve better accuracies with different predictive distribution types. For the baseline model of M3-Laplace-w, a random standard deviation value is selected within the range between the minimum and maximum values of the optimal  $\sigma$ . We repeat this process 30 times to obtain an average sum of pinball loss without optimization.

Table 3.4: Normalized optimal averaged sum of pinball loss

Model	Site							
	C1	C2	C3	C4	C5	C6	C7	C8
QR	2.22	1.76	2.03	1.96	2.56	2.44	1.95	1.68
M3-Gaussian	1.74	1.26	1.44	1.36	1.86	1.69	1.27	1.59
M3-Gamma	1.74	1.26	1.43	1.36	1.87	1.69	1.27	1.58
M3-Laplace	<b>1.72</b>	<b>1.25</b>	1.43	<b>1.35</b>	<b>1.85</b>	<b>1.63</b>	<b>1.26</b>	<b>1.57</b>
M3-ncT	1.74	1.81	2.20	2.21	2.68	3.41	2.56	2.86
M3-Laplace-w	2.94	2.93	2.40	2.39	3.53	3.08	2.46	2.72
PS-Laplace	1.81	1.29	<b>1.34</b>	1.38	1.92	1.70	1.32	1.64

*Note:* The smallest normalized optimal sum of pinball loss at each location is in boldface.

With estimated scale parameters through pinball loss minimization and surrogate modeling, predictive wind speed distributions are determined and the quantiles  $q_1, q_2, \dots, q_{99}$  can be calcu-

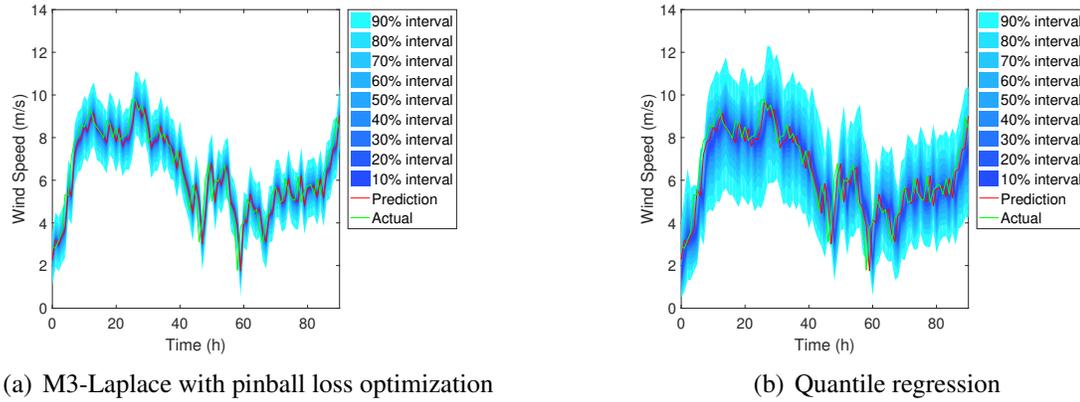


Figure 3.2: M3-Laplace and quantile regression forecasts at the C2 site

lated. To better visualize probabilistic forecasts, the 99 quantiles are converted into nine predictive intervals  $I_\beta$  ( $\beta=10, \dots, 90$ ) in a 10% increment. Fig. 3.2(a) and Fig. 3.2(b) show two examples of probabilistic wind speed forecasts at the C2 site from 2012-02-01 to 2012-02-04. The width of the prediction interval varies with the wind speed variability. When the wind speed fluctuates frequently, the predictive interval tends to be wider, and thereby the uncertainty in wind speed forecasts is relatively higher. Fig. 3.2(b) shows probabilistic forecasts generated from the baseline quantile regression method. The predictive intervals of the proposed M3-Laplace method in Fig. 3.2(a) are **narrower** than those of the quantile regression method. Thus, there is less uncertainty in the M3-Laplace probabilistic wind forecasts.

In addition to pinball loss, two more standard metrics, i.e., reliability and sharpness, are also calculated to assess the performance of the proposed method.

Reliability (RE) stands for the correctness of a probabilistic forecast that matches the observation frequencies (Juban et al., 2007). A reliability plot shows whether a given method tends to systematically underestimate or overestimate the uncertainty. In this study, the nominal coverage rate ranges from 10% to 90% with a 10% increment. Fig. 3.3 shows the reliability plots of the probabilistic forecasts at the eight test sites. A forecast presents better reliability when the curve is closer to the diagonal. Fig. 3.3 shows that overall quantile regression has better reli-

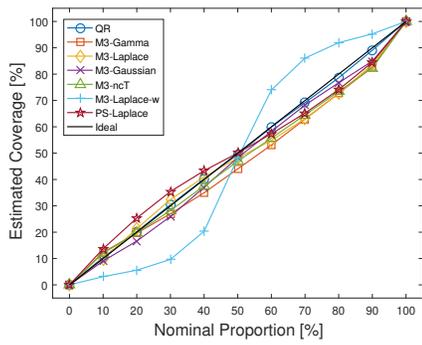
bility performance, because the confidence band of QR is much wider than that of the proposed M3-Laplace method. A wider confidence band indicates that the result takes more errors into consideration; however, note that the reliability over the 90th confidence interval is similar between M3-Laplace and the quantile regression, which is generally more important in probabilistic forecasting applications in power system operations. Also, the M3-Laplace has much better reliability than M3-Laplace-w at all selected locations, which indicates the effectiveness of the pinball loss optimization.

Sharpness indicates the capacity of a forecasting system to forecast wind power with extreme probability (Gallego-Castillo et al., 2016). The sharpness is measured by the average size of the predictive intervals. The sharpness of the proposed M3-Laplace model, quantile regression, M3 with other distribution types, and M3-Laplace-w at the eight sites are compared in Fig. 3.4. The sharpness of the pinball loss-based forecasts is better than that of the baseline quantile regression model. Also, the expected interval size increases with increasing nominal coverage rate, and the M3-Laplace has much better sharpness than that of the M3-Laplace-w. The interval size of the M3-Laplace forecasts ranges from 2% to 18%, which indicates low sharpness.

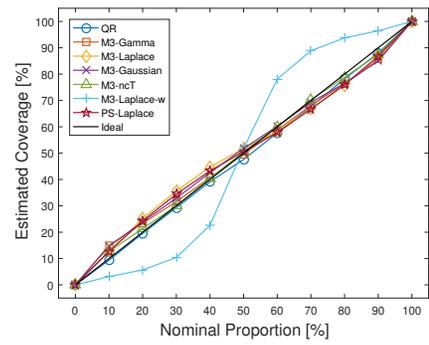
Because the proposed method is a two-step probabilistic forecasting approach, it is interesting to explore the inherent relationship between the first deterministic forecasting step and the second probabilistic forecasting step. To this end, the relationship between a deterministic forecasting metric and a probabilistic forecasting metric is quantified.

NMAE is used to represent the deterministic forecasting accuracy, and normalized pinball loss (NPL) is used to represent the performance of probabilistic forecasts. To generate different NMAE and NPL scenarios, four single machine learning algorithms —i.e., ANN, SVR, GBM, and RF with different kernels are used to produce 14 deterministic forecasts, including:

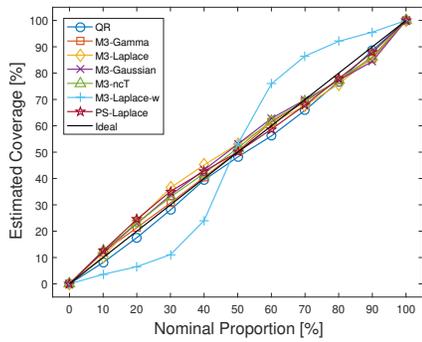
- Three SVR models with linear (SVR\_l), polynomial (SVR\_p), and radial base (SVR\_r) kernels;



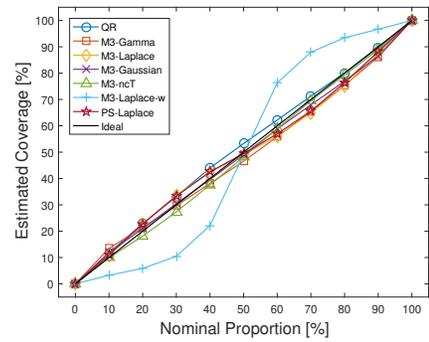
(a) C1



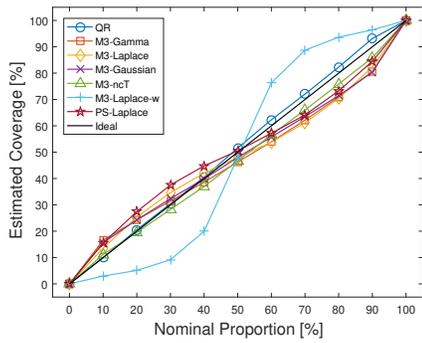
(b) C2



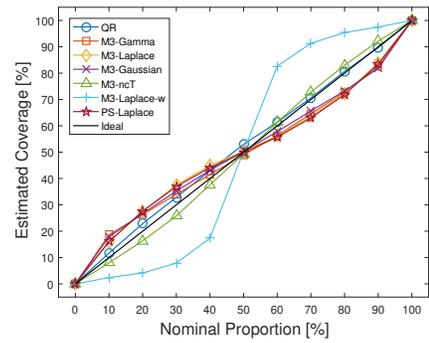
(c) C3



(d) C4

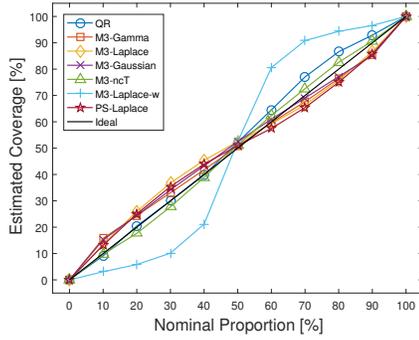


(e) C5

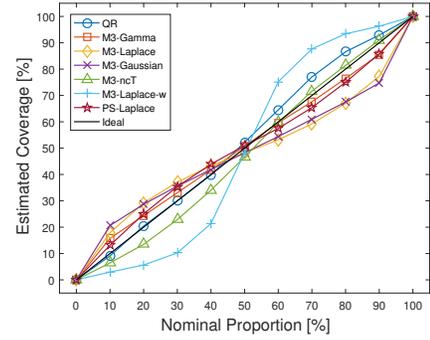


(f) C6

Figure 3.3: Reliability of probabilistic forecasts at selected sites



(g) C7

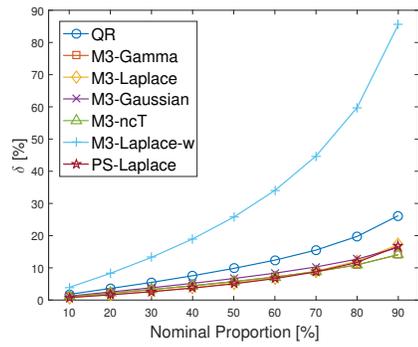


(h) C8

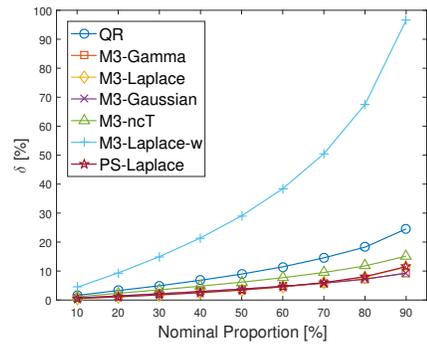
Figure 3.3: (continued) Reliability of probabilistic forecasts at selected sites

- Five ANN models with different numbers of hidden layers ( $n_l$ ), neurons in each layer ( $n_o$ ), and weight decay parameter ( $n_d$ ) values. Our selected models employ the feed-forward back propagation learning function and sigmoid activation function;
- Four GBM models based on different loss functions (Gaussian and Laplacian) and parameters, i.e, number of trees, learning rate ( $\lambda$ ), maximum depth of variable interactions, and minimum number of observations in the terminal nodes;
- Two random forest models with different numbers of variables that are randomly sampled as candidates at each split.

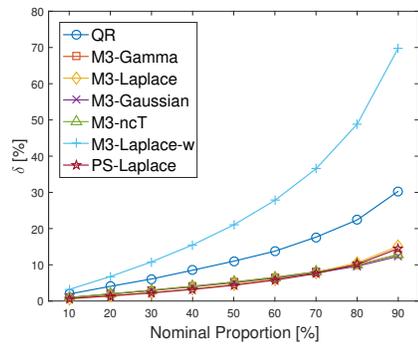
The proposed pinball loss optimization method with Laplace distribution is used to generate probabilistic forecasts. The NMAE values of the 14 deterministic forecasts and their corresponding NPL of probabilistic forecasts are summarized in Table 3.5. A linear regression method (Marquardt, 1963) is used to fit the relationship between NMAE and NPL. Fig. 3.5 shows the relationship between NMAE and NPL at the eight sites. A linear relationship is observed between NMAE and NPL at all test locations, which indicates that a better deterministic forecast model will very likely result in a more accurate probabilistic model with the proposed two-step method. The  $R^2$  values of the eight locations of the linear least squares fit are listed in Table 3.6. The  $R^2$  values are



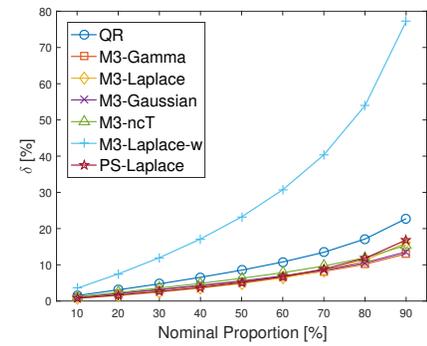
(a) C1



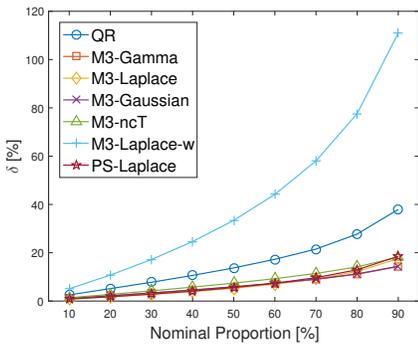
(b) C2



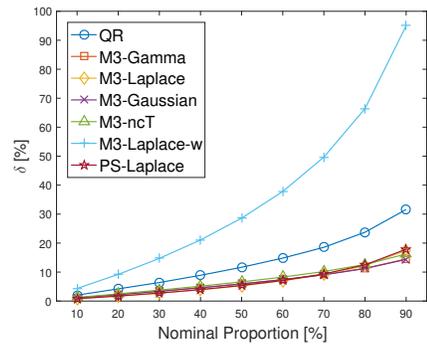
(c) C3



(d) C4

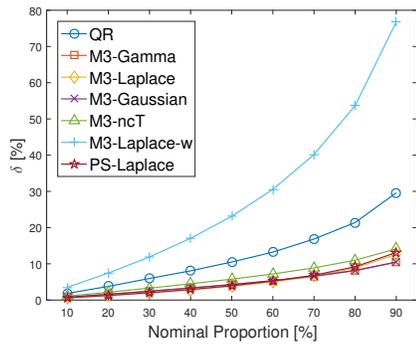


(e) C5

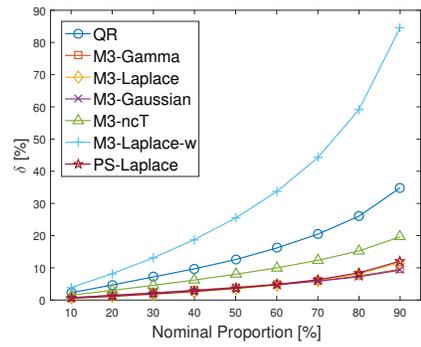


(f) C6

Figure 3.4: Sharpness of probabilistic forecasts at selected sites



(g) C7



(h) C8

Figure 3.4: (continued) Sharpness of probabilistic forecasts at selected sites

close to 1, which also indicates the strong correlation between the deterministic and probabilistic forecasting steps.

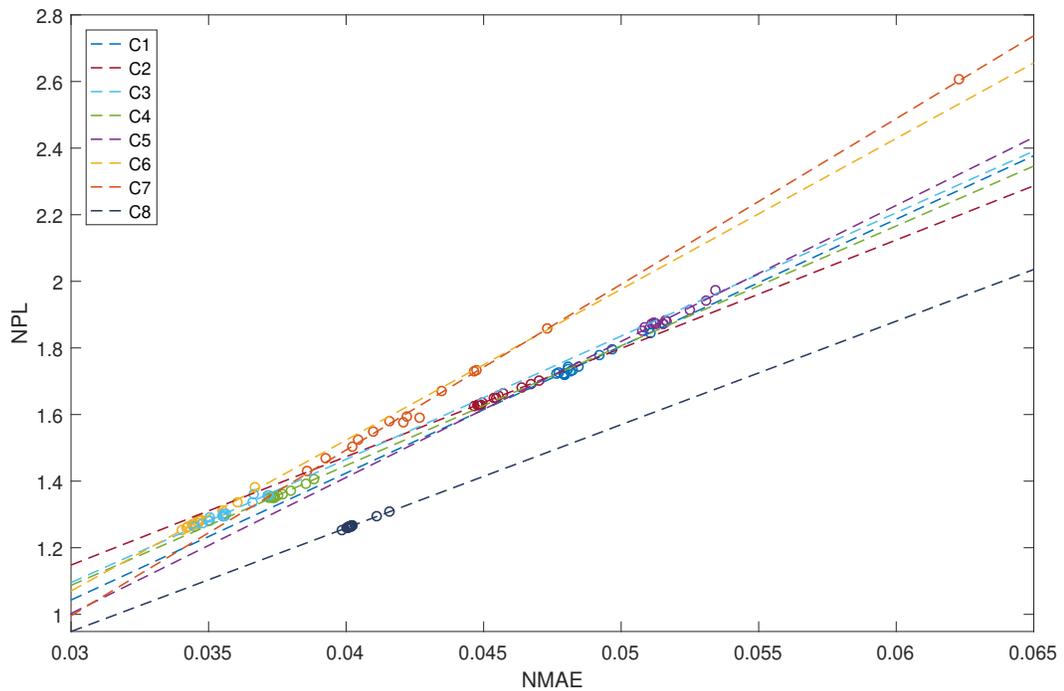


Figure 3.5: The relationship between NMAE and NPL

Table 3.5: One-hour-ahead forecasting NMAE and NPL of single-algorithm models with different kernels

Method	Metrics	Site							
		C1	C2	C3	C4	C5	C6	C7	C8
SVR <sub>r</sub>	NMAE(%)	5.101	3.114	6.229	3.799	5.145	4.554	3.662	4.014
	NPL	1.845	1.381	2.606	1.372	1.974	1.655	1.360	1.974
SVR <sub>l</sub>	NMAE(%)	<b>4.765</b>	<b>2.886</b>	3.927	<b>3.718</b>	<b>4.891</b>	<b>4.466</b>	<b>3.449</b>	<b>3.984</b>
	NPL	1.723	<i>1.253</i>	1.468	1.350	<i>1.853</i>	<i>1.627</i>	<i>1.261</i>	<i>1.853</i>
SVR <sub>p</sub>	NMAE(%)	4.772	2.919	4.267	3.734	4.913	4.572	3.553	4.009
	NPL	1.727	1.270	1.592	1.352	1.859	1.662	1.296	1.859
ANN1	NMAE(%)	4.793	2.921	4.155	3.738	4.936	4.671	3.717	4.007
	NPL	1.721	1.267	1.580	1.353	1.874	1.690	1.356	1.874
ANN2	NMAE(%)	4.789	2.938	4.042	3.735	4.939	4.536	3.560	4.012
	NPL	1.722	1.275	1.524	1.352	1.874	1.650	1.293	1.873
ANN3	NMAE(%)	4.817	2.927	4.096	3.738	4.932	4.494	3.502	4.017
	NPL	1.729	1.270	1.549	1.354	1.875	1.636	1.281	1.875
ANN4	NMAE(%)	4.792	2.906	4.022	3.735	4.924	4.481	3.500	4.005
	NPL	1.722	1.264	1.504	1.351	1.873	1.629	1.281	1.873
ANN5	NMAE(%)	4.793	2.902	<b>3.859</b>	3.727	4.899	4.487	3.480	4.006
	NPL	<i>1.719</i>	1.260	<i>1.431</i>	<i>1.349</i>	1.861	1.628	1.274	1.861
GBM1	NMAE(%)	4.822	2.945	4.468	3.739	4.961	4.479	3.562	4.022
	NPL	1.733	1.284	1.731	1.351	1.872	1.629	1.300	1.873
GBM2	NMAE(%)	4.808	2.941	4.474	3.736	4.963	4.478	3.550	4.020
	NPL	1.735	1.283	1.732	1.350	1.872	1.631	1.295	1.872
GBM3	NMAE(%)	4.806	2.936	4.730	3.768	4.969	4.491	3.504	4.002
	NPL	1.743	1.285	1.860	1.360	1.882	1.628	1.291	1.882
GBM4	NMAE(%)	4.845	2.946	4.348	3.754	4.974	4.544	3.554	4.023
	NPL	1.742	1.282	1.671	1.357	1.878	1.650	1.302	1.878
RF1	NMAE(%)	4.965	3.060	4.207	3.883	5.115	4.703	3.715	4.159
	NPL	1.796	1.338	1.577	1.407	1.941	1.701	1.357	1.941
RF2	NMAE(%)	4.920	3.012	4.221	3.852	5.057	4.637	3.659	4.110
	NPL	1.777	1.312	1.596	1.393	1.913	1.681	1.336	1.913

*Note:* The smallest NMAE (%) at each location is in boldface. The smallest NPL at each location is in red italic.

Table 3.6: R-square of the least square fit

Site	C1	C2	C3	C4	C5	C6	C7	C8
$R^2$	0.97	0.98	0.99	0.98	0.95	0.99	0.95	0.99

### 3.4 Conclusion

This chapter developed a two-step probabilistic wind forecasting method based on pinball loss optimization, in conjunction with a multi-model deterministic forecasting framework. Different types of predictive distributions were compared, and the Laplace distribution was found to be the most suitable predictive distribution type. The optimal shape parameter of the predictive distribution was determined by minimizing the sum of pinball loss in the training stage. A surrogate model of the optimal shape parameter was used to estimate a pseudo-optimal shape parameter in the forecasting stage. Results showed that the M3-Laplace model could reduce the pinball loss score metric by up to 35% compared to benchmark models. Also, M3-Laplace showed better reliability than that of the M3-Laplace-w, which indicates the effectiveness of the pinball loss optimization. The sharpness intervals size of the M3-Laplace forecasts ranges from 2% to 18%, which indicates low sharpness.

Results also showed a linear relationship between the deterministic forecasting metric (i.e., NMAE) and the probabilistic forecasting metric (i.e., NPL). This indicates that a better deterministic model will very likely result in a more accurate probabilistic model with the developed framework.

## CHAPTER 4

### IMPROVING PROBABILISTIC FORECASTING VIA ENSEMBLE LEARNING

Ensemble methods have been used in the literature to better assist probabilistic wind power forecasting due to its high accuracy. This chapter is an extension of Chapter 3, and we seek to improve the probabilistic renewable energy forecasting accuracy through multi-distribution ensemble learning. In this chapter, we present a probabilistic wind power forecasting methods based on multi-distribution ensemble learning. Two ensemble strategies, i.e., competitive ensemble and cooperative ensemble are explored. Non-negative weight parameters are assigned to individual probabilistic forecast at each hour. Then, these weight parameters are estimated through minimizing the pinball loss and assigned to next hour using persistence method. The effectiveness of the proposed algorithm is validated using the Wind Integration National Dataset (WIND) Toolkit. Numerical results of case studies at four locations show that the developed ensemble probabilistic forecasting methodology has improved the pinball loss metric score by up to 20% compared to the best individual model. The proposed method also outperforms over other baseline ensemble models.

#### 4.1 Motivation and Objective

Wind power is considered to be one of the most promising renewable energy source in modern power systems due to the great economic, technological, and environmental incentives they involve (Bracale and De Falco, 2015). Accurate wind forecasts benefit power system stakeholders from different perspectives. For example, the wind power producers are subjected to discounted price or even penalty for the underestimated or overestimated wind power during the market bid. From system operator's viewpoint, accurate wind forecasting helps reduce the amount of operating reserves that are needed to balance generation and load (Botterud et al., 2011).

The forecasting models have gone through extensive improvement and expansion in recent years. Based on the adopted approaches, the classification of forecasting models is summarized in

Fig. 4.1 . They are preliminarily divided into deterministic forecasting and uncertainty analysis. The former can obtain accurate forecasting results after a specific time horizon, while the latter can provide probabilistic and confidence levels for the uncertainty of desired forecasts.

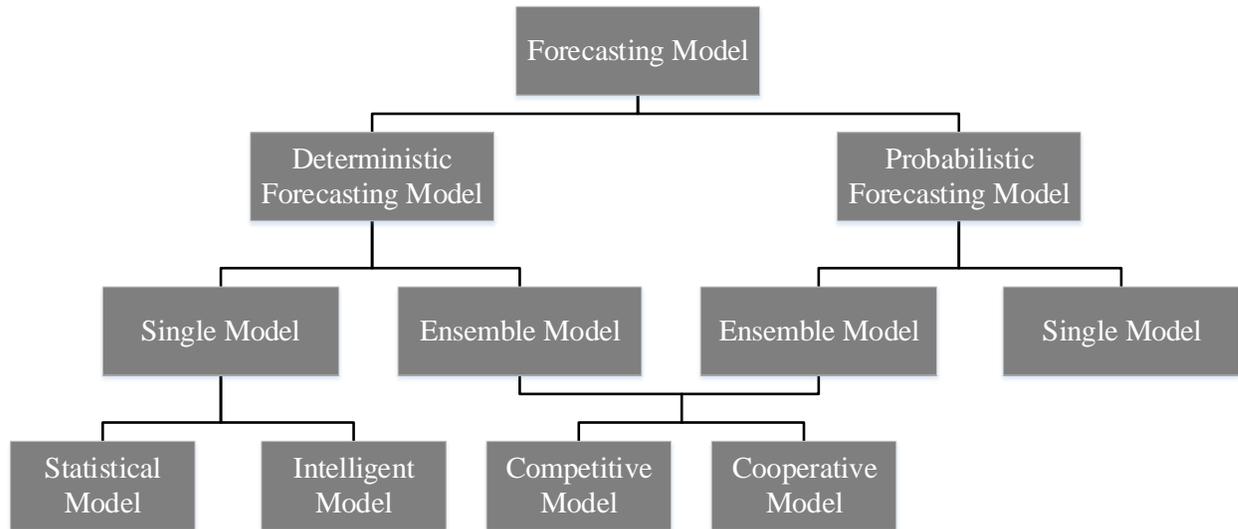


Figure 4.1: Wind forecasting methods categorization

Among the above methods, the most common practice in dealing with uncertainties in wind forecasting are ensemble methods. Ensemble methods have shown to be able to improve the performance of both deterministic and probabilistic wind forecasting (Sun et al., 2020a). However, most of the probabilistic forecasting methods are developed based on a single model, especially for parametric approaches which require a distribution assumption. It is challenging to accurately characterize the associated uncertainty by using a single distribution type, due to the daily and seasonal variability in wind power. A number of wind power forecasting projects including probabilistic forecasts have been conducted in recent year, such as wind forecast improvement project (WFIP) (Freedman et al., 2014), Alberta Electric System Operator (AESO) wind power forecasting pilot project (Jørgensen and Möhrlen, 2008), and ANEMOS project(Giebel et al., 2011). However, the performance of applying ensemble methods to predictive distribution optimization-based probabilistic wind forecasting is still not well studied. In this chapter, we seek to integrate the

advantages of different types of predictive distributions, and develops an ensemble probabilistic wind power forecasting framework that combines multiple predictive distribution types.

## 4.2 Methodology

The overall framework of the proposed ensemble probabilistic forecasting method, named Multi-Distribution Ensemble (MDE), is a two-step forecasting method, which consists of deterministic forecasts generation and ensemble probabilistic forecasts generation. The proposed MDE framework is able to integrate with both competitive and cooperative ensemble strategies. The main differences between competitive and cooperative ensemble models are listed as follows.

1. The competitive ensemble method combines different individual predictive distribution models together to an ensemble distribution model first. Then, probabilistic forecasts are obtained from the ensemble distribution model.
2. The cooperative ensemble method generates different individual probabilistic forecasts first based on each predictive distribution type. Then, ensemble probabilistic forecasts are generated through combining individual probabilistic forecasts to ensemble probabilistic forecasts.

The overall framework of the developed competitive MDE method is illustrated in Fig. 4.2. In the first step, a Q-learning-based forecasting framework is adopted to generate short-term deterministic wind forecasts (i.e., 1-6 hour-ahead (HA)), and a numerical weather prediction (NWP) model is used to generate day-ahead deterministic forecasts. The generated deterministic forecasts are considered as mean values of predictive distributions. In the second step, a set of unknown parameters (i.e., weight parameters and standard deviations) of the ensemble model are determined by minimizing the pinball loss based on training data. Then a surrogate model is constructed to represent each optimal parameter as a function of the deterministic forecast. During online forecasting, a set of pseudo-optimal parameters of the ensemble model are estimated by the surrogate

model and deterministic forecasts. Finally, probabilistic forecasts are generated with the distribution means (i.e., deterministic forecasts) and pseudo-optimal parameters.

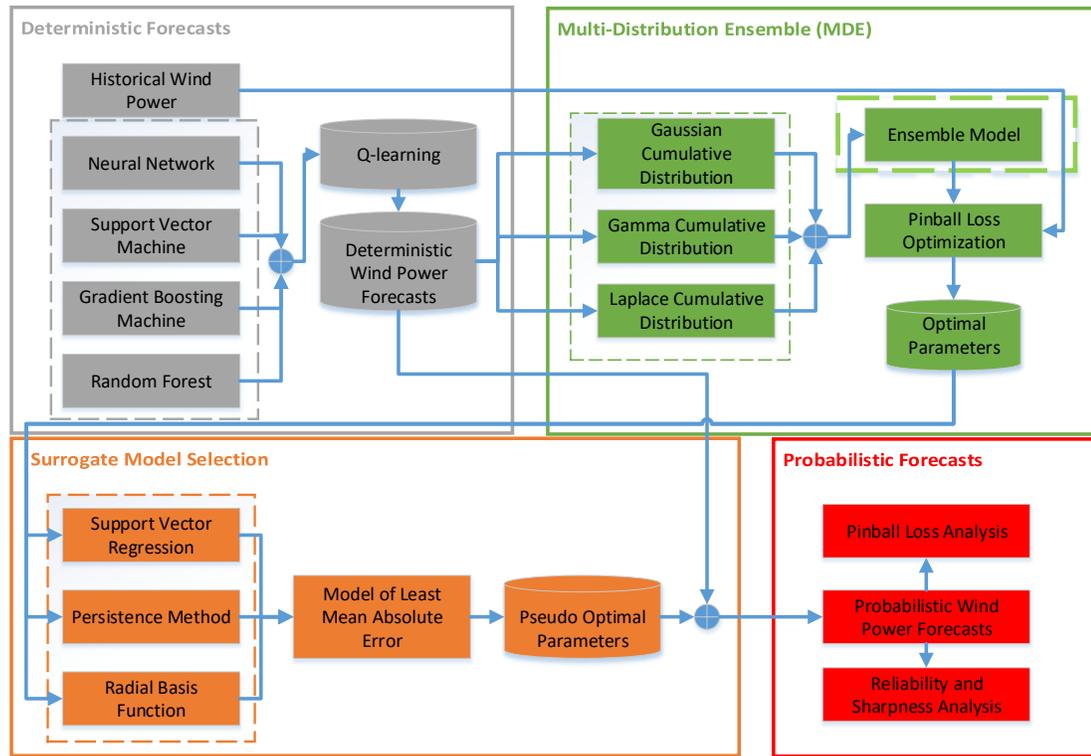


Figure 4.2: Overall framework of the MDE competitive probabilistic wind power forecasting method

The overall framework of the cooperative MDE method is illustrated in Fig. 4.3. The major differences between the cooperative MDE and competitive MDE methods are highlighted in the **blue** dashed square. For the cooperative MDE method, the first step is also to generate deterministic forecasts. In the second step, individual probabilistic forecasts are generated by using each single predictive distribution based on the training dataset (Sun et al., 2019). The unknown parameters (i.e., standard deviations) of each predictive distribution are optimized. A weight parameter is assigned to quantile forecasts from each individual model, and these weight parameters are optimized again by minimizing the pinball loss. Then a surrogate model is developed to represent

each optimal weight as a function of the deterministic forecast. Similar to the competitive MDE method, during the online forecasting, pseudo-optimal weights are estimated and used to generate ensemble probabilistic forecasts. Finally, the method with the minimum pinball loss is chosen to produce the final ensemble probabilistic forecasts. Overall, the competitive MDE method optimizes the unknown weight parameters and standard deviations simultaneously; the cooperative MDE method optimizes the standard deviations when generating the individual model forecasts, and optimize the weight parameters at the ensemble stage.

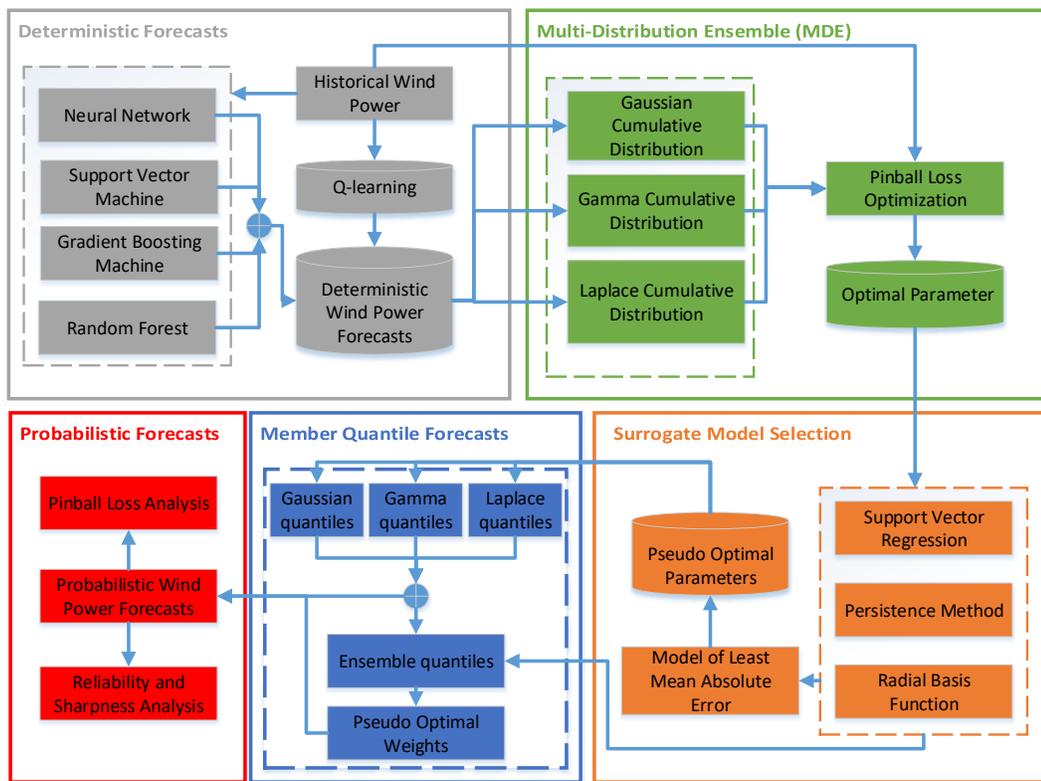


Figure 4.3: Overall framework of the MDE cooperative probabilistic wind power forecasting method

#### 4.2.1 Q-learning enhanced deterministic forecasting

A large number of models have been developed in the literature for deterministic wind forecasting. Most of existing deterministic forecasting methods are either selected based on the overall perfor-

mance or ensembled by multiple models. Ensembling multiple deterministic models enhances the robustness of the forecasting by reducing the risk of unsatisfactory models, but does not guarantee the best accuracy (Feng and Zhang, 2018). Selecting a model based on the overall forecasting performance, generally neglects the local performance of the selected model.

In this study, a Q-learning enhanced deterministic forecasting method is adopted for short-term forecasting (1HA to 6HA), which seeks to choose the best forecasting model from a pool of state-of-the-art machine learning based forecasting models at each forecasting time step (Feng and Zhang, 2019). The developed method trains Q-learning agents based on the rewards of transferring from the current model to the next model. The Q-learning agents converge to the optimal dynamic model selection policy, which will be applied to select the best model for forecasting in the next step based on the current model. The dynamic model selection process is expressed as:

$$S = \{s\} = \{s_1, s_2, \dots, s_I\} \quad (4.1)$$

$$A = \{a\} = \{a_1, a_2, \dots, a_I\} \quad (4.2)$$

$$R^t(s_i, a_j) = \text{ranking}(M_i) - \text{ranking}(M_j) \quad (4.3)$$

where  $S$ ,  $A$ , and  $R$  are state space, action space, and reward function in the dynamic model selection Markov Decision Process, respectively. The parameters  $s$  and  $a$  are possible state and action, respectively.  $I$  is the number of models ( $M$ ) in the model pool. The reward function is defined as the model performance improvement, which ensures the effective and efficient convergence of Q-learning. More details about the Q-learning enhanced deterministic forecasting can be found in Ref. (Feng and Zhang, 2019).

#### **4.2.2 Probabilistic wind power forecasting model**

In this section, multiple predictive distribution types are used to generate different individual probabilistic wind power forecasting models. Probabilistic forecasts usually take the form of probability

density functions (PDFs) or quantiles. Parametric approaches have been widely used in the literature to estimate the density or distribution (Fatemi et al., 2018). For a certain form of predictive distribution, the PDF can generally be characterized by a mean value  $\mu$  and a standard deviation value  $\sigma$  as  $f(x|\mu, \sigma)$ , and the corresponding cumulative distribution function (CDF) can be deduced and denoted as  $F(x|\mu, \sigma)$ . The quantile function is one way of prescribing a probability distribution, which is the inverse function of its corresponding CDF. Therefore, the quantile function of a certain predictive distribution can be denoted as  $q_i = F^{-1}(\frac{i}{100})$ , where  $i$  stands for the  $i$ th quantile. The quantile functions  $q_{i,t}(\cdot)$  are distinct for different predictive distributions, such as Gaussian, gamma, and laplace. Gaussian distribution is one of the most commonly used predictive distributions in parametric approaches (Lange, 2005). The PDF of Gaussian distribution  $f(x|\mu, \sigma)$  is expressed as

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.4)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation, respectively. The laplace distribution has also been applied to describe wind power generation (Sun et al., 2018). The PDF of laplace distribution  $f(x|\mu, b)$  is formulated as

$$f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \quad (4.5)$$

where  $\mu$  and  $b$  are the location and scale parameters, respectively. The relationship between the standard deviation  $\sigma$  and the shape parameter of laplace distribution can be expressed as

$$b = \frac{\sigma}{\sqrt{2}} \quad (4.6)$$

Similarly, for gamma distribution, the PDF is expressed as

$$f(x|k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \quad (4.7)$$

where  $k$  and  $\theta$  are the shape parameter and scale parameter, respectively. The relationship between the mean value  $\mu$ , standard deviation  $\sigma$ ,  $k$ , and  $\theta$  can be expressed as

$$\mu = k\theta \quad (4.8)$$

$$\sigma = \sqrt{k}\theta \quad (4.9)$$

The location parameter and shape parameter from all of the above distributions can be represented by the mean and standard deviation. Therefore, the PDF can be obtained by estimating the mean and standard deviation in the training or forecasting stages. The pseudocode of generating the MDE forecasting member models is illustrated in Algorithm 1.

---

**Algorithm 1:** Generate quantile forecasts through each single distribution

---

**Data:** Deterministic wind power forecasts

**Result:** Quantile forecasts

- 1 Initialization: Obtain PDF of a single model and represent it in the form of mean  $\mu$  and standard deviation  $\sigma$  as  $f(x|\mu, \sigma)$ ;
  - 2 Calculate CDF  $F(x|\mu, \sigma)$  of the predictive distribution;
  - 3 **if** CDF is invertible **then**
  - 4 Calculate quantile function through  $q_i(F^{-1}(x|\mu, \sigma, \frac{i}{100}))$
  - 5 **else**
  - 6 Use Newton-Raphson method to calculate quantile  $q_i$
  - 7 **end**
  - 8 Calculate optimal  $\sigma$ 's through pinball loss optimization;
  - 9 Build a surrogate model between deterministic forecasts and optimal  $\sigma$ s;
  - 10 Estimate pseudo-optimal  $\sigma$  based on deterministic forecasts;
  - 11 Generate quantile forecasts;
-

### 4.2.3 Competitive MDE method

For the competitive MDE method, to combine the advantages of different predictive distributions, a general combined quantile function with  $N$  member models is formulated as follows:

$$q_{i,t}^p(\mu_t, \omega_t^p, \sigma_t) = \sum_{n=1}^N \omega_{t,n}^p q_{i,t,n}(\mu_t, \sigma_{t,n}) \quad (4.10)$$

where  $q_{i,t}^p$  is the combined  $i$ th quantile at the  $t$ th forecasting step,  $q_{i,t,n}(\mu_t, \sigma_{t,n})$  is the  $i$ th member quantile at the  $t$ th forecasting step, and  $\omega_{t,n}^p$  is the weight of the  $n$ th model at time  $t$  satisfying that

$$0 \leq \omega_{t,n}^p \leq 1 \quad (4.11)$$

$$\sum_{n=1}^N \omega_{t,n}^p = 1 \quad (4.12)$$

The unknown parameters in the ensemble quantile function can be solved by minimizing the pinball loss at each time step. Pinball loss is a widely used metric to evaluate probabilistic forecasts, which is defined by

$$L_{i,t}(q_{i,t}, x_t) = \begin{cases} (1 - \frac{i}{100}) \times (q_{i,t} - x_t), & x_t < q_{i,t} \\ \frac{i}{100} \times (x_t - q_{i,t}), & x_t \geq q_{i,t} \end{cases} \quad (4.13)$$

where  $x_t$  represents the observation at time  $t$ . For a given  $i$  percentage, the quantile  $q_{i,t}$  represents the value of a random variable whose CDF is  $i$  percentage at time  $t$ . The pinball loss optimization problem for the competitive MDE model can be formulated as:

$$\begin{aligned} \min_{\omega_t^p, \sigma_t} & \sum_{i=1}^{99} L_{i,t}(\sum_{n=1}^N \omega_{t,n}^p q_{i,t,n}(\mu_t, \sigma_{t,n}), x_t) \\ \text{s.t.} & \sum_{n=1}^N \omega_{t,n}^p = 1 \\ & 0 \leq \omega_{t,n}^p \leq 1, \forall n \in N \\ & \sigma_l \leq \sigma \leq \sigma_u. \end{aligned} \quad (4.14)$$

where  $\sigma_l$  and  $\sigma_u$  represent the lower and upper bounds of the unknown standard deviation, respectively, which are selected based on the forecasting target (Hansen et al., 2017).

#### 4.2.4 Cooperative MDE method

For the cooperative MDE method, with the calculated  $i$ th quantile of the  $n$ th model, a general combined quantile with  $N$  individual models is formulated as:

$$q_{i,t}^o(\mu_t, \omega_t^o) = \sum_{n=1}^N \omega_{t,n}^o q_{i,t,n}(\mu_t) \quad (4.15)$$

$$0 \leq \omega_{t,n}^o \leq 1 \quad (4.16)$$

$$\sum_{n=1}^N \omega_{t,n}^o = 1 \quad (4.17)$$

where  $q_{i,t}^o$  is the combined  $i$ th quantile forecasts at the  $t$ th forecasting step,  $q_{i,t,n}(\mu_t)$  is the  $i$ th member quantile forecasts at the  $t$ th forecasting step from model  $n$ , and  $\omega_{t,n}^o$  is the weight of the quantile forecasts from the  $n$ th model. Different from the competitive MDE method, the unknown parameters in the cooperative MDE method only include weight parameters.

It is seen from Eqs. 4.10 and 4.15 that the weights of individual models at each time step are different, and the location and shape parameters at each time step are also varying. Therefore, we need to adaptively estimate the unknown parameters at each time step. By contrast, for the cooperative MDE method, the weight parameters can be estimated by solving the following optimization problem:

$$\begin{aligned} \min_{\omega_t^o} \quad & \sum_{i=1}^{99} L_{i,t} \left( \sum_{n=1}^N \omega_{t,n}^o q_{i,t,n}(\mu_t), x_t \right) \\ \text{s.t.} \quad & \sum_{n=1}^N \omega_{t,n}^o = 1 \\ & 0 \leq \omega_{t,n}^o \leq 1, \forall n \in N. \end{aligned} \quad (4.18)$$

The optimization problems in Eqs. 4.14 and 4.18 are strictly constrained by three constraints: (i) all the weights of ensemble member models or member quantile forecasts must be nonnegative; (ii) the sum of all weights equals one; and (iii) the standard deviation of each component model must be in the variable range of the corresponding distribution. In this study, the genetic algorithm (GA) (Scrucca et al., 2013) is adopted to solve the two optimization problems. In this study, the

population size is set to be 50, the mutation percentage is 0.1, the crossover percentage is 0.8, and the maximum number of iterations is 200. The optimization stops when the improvement is less than 0.01%.

#### 4.2.5 Surrogate Model

The optimization models in Eqs. 4.14 and 4.18 calculate the optimal parameters of predictive distributions with observations. However, when we generate probabilistic forecasts, we do not know the observations and can only assume the deterministic forecasts as mean values of the predictive distribution. The optimal standard deviations and optimal weights at each forecasting time step are needed to generate probabilistic forecasts. To this end, a surrogate model is built to represent the optimal parameters as functions of the deterministic forecasts in the training stage, which can be expressed as:

$$\hat{\sigma}_{t,n} = f(\hat{x}_t) \quad (4.19)$$

$$\hat{\omega}_{t,n} = g(\hat{x}_t) \quad (4.20)$$

where  $\hat{x}_t$  is the deterministic forecast at time  $t$ ,  $f(\hat{\sigma})$  and  $g(\hat{\omega})$  are the surrogate models of the optimal standard deviation and weight parameter of the predictive distribution, respectively. The surrogate models are used to estimate the pseudo-optimal standard deviation and pseudo-optimal weight parameters at the online forecasting stage.

Several possible surrogate model types can be used, such as support vector regression (SVR), radial basis function (RBF), and persistence model (PS). In this study, a surrogate model selection framework is developed to choose the most suitable surrogate model for estimating weights and standard deviations of each ensemble member models. Figure 4.4 shows the procedure of selecting the optimal surrogate model. At the forecasting training stage, a small portion (i.e., 8% of the training data in this study) of the training data is used to build a surrogate model. The optimal training parameters are fed into a surrogate model pool (that consist of SVR, RBF, and PS) to find

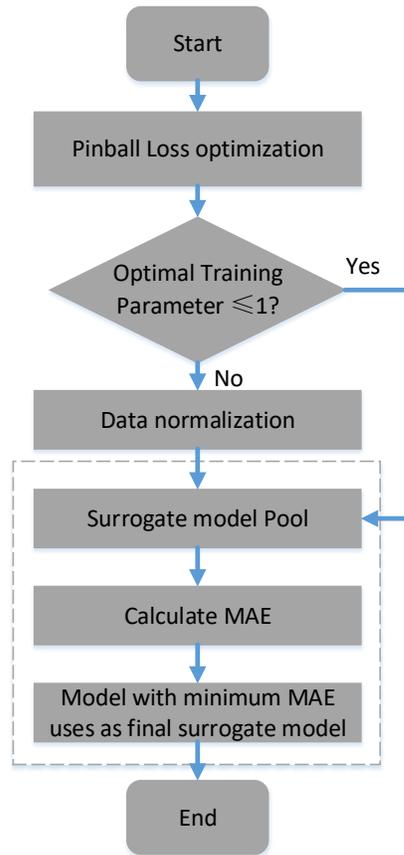


Figure 4.4: Surrogate model selection

the best fitted surrogate model between the point forecast value and each optimal parameters. The model with the minimum mean absolute error (MAE) is selected as the surrogate model for that parameter to be used at the online forecasting stage. In the forecasting stage, the optimal unknown parameters of the predictive distribution are estimated through the surrogate, and thus the estimated parameters are referred to as pseudo-optimal parameters.

### 4.3 Case Studies and Results

The developed MDE probabilistic forecasting framework was evaluated at 7 locations selected from the Wind Integration National Dataset (WIND) Toolkit (Draxl et al., 2015). The WIND

Toolkit includes meteorological information (e.g., wind direction, wind speed, air temperature, surface air pressure, density at hub height), synthetic actual wind power, and wind power forecasts generated by the Weather Research and Forecasting (WRF) model. It covers over 126,000 locations in the United States. In addition to the day-ahead forecasting, very-short-term forecasts (i.e., 1HA to 6HA) are generated by using the Q-learning enhanced deterministic forecasting method. In this study, the duration of the collected data at the selected 7 locations spans two years from January 1<sup>st</sup> 2011 to December 31<sup>st</sup> 2012. The data information at the selected 7 locations is briefly summarized in Table 4.1. For all the locations, the first 3/4 of the data is used as training data, in which the first 11/12 is used to train the deterministic forecast models and the remaining 1/12 of the training data is used to build the surrogate models of the optimal standard deviations and weight parameters. The effectiveness of the forecasts is validated by the remaining 1/4 of the data. The developed MDE method is able to generate probabilistic forecasts at multiple forecasting horizons, and 1HA-6HA and day-ahead wind power forecasts are explored in this study.

Table 4.1: Data summary of the selected 7 WIND Toolkit sites

Case No.	Site ID	Lat.	Long.	Capacity (MW)	State
C1	4816	29.38	-100.37	16	TX
C2	8979	31.53	-95.62	16	TX
C3	10069	32.31	-98.26	16	TX
C4	10526	32.44	-100.55	16	TX
C5	1342	27.12	-97.86	16	TX
C6	2061	27.95	-97.40	14	TX
C7	9572	31.99	-100.118	16	TX

In the study, two different weight averaging methods for model ensemble and three single predictive distribution models are used as baselines in the case studies. The two benchmark weight averaging based ensemble methods are Arithmetic Averaging (AA) and Weighted Averaging (WA).

1. AA: The arithmetic weights apply equally to different models:

$$\omega_{t,n}^{AA} = \frac{1}{N}. \quad (4.21)$$

Then, the combined quantile of AA can be calculated through:

$$q_{i,t}^{AA}(\mu_t, \omega_t^{AA}) = \sum_{n=1}^N \omega_{t,n}^{AA} q_{i,t,n}(\mu_t) \quad (4.22)$$

2. WA: Each quantity to be averaged is assigned a weight that represents the relative importance of that quantity. The model with a higher accuracy is assigned a larger weight:

$$\omega_{t,n}^{WA} = \frac{1 / \sum_{i=1}^{99} L_{i,t,n}(q_{i,t,n}(\mu_t), x_t)}{\sum_{n=1}^N 1 / \sum_{i=1}^{99} L_{i,t,n}(q_{i,t,n}(\mu_t), x_t)} \quad (4.23)$$

Similarly, the combined quantile of WA can be calculated through:

$$q_{i,t}^{WA}(\mu_t, \omega_t^{WA}) = \sum_{n=1}^N \omega_{t,n}^{WA} q_{i,t,n}(\mu_t) \quad (4.24)$$

The three single predictive distribution based optimization models are Q-learning with Gaussian distribution (Q-Gaussian), Q-learning with Gamma distribution (Q-Gamma), and Q-learning with Laplace distribution (Q-Laplace). The details of the pinball loss optimization based model can be found in (Sun et al., 2018).

Normalized indices of standard metrics root mean squared error and mean absolute error, *i.e.*, NMAE and NRMSE, are adopted to evaluate the performance of deterministic forecasts. The forecasting errors by using the Q-learning based deterministic model at the selected locations are summarized in Table 4.2. It is shown that the 1HA NMAE and NRMSE are in the ranges of 5%-8% and 8%-12%, respectively. The numerical weather prediction (NWP) method is used to produce day-ahead deterministic forecasts. It is shown that the day-ahead NMAE and NRMSE are in the range of 11%-14% and 15%-19%, respectively. Overall, the accuracies of the Q-learning

Table 4.2: Deterministic forecasting results by using Q-learning and NWP

Model	LAT	Metric	Site						
			C1	C2	C3	C4	C5	C6	C7
Q-learning	1HA	NMAE(%)	6.63	6.85	6.70	6.74	6.67	5.38	7.76
		NRMSE(%)	10.55	10.93	11.04	10.66	9.93	8.37	11.93
	2HA	NMAE(%)	10.72	11.08	11.20	11.28	10.36	8.34	10.90
		NRMSE(%)	15.86	16.33	16.93	16.64	14.60	12.31	17.14
	3HA	NMAE(%)	13.95	14.22	14.65	14.88	12.44	10.49	14.56
		NRMSE(%)	19.43	19.86	20.92	20.71	16.94	14.93	20.12
	4HA	NMAE(%)	16.43	16.76	17.69	18.00	13.84	12.17	16.28
		NRMSE(%)	21.98	22.22	23.92	23.83	18.56	16.85	21.79
	5HA	NMAE(%)	18.19	18.37	20.36	20.38	15.35	13.53	17.87
		NRMSE(%)	23.65	23.71	26.51	26.01	20.19	18.40	23.15
	6HA	NMAE(%)	19.85	19.60	21.94	22.63	16.51	14.70	18.41
		NRMSE(%)	25.09	24.85	27.82	27.82	21.31	19.47	23.55
NWP	DA	NMAE(%)	12.70	12.59	13.21	13.97	13.97	11.63	13.37
		NRMSE(%)	16.85	17.44	18.07	18.70	18.43	15.41	18.17

*Note:* LAT is the abbreviation for look-ahead time. DA: day-ahead

based 1HA to 6HA deterministic forecasts and the NWP-based day-ahead deterministic forecasts are reasonable.

To show the robustness of the developed MDE probabilistic forecasting framework, the normalized pinball loss values at different look-ahead time of the 7 wind farms are illustrated in Fig. 4.5. Figure 4.6 shows how the normalized pinball loss varies with the look-ahead time at the C7 site. The sum of pinball loss is averaged over all quantiles from 1% to 99% and normalized by the wind farm capacity at each site. A lower pinball loss score indicates a better probabilistic forecast. Results show that the MDE-competitive model has improved the pinball loss by up to 20.5%, and the MDE-cooperative model has improved the pinball loss by up to 8.5% compared to the three individual member models (i.e., Q-Gaussian, Q-Gamma, and Q-Laplace) and two baseline weight averaging ensemble models (i.e., AA and WA). It is seen from Fig. 4.6 that the MDE-competitive model has the smallest pinball loss value at all locations for different look-ahead times except for 1HA forecasts. For 1HA forecasts, the MDE-cooperative method has the smallest pinball loss value at all locations. Note that for all the look-ahead hours, the MDE-cooperative model has

also shown a better accuracy than individual member models, which validates the effectiveness of the MDE ensemble framework. The MDE-competitive model has shown a better accuracy than individual member models for 2HA-6HA and day-ahead forecasts.

The reason why the MDE-competitive method has a worse accuracy than the MDE-cooperative model and some single-distribution methods in 1HA forecasts is that, the accuracy of 1HA deterministic forecasts is significantly better than that of other look-ahead times. Thus there is less improvement space for the ensemble model to further improve the performance by optimizing the standard deviation and weight parameters simultaneously. Instead, the cooperative model that refines the 1HA forecasts from different single models performs better. By contrast, the competitive method performs better when the deterministic forecasting accuracy is relatively worse. The accuracy of deterministic forecasts is a major factor that affects the performance (i.e., pinball loss) of the final probabilistic forecasts. In addition, the one-step optimization in MDE-competitive and two-step optimization in MDE-cooperative also affects the probabilistic forecasting performance. The MDE-competitive method first ensembles the distribution members together, and then uses this combined distribution model to generate probabilistic forecasts, which optimizes the standard deviations and weight parameters simultaneously. The MDE-cooperative method first generates different single quantile forecasts, and then ensembles the single quantile forecasts to refined quantile forecasts. This two-step optimization (i.e., determine standard deviations first and then weights) may introduce more uncertainty compared to the one-step optimization. Similarly, the surrogate model accuracy and look-ahead forecasting times may also affect the final probabilistic forecasting results.

Figures 4.7(a) and 4.7(b) show the MDE-competitive and MDE-cooperative based day-ahead probabilistic wind power forecasts at the C3 site from 2012-08-09 to 2012-08-19. It is seen from Fig. 4.7 that the PIs of the MDE-competitive method cover the actual and forecasted wind power better than the MDE-cooperative method. For the MDE-cooperative method, even it has narrower PIs, some part of the PIs could not cover the observations. Therefore, the MDE-competitive

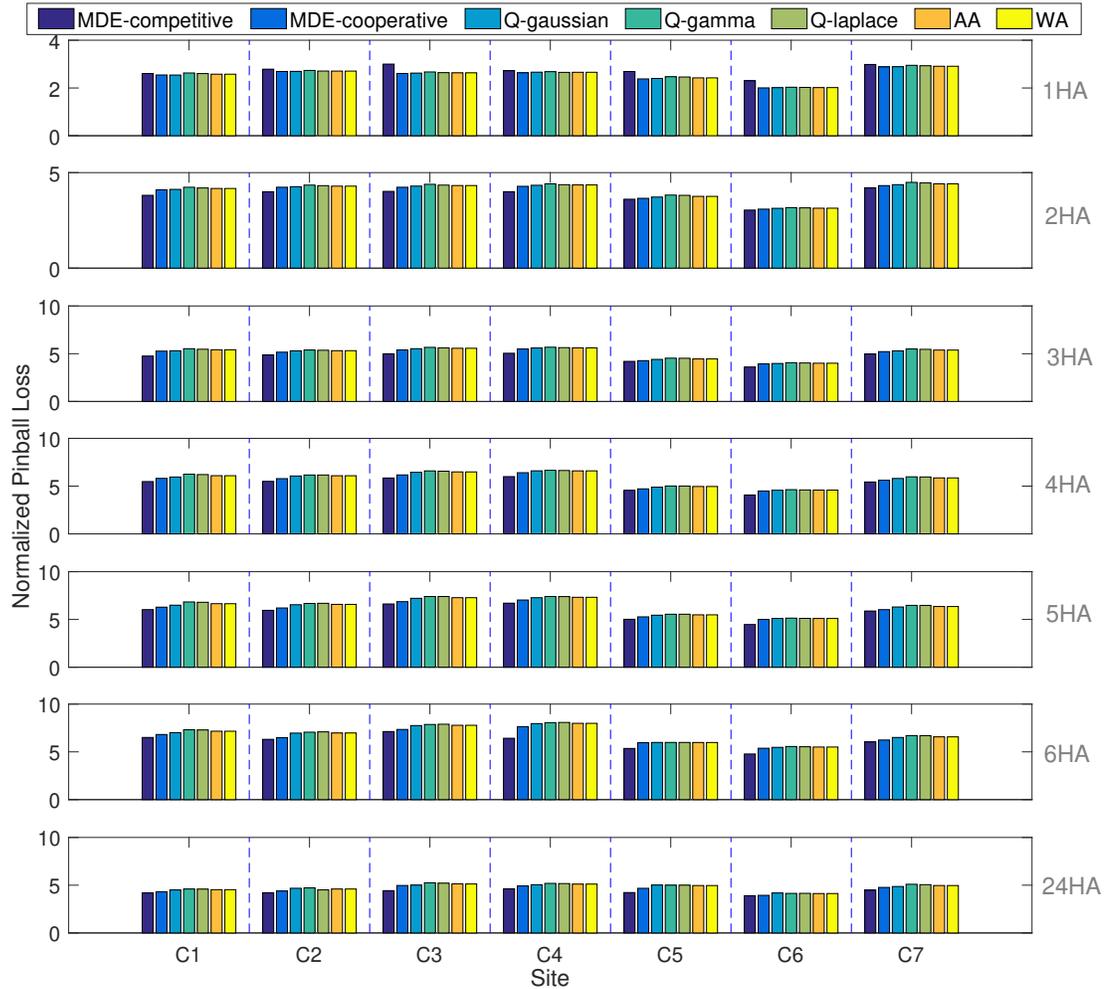


Figure 4.5: Normalized pinball loss of different ensemble methods and single methods

method is more reliable than the MDE-cooperative method in day-ahead forecasts. The width of the predictive interval varies with the wind power variability. When the wind power fluctuates more frequently, the predictive interval tends to be wider, and thereby the uncertainty in wind power forecasts is relatively higher.

In addition to pinball loss, two more standard metrics, i.e., reliability and sharpness, are also calculated to assess the probabilistic forecasting accuracy. Fig. 4.8 show the 1- to 6-HA and 24-HA reliability plots of the probabilistic forecasts at the selected C7 site with different forecasting mod-

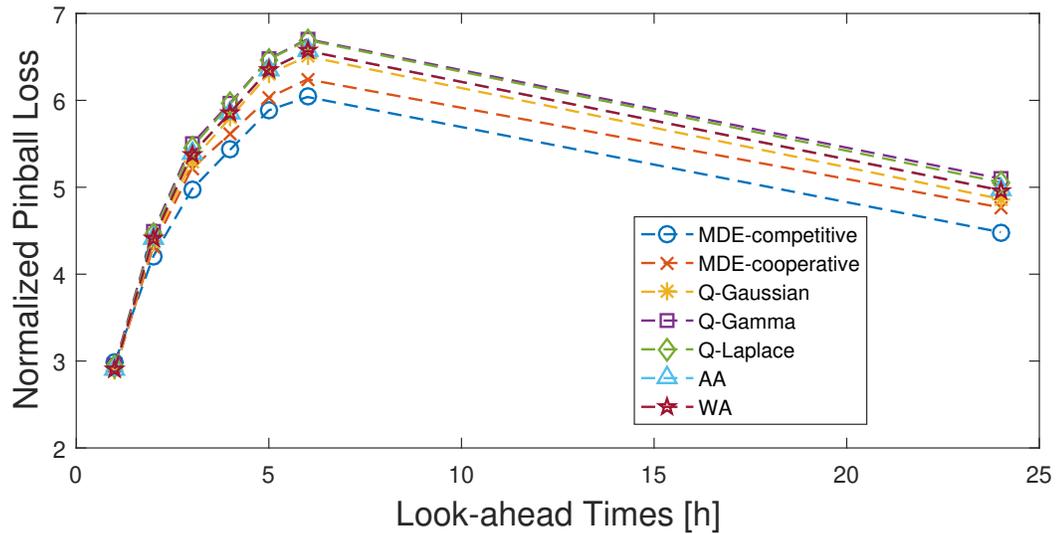
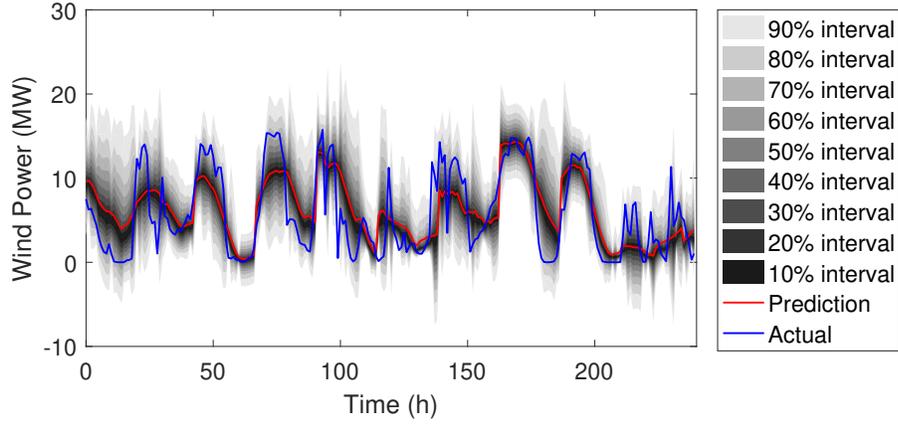


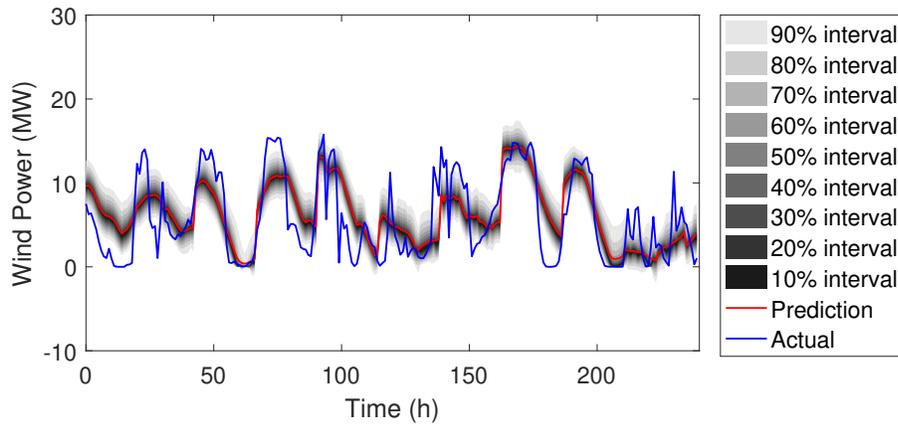
Figure 4.6: Normalized pinball loss of different models with different look-ahead hours at the C7 site

els. A forecast presents better reliability when the curve is closer to the diagonal. Overall it is seen from Fig. 4.8 that the proposed MDE ensemble framework (both competitive and cooperative) has better reliability performance than the baseline methods. Note that for 1-HA forecasts, the MDE-cooperative method has the best reliability performance and the MDE-competitive method has the best reliability performance for 2HA-6HA and 24HA forecasts. It is also seen that the two baseline ensemble methods (AA and MA) have better performance than the single-distribution models. Overall the results show the effectiveness of ensemble modeling for enhancing probabilistic forecasts. The reliability plots of all the case study sites are provided in Fig. 4.9-4.15.

The sharpness plots of the proposed ensemble system (i.e., MDE-competitive method and MDE-cooperative method); and baseline models (Q-Laplace, Q-Gaussian, Q-Gamma, AA, and WA ensemble methods) at the C7 site are compared in Fig. 4.8. The expected interval score decreases with increasing nominal coverage rate, and the sharpness of the MDE-competitive and MDE-cooperative models are slightly worse than some of the baseline models (e.g., Q-laplace and Q-gamma). It is mainly because the reliability and sharpness are two complementary metrics, and the improvement of reliability will sacrifice the sharpness to some extent. Overall, the interval



(a) MDE-competitive



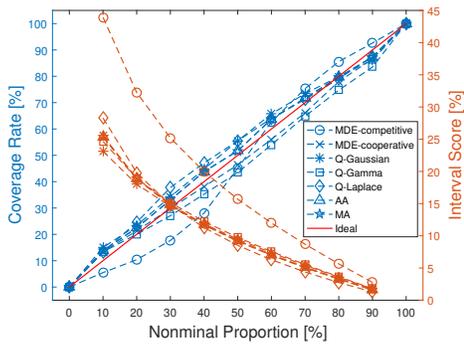
(b) MDE-cooperative

Figure 4.7: Day-ahead MDE-competitive and MDE-cooperative forecasts at the C3 site

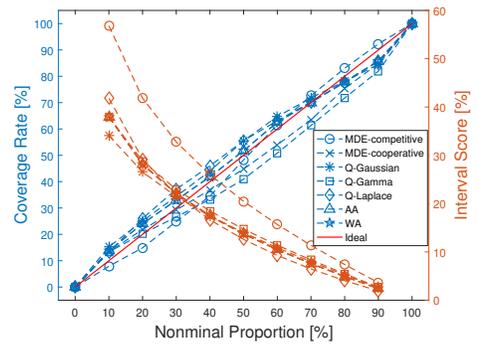
size of the MDE forecasts ranges from 10% to 120%, which indicates reasonable sharpness. The sharpness plots of all the case study sites are provided in Fig. 4.9-4.15 .

#### 4.4 Conclusion

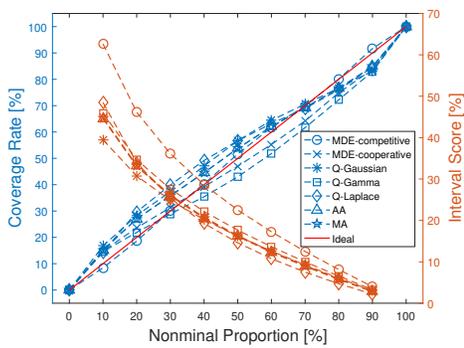
In this chapter, a multi-distribution ensemble (MDE) probabilistic wind forecasting framework was developed, with both competitive and cooperative ensemble strategies. Three types of predictive distributions (i.e., Gaussian, Gama, and Laplace) were adopted as ensemble members. The optimal ensemble weight parameters and standard deviations of different predictive distributions



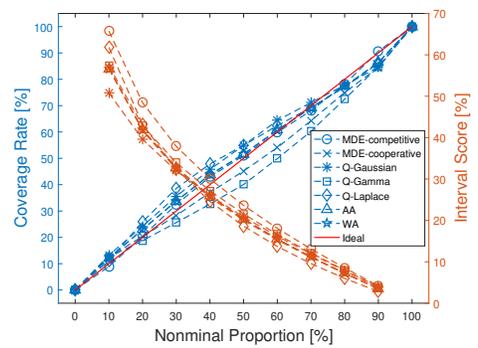
(a) 1HA



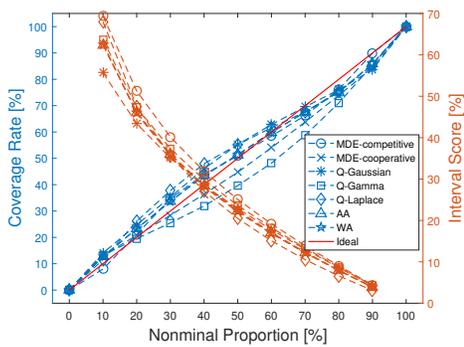
(b) 2HA



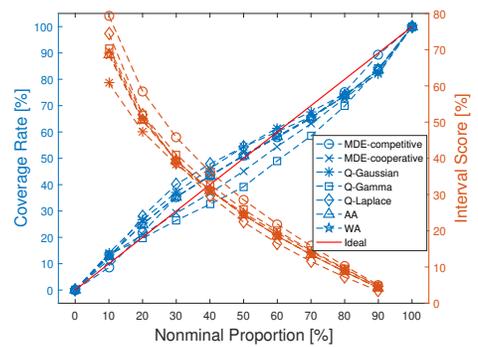
(c) 3HA



(d) 4HA

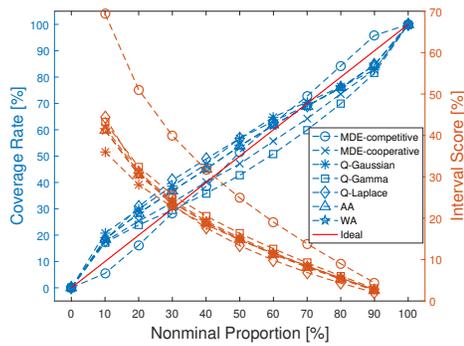


(e) 5HA



(f) 6HA

Figure 4.8: Reliability and sharpness comparison of different models at the C7 site with different look-ahead times (blue lines: reliability; orange lines: sharpness)



(g) 24HA

Figure 4.8: (continued) Reliability and sharpness comparison of different models at the C7 site with different look-ahead times (blue lines: reliability; orange lines: sharpness)

were determined by minimizing the sum of pinball loss in the training stage. A set of surrogate models of the optimal parameters were constructed to estimate the pseudo-optimal parameters in the forecasting stage. Case studies at 7 selected sites show that:

1. The developed MDE probabilistic forecasting framework could reduce the pinball loss score by up to 20.5% compared to benchmark models.
2. The MDE framework is robust under different forecasting time horizons at different locations.
3. The MDE-competitive method performs better in 2HA-6HA and 24HA forecasts and the MDE-cooperative method performs better in 1HA forecasts.
4. Both competitive and cooperative MDE methods have shown better reliability than single-distribution models and benchmark ensemble models. The sharpness intervals size of MDE forecasts ranges from 10% to 120%, which indicates reasonable sharpness.

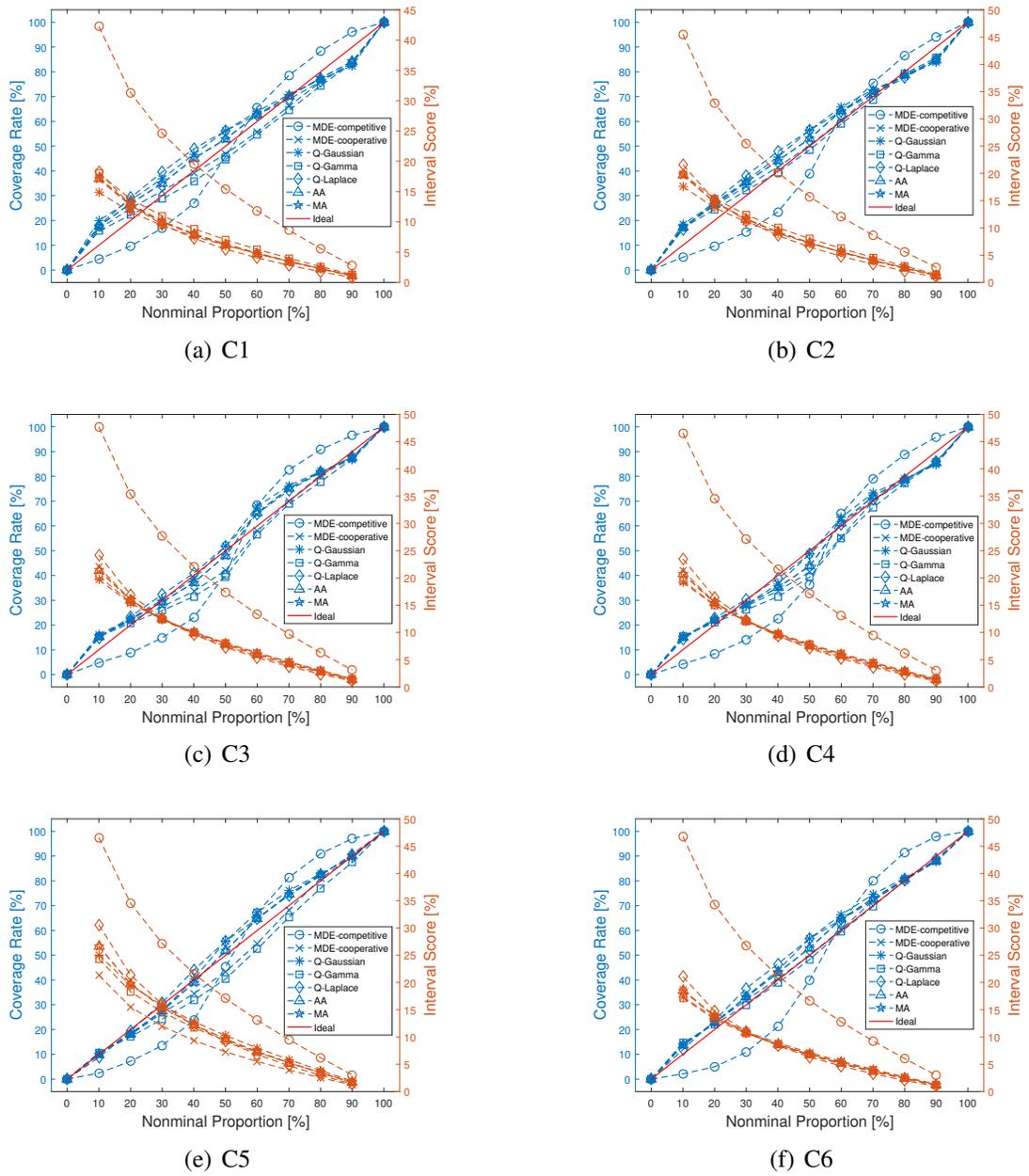
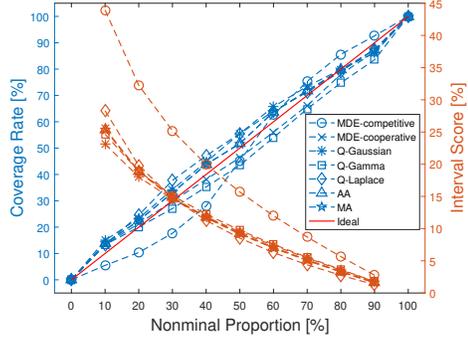
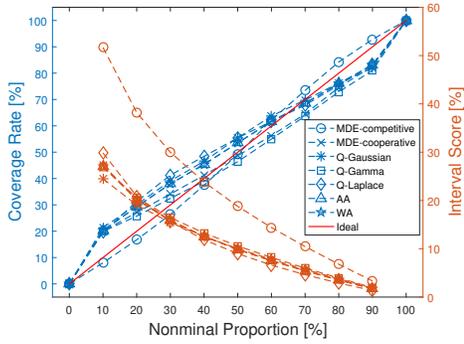


Figure 4.9: 1HA reliability and sharpness comparison of different models at the 7 sites

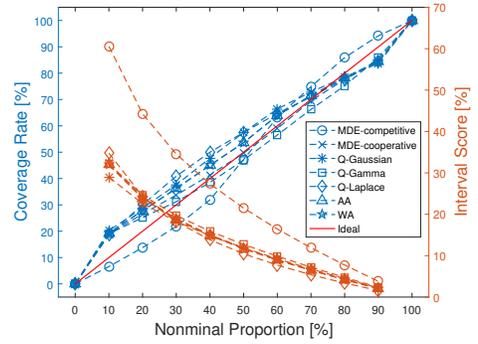


(g) C7

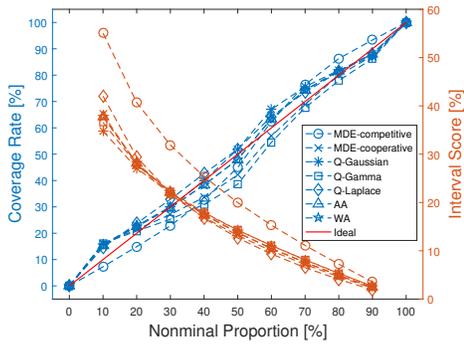
Figure 4.9: (continued) 1HA reliability and sharpness comparison of different models at the 7 sites



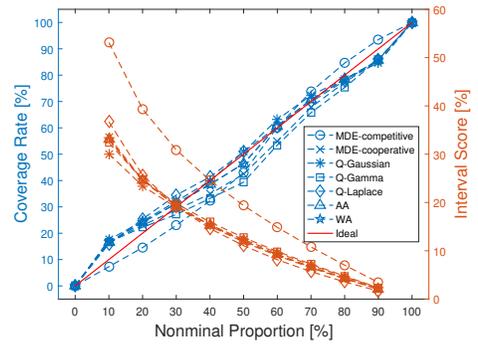
(a) C1



(b) C2

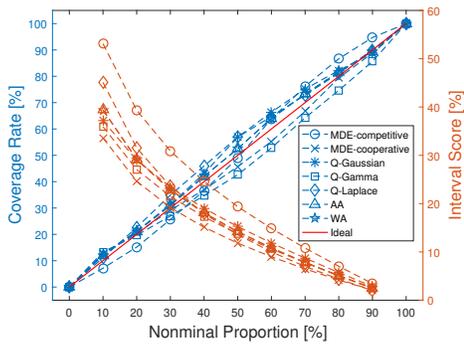


(c) C3

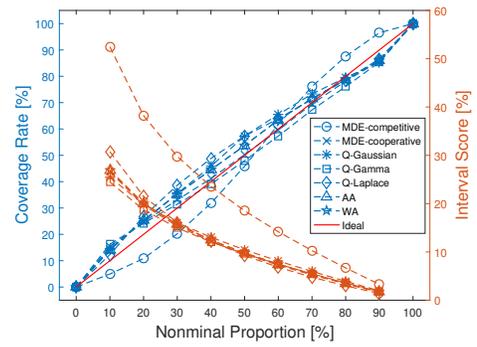


(d) C4

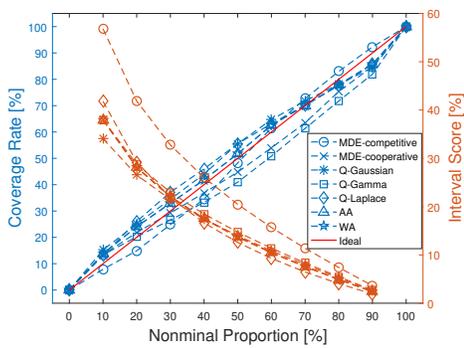
Figure 4.10: 2HA reliability and sharpness comparison of different models at the 7 sites



(e) C5

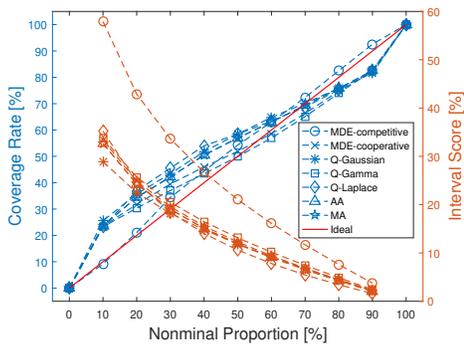


(f) C6

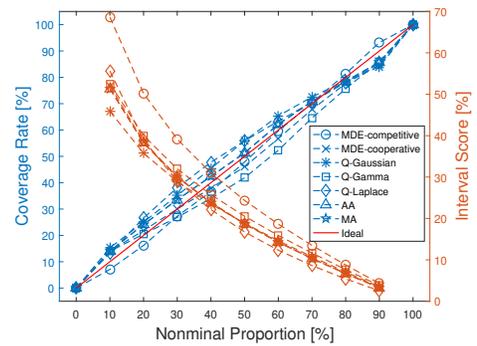


(g) C7

Figure 4.10: (continued) 2HA reliability and sharpness comparison of different models at the 7 sites

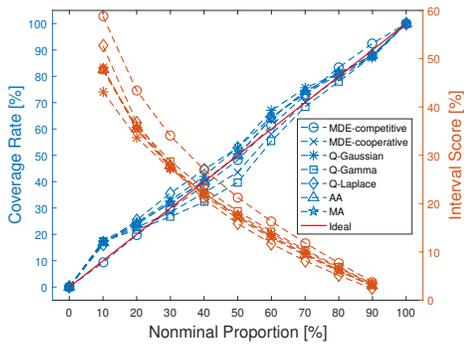


(a) C1

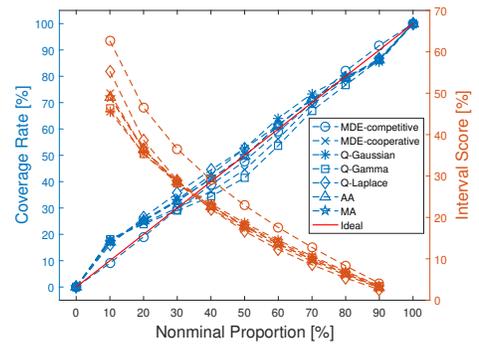


(b) C2

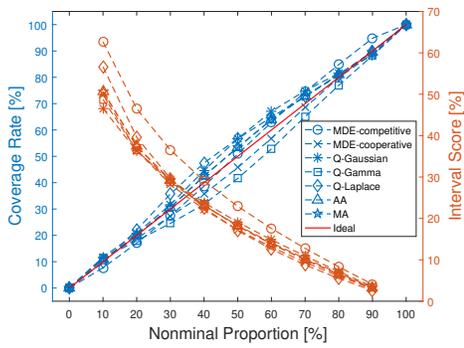
Figure 4.11: 3HA reliability and sharpness comparison of different models at the 7 sites



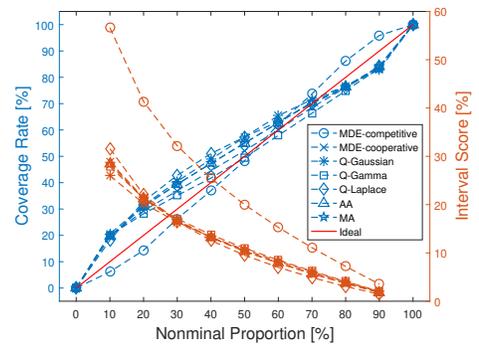
(c) C3



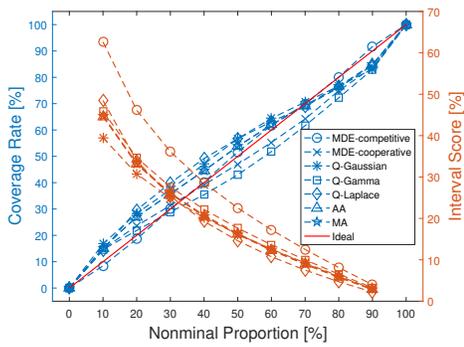
(d) C4



(e) C5

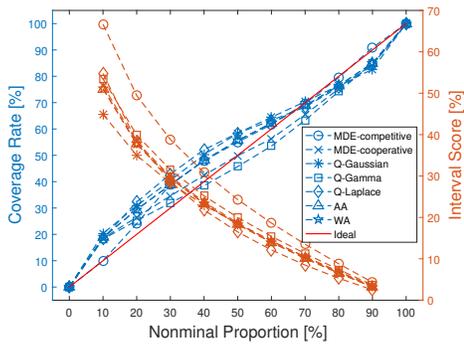


(f) C6

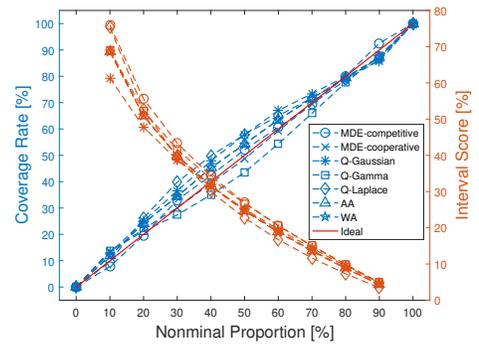


(g) C7

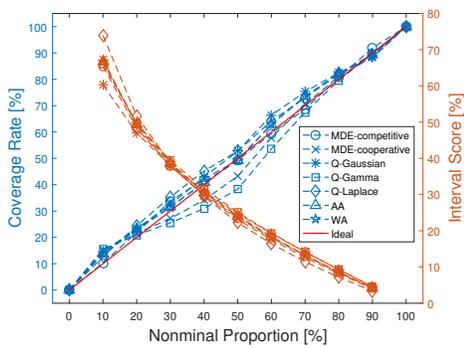
Figure 4.11: (continued) 3HA reliability and sharpness comparison of different models at the 7 sites



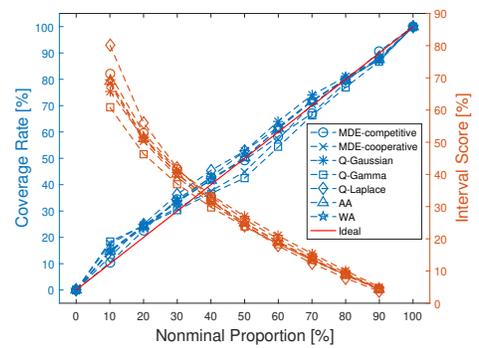
(a) C1



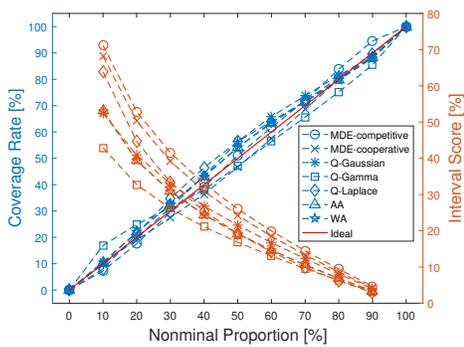
(b) C2



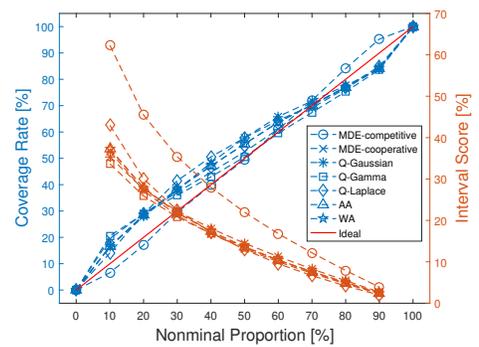
(c) C3



(d) C4

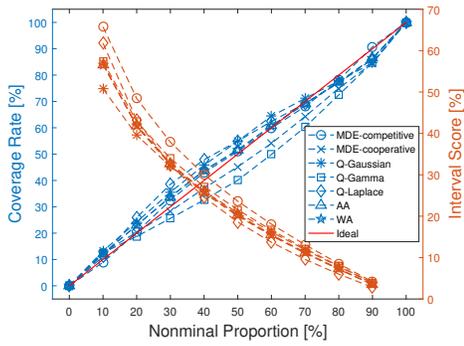


(e) C5



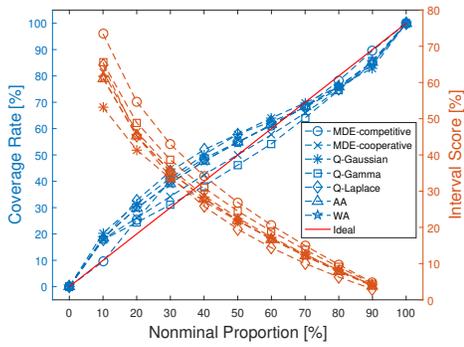
(f) C6

Figure 4.12: 4HA reliability and sharpness comparison of different models at the 7 sites

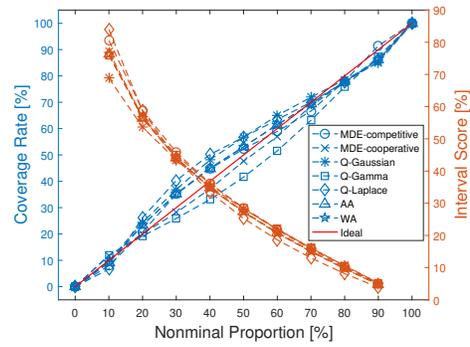


(g) C7

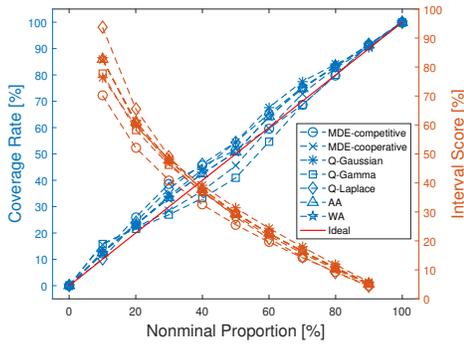
Figure 4.12: (continued) 4HA reliability and sharpness comparison of different models at the 7 sites



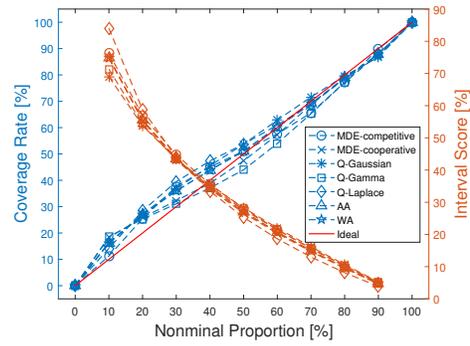
(a) C1



(b) C2

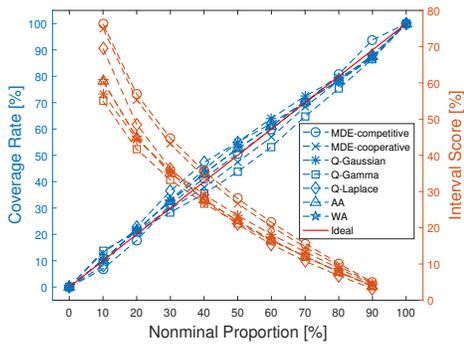


(c) C3

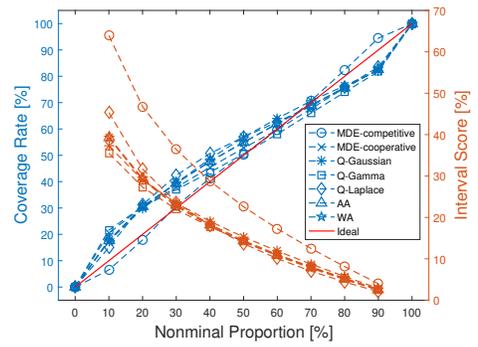


(d) C4

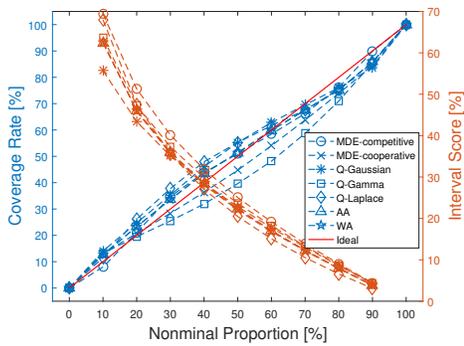
Figure 4.13: 5HA reliability and sharpness comparison of different models at the 7 sites



(e) C5

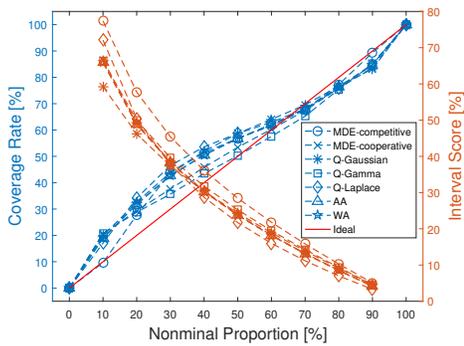


(f) C6

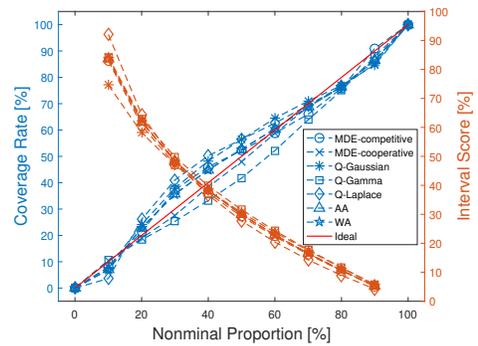


(g) C7

Figure 4.13: (continued) 5HA reliability and sharpness comparison of different models at the 7 sites

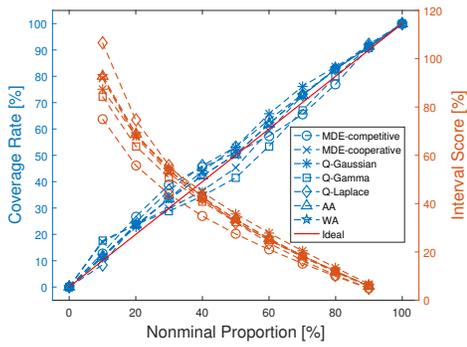


(a) C1

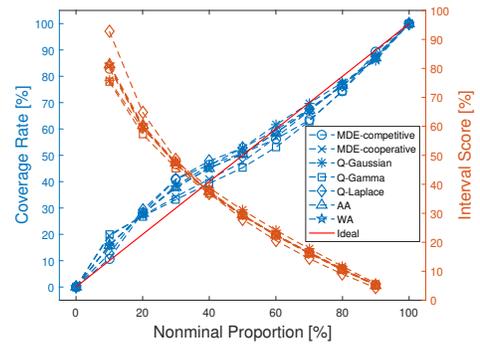


(b) C2

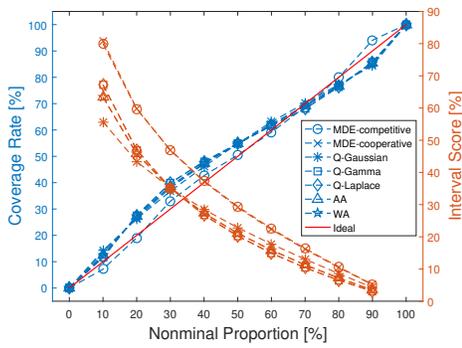
Figure 4.14: 6HA reliability and sharpness comparison of different models at the 7 sites



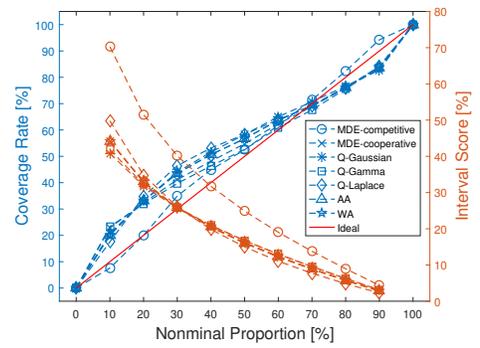
(c) C3



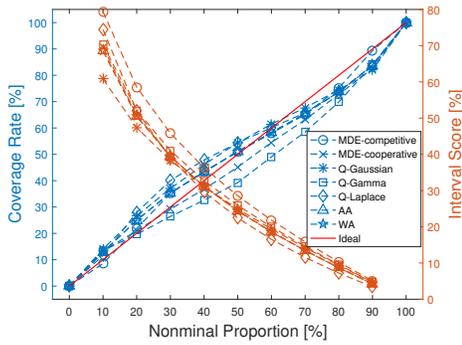
(d) C4



(e) C5

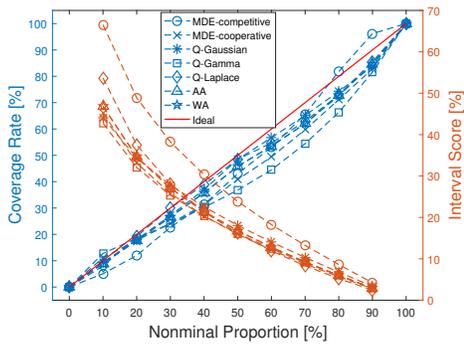


(f) C6

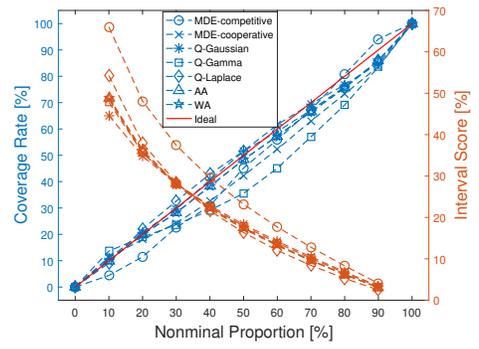


(g) C7

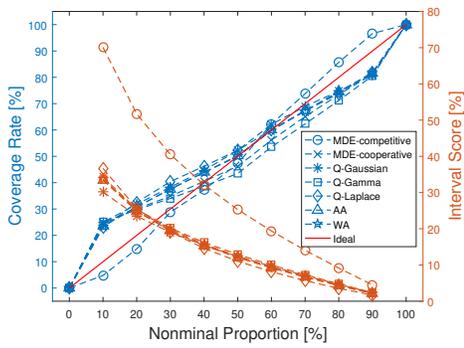
Figure 4.14: (continued) 6HA reliability and sharpness comparison of different models at the 7 sites



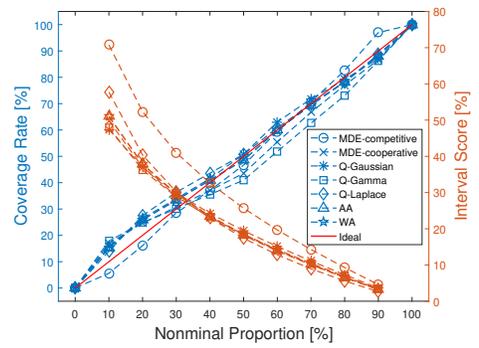
(a) C1



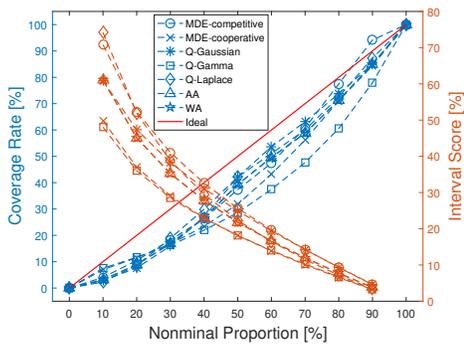
(b) C2



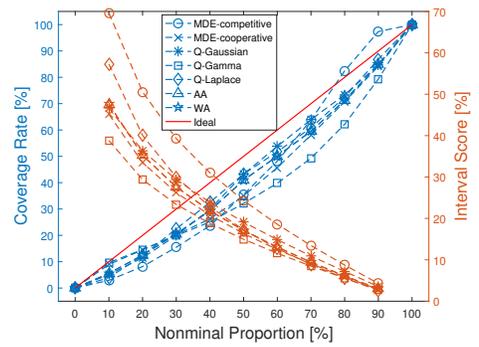
(c) C3



(d) C4

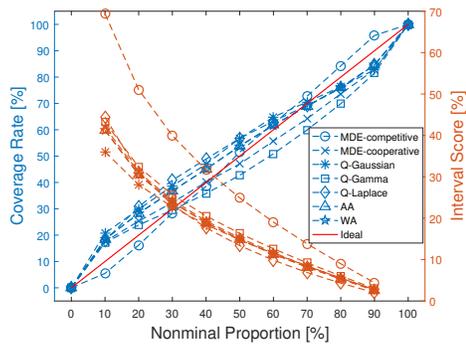


(e) C5



(f) C6

Figure 4.15: 24HA reliability and sharpness comparison of different models at the 7 sites



(g) C7

Figure 4.15: (continued) 24HA reliability and sharpness comparison of different models at the 7 sites

## CHAPTER 5

### IMPROVING PROBABILISTIC RENEWABLE FORECASTING VIA SPATIAL-TEMPORAL CORRELATION MODELING

Leveraging both temporal and spatial correlations to predict wind power remains one of the most challenging and less studied areas of deterministic wind power prediction. Instead of deterministic wind power forecasting, probabilistic forecasting models have been adopted in the literature to generate wind power forecast scenarios and provide insights on the uncertainty of forecasts. Moreover, for large-scale wind farms, modeling aggregated wind power takes the advantage of fast computation and simplified implementation compared to detailed modelling of single wind farm.

In this chapter, the problem of aggregated probabilistic wind power forecasting of multiple sites is investigated by considering the spatial-temporal correlation. A Q-learning enhanced deterministic wind power forecasting method is used to generate deterministic wind power forecasts for individual wind farms. The spatial-temporal correlation between the member wind farms and the aggregated wind power is modeled by using a joint distribution model based on the copula theory. The marginal distributions of actual aggregated wind power and forecasted power of member wind farms are built from Gaussian mixture models. Then, a conditional distribution of the aggregated wind power is deduced through the Bayesian theory, which is used for aggregated probabilistic forecasts. The effectiveness of the proposed aggregated probabilistic wind power forecasting framework is validated by using the Wind Integration National Dataset Toolkit. Numerical results of case studies at nine locations show that the developed aggregated probabilistic forecasting methodology has improved the pinball loss metric score by up to 54% compared to three benchmark models.

## 5.1 Motivation and Objective

Studies have shown that the integration of geographically dispersed wind farms could reduce extreme power output, which is referred to as smoothing effect (Tastu et al., 2011). In addition, power produced from one wind farm at different times is typically temporally correlated (Malvaldi et al., 2017). It would be interesting to explore the impacts of spatial-temporal correlation on the performance of aggregated wind power forecasting. The benefits of spatial-temporal modeling for wind power forecasting at aggregated levels have been briefly discussed in (Lenzi et al., 2018). In addition to wind power, spatial-temporal correlation modeling has also been applied to wind speed forecasting (Khodayar and Wang, 2018), load forecasting (Zhao et al., 2018; Mohan et al., 2018), and solar power forecasting (Agoua et al., 2018). Therefore, in this study, we seek to develop an aggregated probabilistic wind power forecasting method by considering spatial-temporal correlation among wind farms with improved marginal distribution modeling of wind power and clustering. Specifically, a Copula model is adopted to build a spatial-temporal correlated joint model between the aggregated wind power and forecasted wind power of each wind farm, and the conditional distribution of aggregated wind power is used to generate aggregated probabilistic wind power forecasts.

## 5.2 Methodology

The overall framework of the developed conditional probabilistic aggregated wind power forecasting (cp-AWPF) framework is illustrated in Fig. 5.1. The five major steps are briefly described as follows:

1. Step 1: A Q-learning based ensemble deterministic forecasting method is adopted to select the best forecasting model from a pool of state-of-the-art machine learning based forecasting models at each time step, thus generating deterministic wind power forecasts for individual wind farms.

2. Step 2: GMM is used to fit the probability density functions (PDFs) of the historical aggregated actual wind power and the wind power forecasts at each wind farm.
3. Step 3: A prototype-based k-means clustering method is used to cluster wind farms into multiple non-overlapping clusters based on their spatial similarity.
4. Step 4: For each cluster, the joint distribution of historical aggregated actual wind power and wind power forecasts at each member wind farm is constructed based on the Copula theory.
5. Step 5: Aggregated probabilistic wind power forecasts are generated through the conditional distribution.

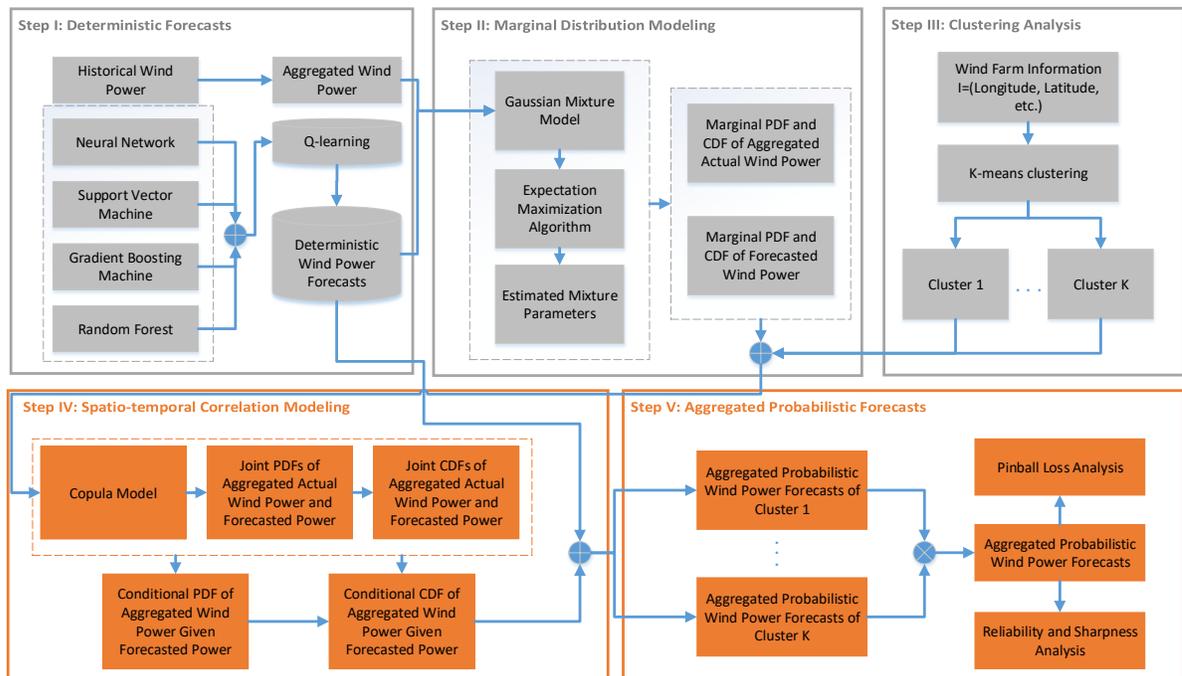


Figure 5.1: Overall framework of the conditional probabilistic aggregated wind power forecasting

### 5.2.1 Q-learning Enhanced Deterministic Forecasting

The developed spatial-temporal correlation based cp-AWPF framework is constructed based on deterministic forecasts. A large collection of methods have been developed in the literature to effectively generate deterministic wind forecasts. However, most of existing deterministic methods are either selected based on the overall forecasting performance or ensembled by multiple models. Selecting a model based on the overall forecasting performance generally neglects the local performance of the selected model.

In this study, a Q-learning enhanced deterministic forecasting method, developed in our previous work (Feng and Zhang, 2019), is adopted. This method can choose the best forecasting model from a pool of state-of-the-art machine learning based forecasting models (i.e., artificial neural network, support vector machine, gradient boosting machine, and random forest) at each time step. To be more specific, the developed method trains Q-learning agents based on the rewards of transferring from the current model to the next model. For example, a Q-learning agent will receive a reward by transferring from the current forecasting model  $M_i$  to the next forecasting model  $M_j$  in each training step, from which the Q-learning agent will learn the optimal policy of the model selection. Then, this optimal policy will be applied to select the best model for forecasting in the next step based on the current model in the forecasting stage. The dynamic model selection process is expressed as (Feng and Zhang, 2019):

$$S = \{s\} = \{s_1, s_2, \dots, s_I\} \quad (5.1)$$

$$A = \{a\} = \{a_1, a_2, \dots, a_I\} \quad (5.2)$$

$$R^t(s_i, a_j) = \text{ranking}(M_i) - \text{ranking}(M_j) \quad (5.3)$$

$$Q^{(e+1)}(s^e, a^e) = (1 - \alpha)Q^e(s^e, a^e) + \alpha[R^e(s^e, a^e) + \gamma \max_{a \in A} Q^e(s^{(e+1)}, a)] \quad (5.4)$$

where  $S$ ,  $A$ ,  $R$ , and  $Q$  are state space, action space, reward function, and Q-table in the dynamic model selection Markov Decision Process, respectively.  $s$  and  $a$  are possible state and action,

respectively.  $I$  is the number of models ( $M$ ) in the model pool.  $e$  is the episode index with the maximum of 100.  $\alpha = 0.1$  is the learning rate that controls the aggressiveness of learning.  $\gamma = 0.8$  is a discount factor that weights the future reward. The reward function is defined as the model performance improvement, which ensures the effective and efficient convergence of Q-learning. More details about the Q-learning enhanced deterministic forecasting can be found in Ref. (Feng and Zhang, 2019).

### 5.2.2 Wind Power Distribution Modeling

The large variability in wind power imposes challenges to accurately model the wind power through a unimodal distribution. Mixture distributions have been widely utilized in statistics to approximate multi-modal distributions. To accurately characterize the variability of wind power, GMM is adopted in this study to model the aggregated actual wind power and forecasted wind power. The probabilistic density function (PDF) of GMM is formulated as follows:

$$f_G(x|N_G, \omega_i, \mu_i, \sigma_i) = \sum_{i=1}^{N_G} \omega_i g_i(x|\mu_i, \sigma_i) \quad (5.5)$$

where  $N_G$  is the number of mixture components,  $U(\mu_i \in U)$  is the expected value vector,  $\Sigma(\sigma_i \in \Sigma)$  is the standard deviation vector, and the  $\Omega(\omega_i \in \Omega)$  is the weight vector. Each component  $g(x; \mu_i, \sigma_i)$  follows a normal distribution, which can be expressed as:

$$g(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5.6)$$

The GMM distribution has two constraints: (1) the integral of Eq. 5.5 equals unity, and (2) the summation of weight parameters equals unity as well, which are expressed as follows:

$$\int_{-\infty}^{+\infty} f_G(x|N_G, \omega_i, \mu_i, \sigma_i) dx = \int_{-\infty}^{+\infty} \sum_{i=1}^{N_G} \omega_i g_i(x|\mu_i, \sigma_i) dx = 1 \quad (5.7)$$

$$\sum_{i=1}^{N_G} \omega_i = 1 \quad (5.8)$$

The parameters of GMM are estimated through the expectation maximization (EM) algorithm. The goal of EM is to maximize the likelihood function with respect to the parameters. More details about EM can be found in (Hartley, 1958). The CDF ( $\hat{F}$ ) corresponding to the estimated PDF is expressed as:

$$F_G(x|N_G, \omega_i, \mu_i, \sigma_i) = \sum_{i=1}^{N_G} \left[ \frac{\sqrt{\pi}}{2} \omega_i \sigma_i \operatorname{erf}\left(\frac{\mu_i - x}{\sigma_i}\right) \right] + C \quad (5.9)$$

where  $C$  is an integral constant, and  $\operatorname{erf}(\cdot)$  is the Gaussian error function which is expressed as:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (5.10)$$

### 5.2.3 Clustering Analysis

Clustering unlabelled wind farms is an unsupervised problem, which distinguishes and labels the type of training data based on the characteristics of the data itself. In this study, we aim to reduce the dimensionality due to the large number of wind farms by dividing wind farms into multiple clusters, thereby reducing the computational cost. In addition, wind farm clustering could also strengthen the intra-cluster correlation, i.e., wind farms in the same cluster are more strongly correlated, which could potentially increase the accuracy of the aggregated wind power forecasting for each cluster and thus the overall forecasting accuracy.

### 5.2.4 Unsupervised Clustering based on Spatial Correlation

In this study, K-means is adopted to divide the member wind farms into multiple non-overlapping clusters based on wind farm characteristics. K-means is a widely used unsupervised clustering

algorithm. Given a dataset  $\psi = x_1, x_2, \dots, x_n$  with  $n$  instances, the K-means algorithm divides the dataset into  $K$  disjoint clusters as  $\psi = C_1, C_2, \dots, C_k$ . The objective function of K-means is:

$$\min \sum_{k=1}^K \sum_{x_k} d(x, m_k) \quad (5.11)$$

where  $m_k$  is the centroid of cluster  $C_k$ ,  $d(\cdot)$  measures the distance between  $x$  and the centroid of cluster  $C_k$ . In this study, the Pearson's correlation coefficient is used as a distance metric to characterize the spatial-temporal correlation of wind farms.

### Clustering Assessment Metric

It is challenging to validate unsupervised clustering results since the objects are unlabeled. Unsupervised clustering can be measured quantitatively based on cohesion and separation (Arbelaitz et al., 2013). Cohesion measures how closely related the objects are in a cluster, while Separation measures how distinct or well-separated a cluster (or centroid) is from other clusters. In this study, the Silhouette width (SW) is used to evaluate the clustering performance, which quantifies both cohesion and separation. It can be expressed as:

$$SW = \frac{1}{N} \sum_{i=1}^N \frac{\eta_b(i) - \eta_a(i)}{\max(\eta_b(i) - \eta_a(i))} \quad (5.12)$$

where  $\eta_a(i)$  is the average distance between object  $i$  and all other data in the same cluster, and  $\eta_b(i)$  is the smallest average distance between object  $i$  to other objects in the neighbour cluster. The SW value ranges from -1 to +1, where  $SW = +1$  indicates desired clustering, while  $SW = -1$  indicates undesired clustering.

### 5.2.5 Wind Farms spatial-temporal Correlation Modeling

Once the marginal wind power distribution of a single wind farm is defined, the spatial-temporal correlation among wind farms in each cluster can be modeled by Copula. Copula is one of the

most widely used methods for modeling the dependency among random variables. Given a cluster, the aggregated actual wind power  $p^\Sigma$  is expressed as

$$p^\Sigma = \sum_{i=1}^N p_i \quad (5.13)$$

where  $p_i$  is the actual wind power of the  $i$ th wind farm, and  $N$  is the total number of wind farms to be aggregated in the cluster.

Suppose  $\hat{p}_i^t$  is the  $t$ -hour-ahead (tHA) forecasted wind power at the  $i$ th wind farm,  $f(\hat{p}_i^t)$  and  $F(\hat{p}_i^t)$  denote the corresponding marginal PDF and marginal cumulative distribution function (CDF), respectively. The spatial correlation  $\Theta_S$  among the  $N$  wind farms is modeled as:

$$F(\hat{P}^t) = C(F(\hat{p}_1^t), \dots, F(\hat{p}_N^t)) \quad (5.14)$$

where  $\hat{P}^t = \{\hat{p}_1^t, \hat{p}_1^t, \dots, \hat{p}_N^t\}$  is a  $N$ -dimension vector denoting the  $t$ HA wind power forecasts of the  $N$  wind farms, and  $C(\cdot)$  is the Copula function.

In addition to the spatial correlation among wind farms, the forecasted wind power with different look ahead times are typically temporally correlated (Wang et al., 2018). Such temporal correlation  $\Theta_T$  is modeled as:

$$F(\hat{P}^t, \hat{P}^{t+1}) = C(F(\hat{p}_1^t), \dots, F(\hat{p}_N^t), F(\hat{p}_1^{t+1}), \dots, F(\hat{p}_N^{t+1})) \quad (5.15)$$

where,  $\hat{P}^{t+1} = \{\hat{p}_1^{t+1}, \hat{p}_1^{t+1}, \dots, \hat{p}_N^{t+1}\}$  is a  $N$ -dimension vector denoting the  $(t+1)$ HA wind power forecasts of the  $N$  wind farms.

Based on Eqs. 5.13–5.15, the joint CDF of the forecasted wind power at individual wind farms and the aggregated actual wind power (of all farms),  $F(p^\Sigma, \hat{P}^t, \hat{P}^{t+1})$  can be modeled through their marginal CDFs and the Copula function with a spatial-temporal joint structure  $\Theta_{S-T}$ , which is expressed as:

$$\begin{aligned} F(p^\Sigma, \hat{P}^t, \hat{P}^{t+1}) &= C(p^\Sigma, F(\hat{p}_1^t), \dots, F(\hat{p}_N^t), F(\hat{p}_1^{t+1}), \dots, F(\hat{p}_N^{t+1})) \\ &= C(p^\Sigma, J, K) \end{aligned} \quad (5.16)$$

where,

$$J = F(\hat{p}_1^t), \dots, F(\hat{p}_N^t) \quad (5.17)$$

$$K = F(\hat{p}_1^{t+1}), \dots, F(\hat{p}_N^{t+1}) \quad (5.18)$$

Similarly, the joint PDF of the forecasted wind power at individual member farms and the aggregated actual wind power (of all farms) is expressed as:

$$f(p^\Sigma, \hat{P}^t, \hat{P}^{t+1}) = c(F(p^\Sigma), J, K) \cdot f(p^\Sigma) \cdot \prod_{i=1}^N f(\hat{p}_i^t) \quad (5.19)$$

where the marginal PDF is modeled by using the aforementioned GMM distribution based on historical actual and forecasting data. Then, the conditional joint PDF of the aggregated wind power given the power forecasts of all the member farms is deduced from the Bayesian formula, given by:

$$f(p^\Sigma | \hat{P}^t, \hat{P}^{t+1}) = \frac{c(F(p^\Sigma), J, K)}{c(J, K)} \cdot f(p^\Sigma) \quad (5.20)$$

The conditional distribution of the aggregated wind power given all the member farms forecasts can be trained by using historical actual and forecasting data. With any given deterministic forecasts of individual wind farms, we can use the copula model and the trained conditional PDF in Eq. 5.20 to calculate the conditional CDF. Then a large number of scenarios of the aggregated wind power of each cluster can be generated by sampling from the conditional CDF. The PDF of the aggregated wind power in each cluster can be deduced based on the scenarios. The final PDF of all aggregated wind farms is calculated through convolution. One prerequisite of convolution is that the variables being convoluted should be independent. If the clusters are cross-correlated with each other, a copula based dependent convolution (such as the one proposed by Zhang et al. (Zhang et al., 2016)) could be used to characterize the cross-correlation among clusters. The probabilistic forecasts could also be represented in the form of quantiles and confidence intervals based on the

PDF. The pseudocode of quantile forecasts based on inverse transform is illustrated in Algorithm 2.

---

**Algorithm 2:** Generate quantile forecasts based on inverse transform

---

**Data:** Deterministic wind power forecasts

**Result:** Quantile forecasts

- 1 Initialization: Obtain PDF of each single wind farm and aggregated wind power through GMM;
  - 2 Model spatial-temporal correlation among wind farms through a joint distribution by Copula;
  - 3 Calculate conditional PDF and CDF of the aggregated wind power;
  - 4 Sample from conditional CDF through inverse transform to generate a large number of aggregated wind power scenarios;
  - 5 Generate quantile forecasts based on the distribution of generated scenarios;
- 

### 5.3 Case Studies and Results

The developed cp-AWPF framework was evaluated at 9 wind farms in Texas that were selected from the Wind Integration National Dataset (WIND) Toolkit (Draxl et al., 2015). The WIND Toolkit includes meteorological information (e.g., wind direction, wind speed, air temperature, surface air pressure, density at hub height), synthetic actual wind power, and wind power forecasts generated by the Weather Research and Forecasting (WRF) model. It covers over 126,000 locations in the United States. In addition to day-ahead (DA) forecasts from WIND Toolkit, very-short-term forecasts (i.e., 1HA to 6HA) are generated by using the Q-learning enhanced deterministic forecasting method. In this study, the duration of the collected data at the selected 9 wind farms spans two years from January 1<sup>st</sup> 2011 to December 31<sup>st</sup> 2012. The data information at the selected 9 wind farms is briefly summarized in Table 5.1. For all the 9 locations, the first 3/4 of the data is used as training data. The number of scenarios generated from the conditional distribution is set as  $N_s=5,000$ . The accuracy of the forecasts is evaluated by the remaining 1/4 of data. Though the

developed cp-AWPF method is capable of generating forecasts at multiple forecasting horizons, only 1HA-6HA and day-ahead wind power forecasts are generated in this study.

Table 5.1: Data summary of the selected 9 WIND Toolkit sites

Site Name	Site ID	Lat.	Long.	Capacity (MW)	State
S1	4816	29.38	-100.37	16	TX
S2	8979	31.53	-95.62	16	TX
S3	10069	32.31	-98.26	16	TX
S4	10526	32.44	-100.55	16	TX
S5	1342	27.12	-97.86	16	TX
S6	2061	27.95	-97.40	14	TX
S7	9572	31.99	-100.118	16	TX
S8	10527	32.45	-100.41	2	TX
S9	11038	32.72	-100.92	6	TX

Normalized indices of standard metrics like root mean squared error and mean absolute error, *i.e.*, NRMSE and NMAE, are adopted to evaluate the performance of Q-learning enhanced deterministic forecasts. The forecasting errors by using the Q-learning based deterministic forecasting model at the selected locations are summarized in Table 5.2. It is shown that the 1HA NMAE and NRMSE are in the ranges of 5%-8% and 8%-12%, respectively. Day-ahead deterministic forecasts are provided in the WIND Toolkit dataset, which are generated from numerical weather prediction (NWP) models (Zhang et al., 2015). It is shown that the day-ahead NMAE and NRMSE are in the range of 11%-14% and 15%-19%, respectively. Overall, the accuracies of the Q-learning based 1HA to 6HA deterministic forecasts and the NWP-based day-ahead deterministic forecasts are reasonable.

Fig. 5.2(a) shows the marginal probability distributions of wind power forecasts from six distribution types at the S4 site (*i.e.*, Gaussian, Gamma, Logistic, Rayleigh, KDE, and GMM distributions). For the Gaussian, Gamma, Logistic, and Rayleigh distributions, the parameters are estimated through the maximum likelihood (ML) method. The parameters of the KDE and GMM distributions are estimated by using the expectation maximization (EM) algorithm. Fig. 5.2(b) illustrates the mixture components of the GMM distribution at the S4 site. The Akaike

Table 5.2: Deterministic forecasting results by using Q-learning and NWP

Model	LAT	Metric	Site								
			S1	S2	S3	S4	S5	S6	S7	S8	S9
QL	1HA	NMAE(%)	6.63	6.85	6.70	6.74	6.67	5.38	7.76	6.85	7.40
		NRMSE(%)	10.55	10.93	11.04	10.66	9.93	8.37	11.93	10.72	11.51
	2HA	NMAE(%)	10.72	11.08	11.20	11.28	10.36	8.34	10.90	11.15	11.92
		NRMSE(%)	15.86	16.33	16.93	16.64	14.60	12.31	17.14	16.31	17.32
	3HA	NMAE(%)	13.95	14.22	14.65	14.88	12.44	10.49	14.56	14.27	14.82
		NRMSE(%)	19.43	19.86	20.92	20.71	16.94	14.93	20.12	19.75	20.50
	4HA	NMAE(%)	16.43	16.76	17.69	18.00	13.84	12.17	16.28	17.16	16.87
		NRMSE(%)	21.98	22.22	23.92	23.83	18.56	16.85	21.79	22.68	22.41
	5HA	NMAE(%)	18.19	18.37	20.36	20.38	15.35	13.53	17.87	19.33	18.43
		NRMSE(%)	23.65	23.71	26.51	26.01	20.19	18.40	23.15	24.66	24.03
	6HA	NMAE(%)	19.85	19.60	21.94	22.63	16.51	14.70	18.41	20.93	19.60
		NRMSE(%)	25.09	24.85	27.82	27.82	21.31	19.47	23.55	26.03	24.78
NWP	DA	NMAE(%)	12.70	12.59	13.21	13.97	13.97	11.63	13.37	13.44	13.45
		NRMSE(%)	16.85	17.44	18.07	18.70	18.43	15.41	18.17	18.43	18.21

*Note:* QL is the abbreviation for Q-learning. LAT is the abbreviation for look-ahead time. DA: day-ahead

information criterion (AIC) and the Bayesian information criterion (BIC) are used to evaluate the estimated wind power distribution accuracy. For a given model with parameters  $\theta$ , the AIC and BIC are defined as:

$$AIC = -2L(\hat{\theta}|x) + 2k \quad (5.21)$$

$$BIC = -2L(\hat{\theta}|x) + k \log n \quad (5.22)$$

where  $\hat{\theta}$  is the maximum likelihood estimation of the distribution parameter,  $L(\hat{\theta}|x)$  is the log-likelihood function of  $\hat{\theta}$ ,  $k$  is the number of parameters, and  $n$  is the length of observed data. The preferred model is the one with the lowest AIC and BIC (Burnham and Anderson, 2003). The AIC and BIC values of the nine wind farms with different distribution types are summarized in Table 5.3. Results show that the GMM distribution outperforms other single distributions and KDE for modeling wind power forecasts. Therefore, GMM is adopted to fit the marginal wind power distribution within the Copula model.

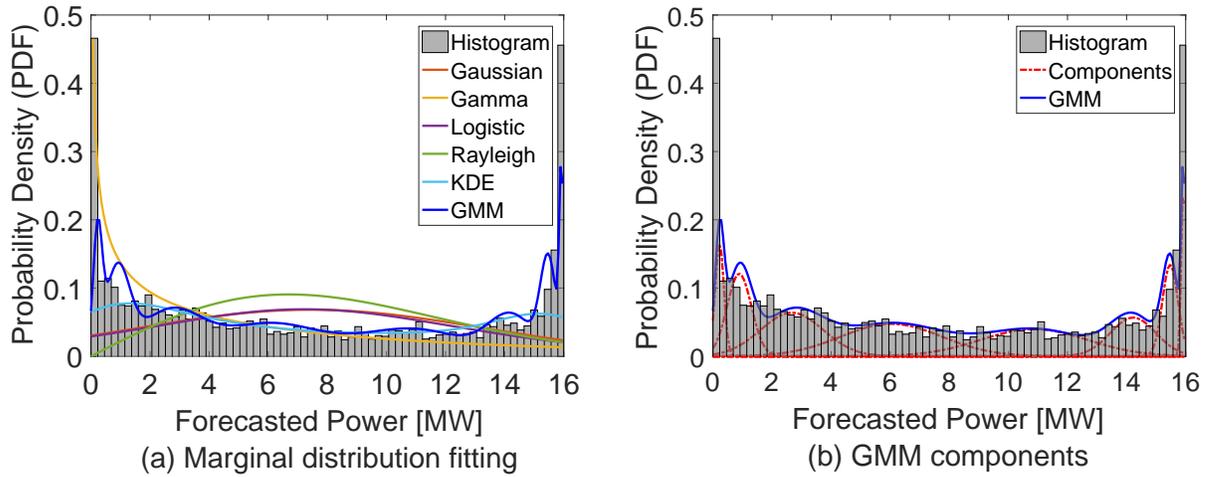


Figure 5.2: Distributions of wind power forecasts at the S4 site

Table 5.3: Information criteria of the estimated distribution

Site	Metric	Distribution type					
		Gaussian	Gamma	Logistic	Rayleigh	KDE	GMM
S1	AIC	23760	23910	24070	23810	21367	<b>18512</b>
	BIC	23768	23918	24078	23818	21374	<b>18548</b>
S2	AIC	23840	22980	24080	24210	20081	<b>16802</b>
	BIC	23848	22988	24088	24218	20105	<b>16846</b>
S3	AIC	27450	26010	27960	29910	21900	<b>17664</b>
	BIC	27458	26018	27968	29918	21908	<b>17708</b>
S4	AIC	27140	25940	27620	29130	22180	<b>17026</b>
	BIC	27148	25948	27628	29138	22196	<b>17074</b>
S5	AIC	23870	23776	24160	23871	20124	<b>18442</b>
	BIC	23878	23784	24168	23819	20136	<b>18498</b>
S6	AIC	22360	21090	21610	22000	18900	<b>16430</b>
	BIC	22368	21098	21618	22008	18928	<b>16496</b>
S7	AIC	23610	23540	23920	24200	21103	<b>18900</b>
	BIC	23618	23548	23928	24208	21119	<b>18948</b>
S8	AIC	7338	6987	7699	7307	6692	<b>5483</b>
	BIC	7346	6995	7677	7315	6708	<b>5514</b>
S9	AIC	15660	15280	15970	15660	13214	<b>12090</b>
	BIC	15668	15288	15978	15668	13228	<b>12136</b>

*Note:* The best information criterion at each location is in boldface.

The correlation among the selected 9 wind farms are visualized in Fig. 5.4(a). It is seen that the selected wind farms are strongly correlated. Fig. 5.4(b) shows the SW among all the wind farms in year 2011 and year 2012. It is seen that  $K = 2$  always results in the highest SW, which indicates desired clustering. To better visualize the two clusters, the wind farms are plotted on a Texas map according to their geographical dispersion in Fig. 5.3. The two wind farm clusters are differentiated by red and blue colors. The red cluster is mainly located in north Texas and the blue cluster is located in south Texas. The mean correlation coefficient in the red cluster and the blue cluster are 0.65 and 0.58, respectively. In addition, the mean correlation coefficient without clustering is 0.54, which is lower than both of the intra-cluster correlation, further showing the effectiveness of clustering.

Within each cluster, a joint distribution of the aggregated wind power and the forecasted power of member wind farms is determined. Then the conditional distribution of the aggregated wind power can be calculated, which is used to generate aggregated wind power forecasting scenarios (e.g., 5,000). The quantiles of the aggregated wind power are calculated based on the empirical distribution of the generated scenarios. To evaluate the performance of cp-AWPF, three baseline models are selected for comparison, which are: quantile regression (QR), cp-AWPF without clustering (cp-AWPF-W/), and a data-driven two-step aggregated probabilistic wind power forecasting method with clustering (dd-AWPF). The reasons for choosing these baseline models are: (i) QR is a widely used method in probabilistic forecasting, which allows us to explore the forecasting enhancement by considering spatial-temporal correlation; (ii) since a clustering model is included in cp-AWPF, it is important to compare the accuracy of the proposed cp-AWPF method with the method without clustering; (iii) for the baseline model of dd-AWPF, the forecasting error distribution could be fitted through GMM based on historical deterministic aggregated wind power forecasts. Then, Monte Carlo sampling is used to generate a large number of forecasting error scenarios. These aggregated error scenarios could be used together with deterministic aggregated wind power forecasts to generate probabilistic aggregated wind power forecasts. Details of the GMM fitting and scenario generation for dd-AWPF could be found in (Cui et al., 2017).

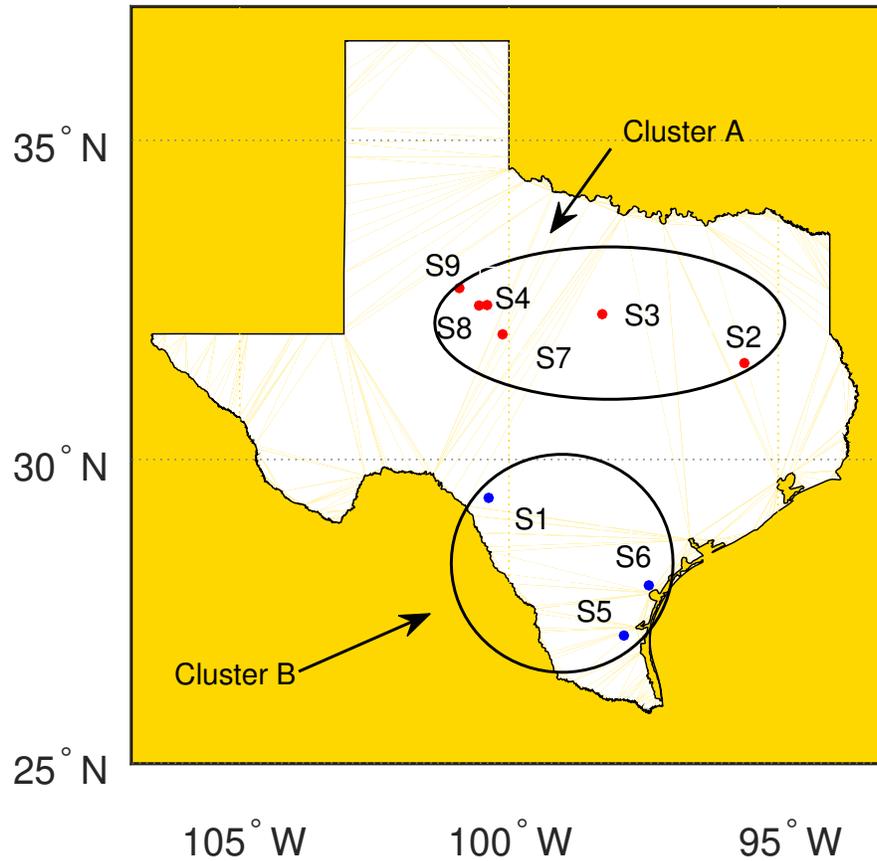


Figure 5.3: Locations of clustered wind farms

Fig. 5.5(a) shows the day-ahead aggregated probabilistic wind power forecasts of all the 9 wind farms from 2012-07-05 to 2012-07-08, which are generated from the proposed cp-AWPF model. It is observed that at the entire representative period, the aggregated wind power reasonably lies within the PIs. Fig. 5.6(b), Fig. 5.6(c), and Fig. 5.6(d) show the aggregated probabilistic forecasts generated from the baseline cp-AWPF-W/, dd-AWPF, and QR methods, respectively. It is seen that the PIs of the probabilistic forecasting with clustering in Fig. 5.5(a) is narrower than the PIs without clustering in Fig. 5.6(b). This is due to that clustering has enhanced the intra-cluster correlation among wind farms. In addition, the PIs of both cp-AWPF and cp-AWPF-W/ are significantly narrower than those of QR in Fig. 5.6(d), which is due to the consideration of spatial-

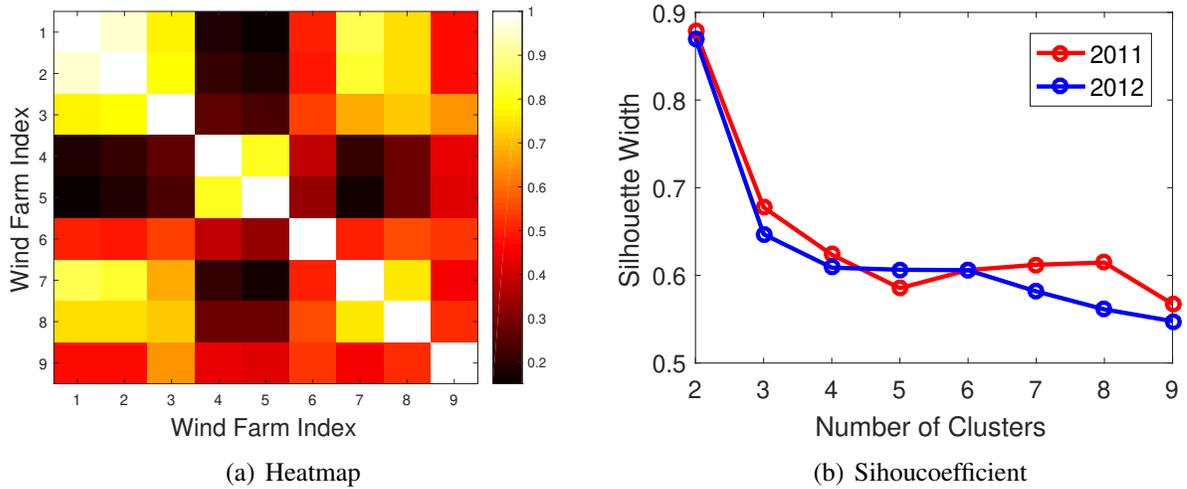


Figure 5.4: Clustering analysis

temporal correlation. The dd-AWPF has a similar width of PIs as cp-AWPF-W/. Nevertheless, the PIs of cp-AWPF-W/ are smoother than those of dd-AWPF, which indicates a stable and reliable probabilistic forecast. It is also observed that the width of the PIs varies with the variability of the aggregated wind power. For example, when the wind power fluctuates more frequently, the PI tends to be wider, and thereby the uncertainty in wind power forecasts is relatively higher.

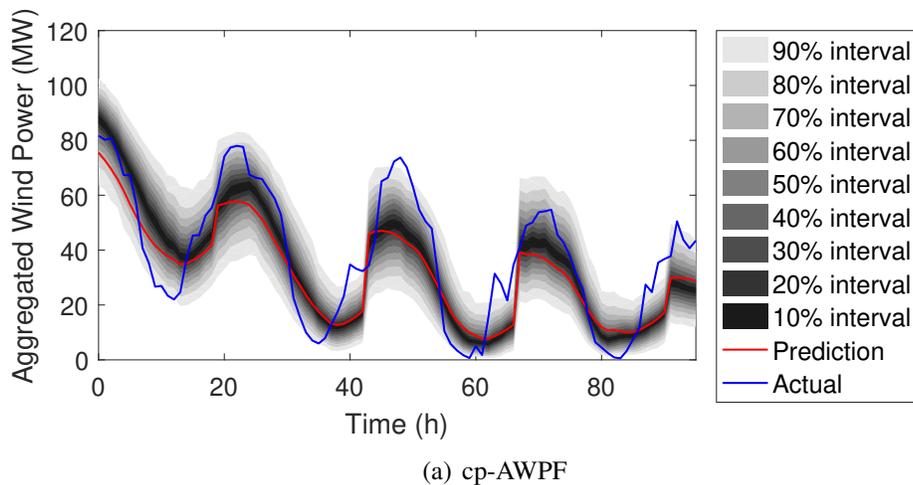
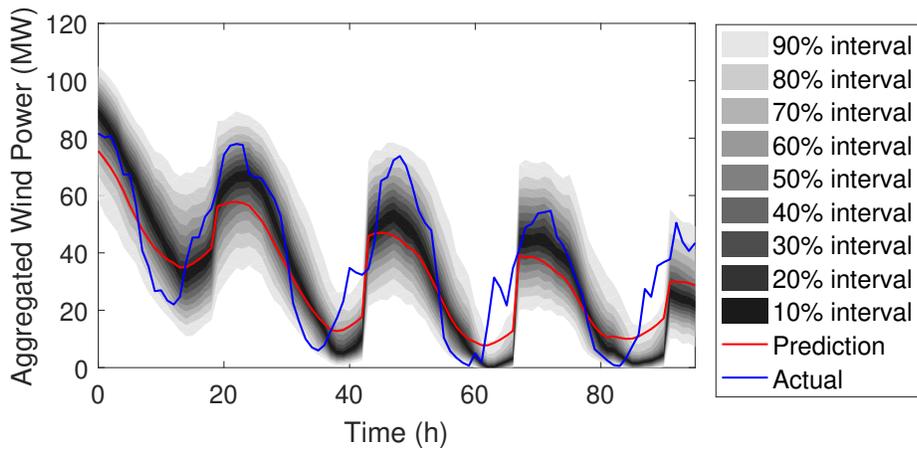
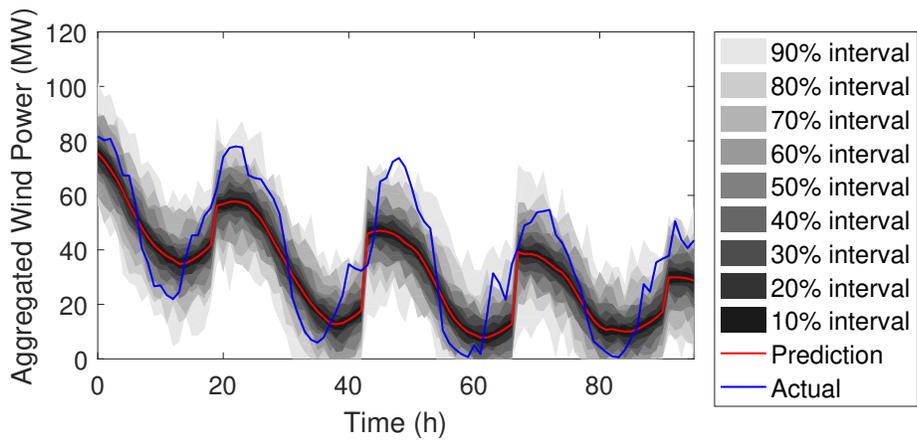


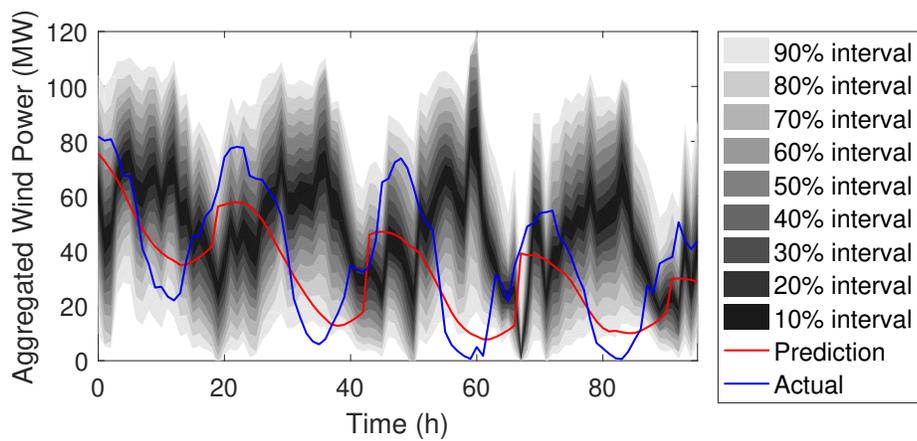
Figure 5.5: Day-ahead aggregated probabilistic wind power forecasts



(b) cp-AWPF-W/



(c) dd-AWPF



(d) QR

Figure 5.5: (continued) Day-ahead aggregated probabilistic wind power forecasts

To show the effectiveness of the developed spatial-temporal based probabilistic forecasting framework, the normalized pinball loss values of different models with different look-ahead time are compared in Table 5.4. The sum of pinball loss is averaged over all quantiles from 1% to 99% and normalized by the aggregated wind farm capacity. A lower pinball loss score indicates a better probabilistic forecast. Results show that the proposed cp-AWPF has improved the pinball loss by up to 54% compared to the three benchmark models, which validates the effectiveness of spatial-temporal correlation and clustering. Note that the cp-AWPF method has shown a better accuracy than baseline models for both 1HA-6HA and day-ahead forecasts, which validates the robustness of the methodology. Furthermore, cp-AWPF without clustering also has better pinball loss than QR, which shows the improvement from spatial-temporal correlation modeling. The dd-AWPF model outperforms both QR and cp-AWPF-W/, which indicates the enhancement resulted from clustering. In addition to pinball loss, two more standard metrics, i.e., reliability and sharpness, are also calculated to assess the performance of aggregated probabilistic forecasting.

Table 5.4: Normalized pinball loss of different models with different look-ahead hours

Model	Look-ahead Times						
	1HA	2HA	3HA	4HA	5HA	6HA	DA
cp-AWPF	<b>1.91</b>	<b>2.03</b>	<b>2.31</b>	<b>2.81</b>	<b>3.46</b>	<b>4.10</b>	<b>3.28</b>
cp-AWPF-W/	2.92	2.93	3.06	3.40	3.94	4.63	4.26
dd-AWPF	2.20	2.38	2.71	3.19	3.83	4.55	3.47
QR	2.98	3.34	4.76	5.80	6.53	7.02	7.27

*Note:* The smallest normalized pinball loss value is in boldface.

Fig. 5.6 shows the 1HA to 6HA and DA reliability plots of the aggregated probabilistic forecasts with different forecasting models. A forecast presents better reliability when the curve is closer to the diagonal. It is seen from Fig. 5.6 that overall QR has better reliability performance. It is mainly because the PIs of QR are much wider than those of the proposed cp-AWPF method. A wider PI indicates that the result takes more errors into consideration; however, note that the reliability over the 90th confidence interval is similar between cp-AWPF and QR, which

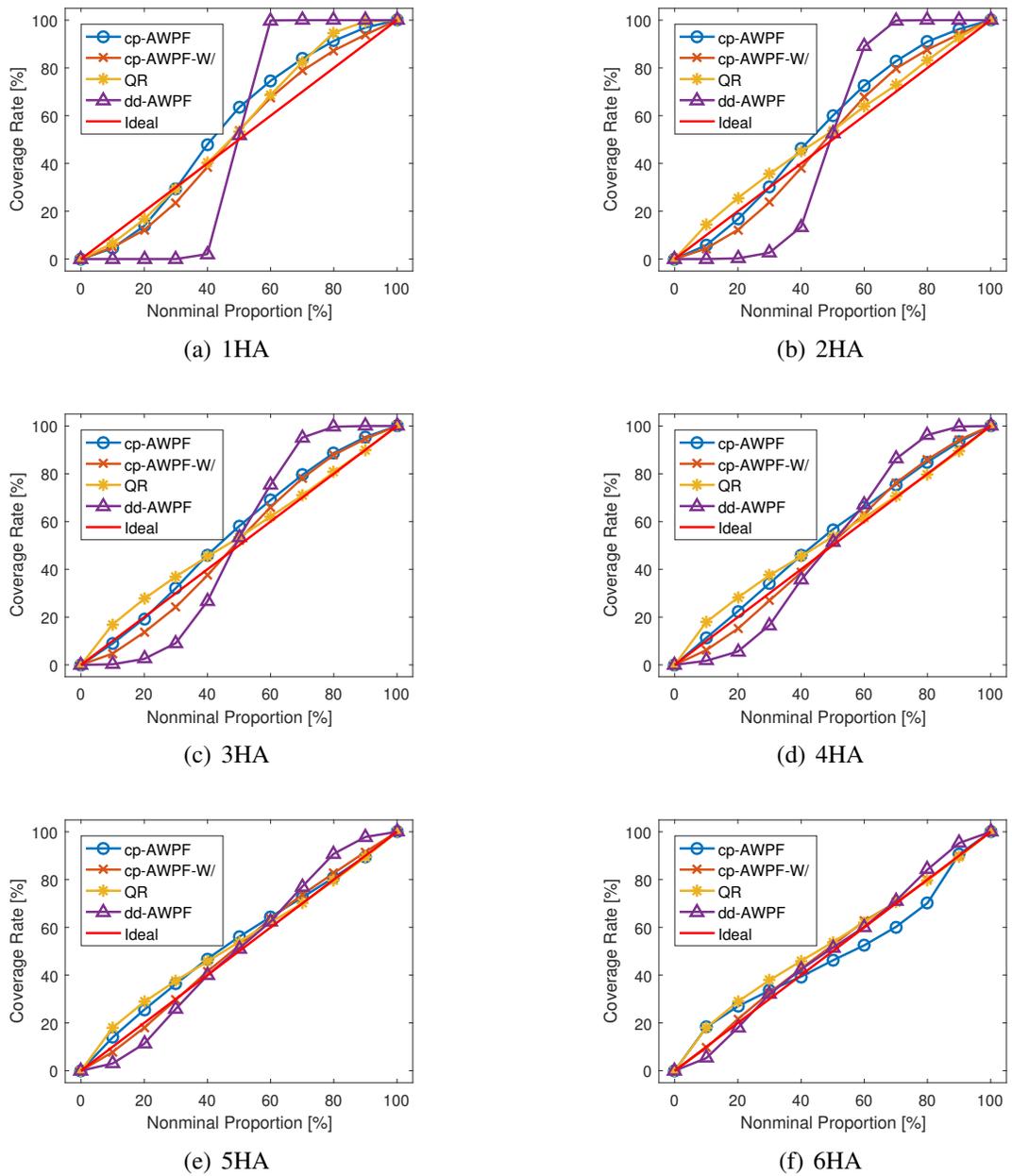
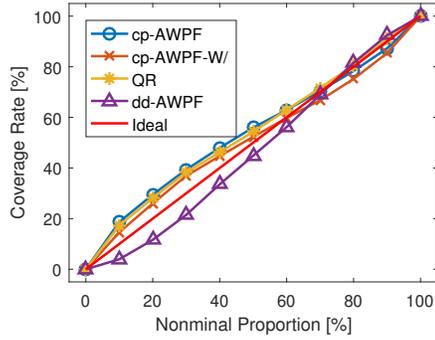


Figure 5.6: Reliability comparison of different models with different look-ahead times

is generally more important in probabilistic forecasting applications. In addition, the proposed cp-AWPF model has shown better reliability than dd-AWPF, indicating the enhancement resulted from spatial-temporal correlation modeling.



(g) DA

Figure 5.6: (continued) Reliability comparison of different models with different look-ahead times

The sharpness plots of cp-AWPF and baseline models (i.e., cp-AWPF-W/, dd-AWPF, and QR) with different look-ahead hours are compared in Fig. 5.7. The expected interval size increases with increasing the nominal coverage rate, and the sharpness of the proposed cp-AWPF model is significantly better than that of the baseline models (i.e., cp-AWPF-W/ and QR) except at the 1HA horizon. It is mainly because the reliability and sharpness are two complementary metrics, and the better reliability of cp-AWPF at 1HA sacrifices the sharpness to some extent. Note that the cp-AWPF model has shown better sharpness than dd-AWPF, indicating the enhancement resulted from spatial-temporal correlation modeling. Overall, the interval size of the proposed cp-AWPF model ranges from 2% to 45%, which indicates low sharpness. In addition, cp-AWPF has significantly better sharpness than cp-AWPF-W/ at all look-ahead times, which validates the effectiveness of clustering.

## 5.4 Conclusion

In this chapter, an aggregated conditional probabilistic wind power forecasting (cp-AWPF) framework was developed by considering spatial-temporal correlation and wind farm clustering. The K-means clustering approach was applied to cluster 9 wind farms into two clusters. For each cluster, GMM was adopted to accurately model the marginal wind power distribution. Then, the

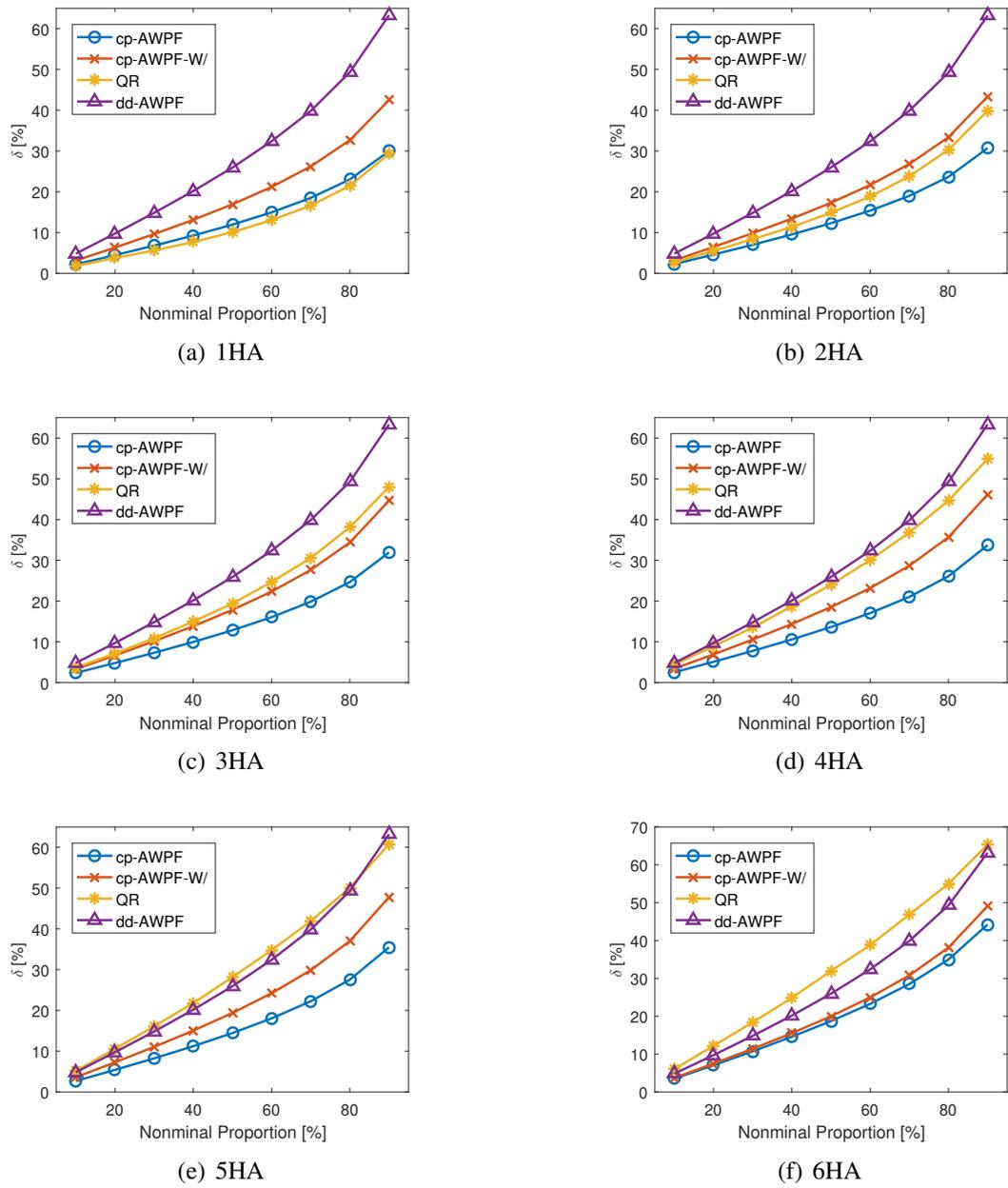
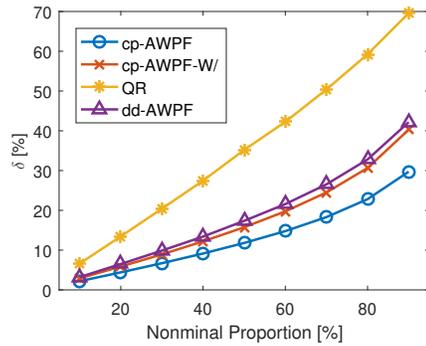


Figure 5.7: Sharpness comparison of different models with different look-ahead times



(g) DA

Figure 5.7: (continued) Sharpness comparison of different models with different look-ahead times. The spatial-temporal correlation between the member wind farms and the aggregated wind power was modeled through a high-dimensional joint distribution based on Copula theory. Inverse sampling was applied on the conditional CDF of the joint distribution to generate aggregated probabilistic forecasts. Results at 9 selected wind farms showed that:

1. cp-AWPF could reduce the pinball loss score by up to 54% compared to three benchmark models.
2. The GMM model has shown better goodness-of-fit to wind power distribution than single-distribution models and KDE.
3. Clustering could enhance intra-cluster correlation among member wind farms, thus providing better probabilistic forecasting accuracy.
4. The developed cp-AWPF framework is robust at different forecasting time horizons and locations.
5. cp-AWPF has shown better sharpness than models without considering spatial-temporal correlation and clustering. The reliability of cp-AWPF is close to the ideal diagonal, which indicates reasonable reliability.

## **CHAPTER 6**

### **IMPROVING PROBABILISTIC RENEWABLE FORECASTING**

#### **VIA SCENARIO GENERATION**

Probabilistic solar power forecasting plays an important role in solar power grid integration and power system operations. One of the most popular probabilistic solar forecasting methods is to feed simulated explanatory weather scenarios into a deterministic forecasting model. However, the correlation among different explanatory weather variables are seldom considered during the scenario generation process.

In previous chapter, the spatial-temporal correlation modeling is used to improve aggregated probabilistic forecasting. However, the correlation between different weather explanatory variables is not quantified, which may affect spatial-temporal modeling. This chapter presents an improved probabilistic solar power forecasting framework based on correlated weather scenario generation. Copula is used to model a multivariate joint distribution between predicted weather variables and observed weather variables. Massive weather scenarios are obtained by deriving a conditional probability density function given a current weather prediction by using the Bayesian theory. The generated weather scenarios are used as input variables to a machine learning-based multi-model solar power forecasting model, where probabilistic solar power forecasts are obtained. The effectiveness of the proposed probabilistic solar power forecasting framework is validated by using seven solar farms from the 2000-bus synthetic grid system in Texas. Numerical results of case studies at the seven sites show that the developed probabilistic solar power forecasting methodology has improved the pinball loss metric score by up to 140% compared to benchmark models.

### **6.1 Motivation and Objective**

PV power output is highly dependent on external weather conditions such as solar radiation and temperature (Mandal et al., 2012). Therefore, it is challenging to get accurate forecasts under

different weather conditions. To better account for the solar power uncertainty and variability, probabilistic solar power forecasts are needed. Quantile regression-based methods and simulating predictors with massive input scenarios, are the most popular used non-parametric methods for probabilistic forecasting (Sun et al., 2020b). For methods with simulating predictors, massive scenarios are generated as inputs of regression models. Weather scenario generation has been widely used in the literature for probabilistic forecasts due to its simplicity and accessibility. Weather scenario generation methods can be generally classified into three categories (Xie and Hong, 2016): (i) fixed-date method, (ii) shifted-date method, and (iii) bootstrap method. The fixed-date method assigns the weather profile of historical years to the current year. The number of scenarios equals to the number of years the weather profile is available. Liu et al. (Liu et al., 2015) used six years historical weather data as input scenarios to a quantile regression averaging model. The shifted-date method generally shifts the historical weather profile with a number of days. Then these shifted weather profiles are treated as weather scenarios of the current year. Bootstrap is a method of computational inference based on resampling a dataset. Breinl et al. (Breinl et al., 2015) adopted a block bootstrap method to generate precipitation and temperatures scenarios. The three weather scenario generation methods mentioned above were compared in (Xie and Hong, 2016) through quantile score, complexity, and number of scenarios based on the GEFCom2014 dataset. Results showed that among the bootstrap, fixed-date, and shifted-date methods, no single method outperforms others in all aspects. Nevertheless, several challenges present in existing methods of weather scenario generation: (i) the number of scenarios generated by the fixed-date and shifted-date methods is limited by the length of the weather profile; (ii) scenarios from bootstrap are heavily relied on the data itself, where bootstrap tends to undervalue extreme weather observations; (iii) the correlation between different weather variables is not considered, which may not be reliable for scenario generation; (iv) high dimensional matrices are involved in traditional weather scenario generation methods, which adds additional computational burden.

In this study, we seek to develop a probabilistic solar power forecasting method based on weather scenario generation, where the correlation among different weather variables are considered.

## 6.2 Methodology

The overall framework of the developed weather scenario generation-based probabilistic solar power forecasting (wsp-SPF) method is illustrated in Fig. 6.1. The two major steps are weather scenario generation and probabilistic solar power forecasting. In each major step, there are several sub-steps which are briefly described as follows:

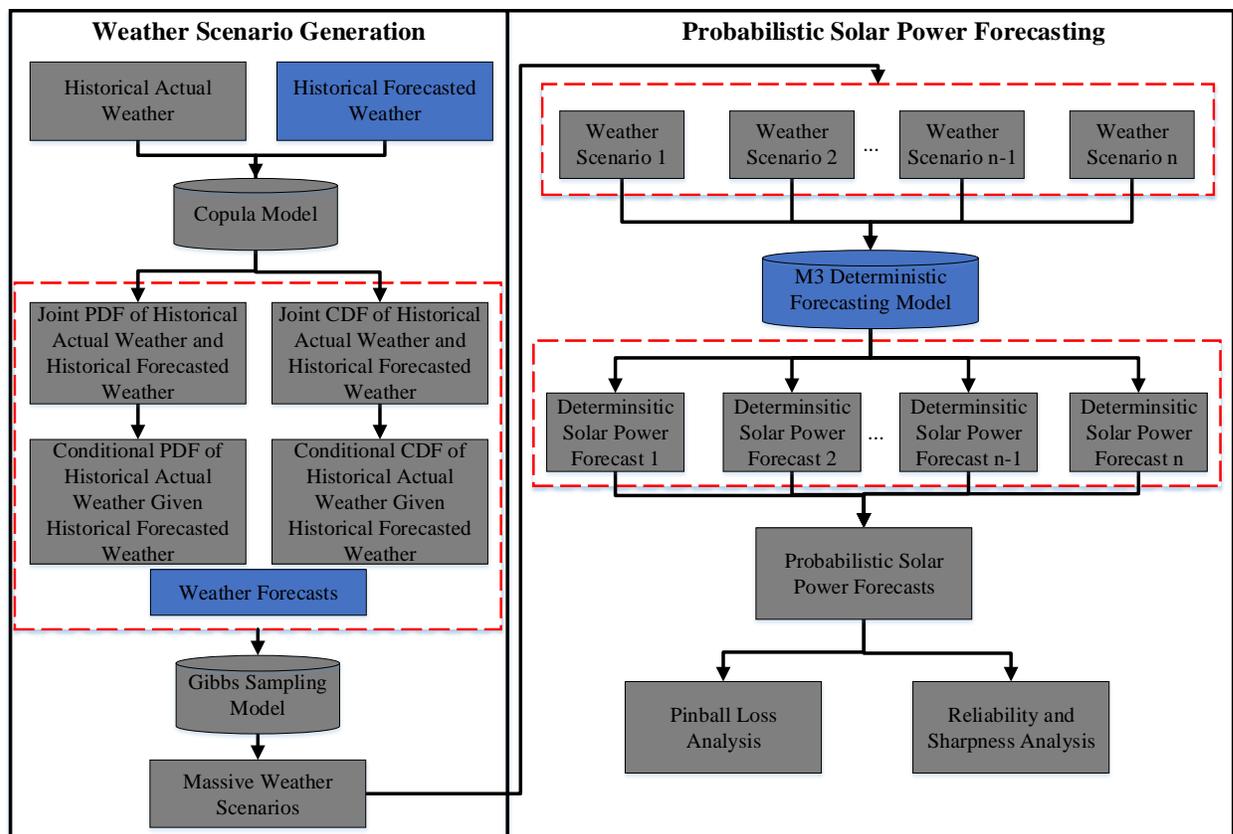


Figure 6.1: Overall framework of weather scenario generation-based probabilistic solar power forecasting

1. Step 1.1: A Machine Learning-based Multi-Model (M3) forecasting framework is adopted to generate short-term deterministic weather forecasts (blue box) (i.e., 1-h-ahead (1HA)), and train deterministic solar power forecasting model (blue box).
2. Step 1.2: Gaussian mixture model (GMM) is used to fit the PDFs of historical actual weather and historical forecasted weather.
3. Step 1.3: The conditional joint distribution of historical actual weather given historical weather forecasts is constructed based on the Copula theory.
4. Step 1.4: A large number of weather scenarios are generated through the conditional distribution via Gibbs sampling.
5. Step 2: The simulated weather predictors are fed into the M3 deterministic solar power forecasting model to generate probabilistic solar power forecasts.

### **6.2.1 Machine Learning-based Multi-Model (M3) Deterministic Forecasting**

In this study, the M3 model developed in our previous work (Feng et al., 2017) is adopted to train deterministic forecasting models (i.e., deterministic weather forecasting model and deterministic solar power forecasting model). M3 is ensembled by multiple models from a pool of state-of-the-art machine learning-based forecasting models. To be more specific, multiple sets of deterministic forecasts are generated by using different machine learning algorithms with different kernels in the first layer. Then, the forecasts are blended by another machine learning algorithm in the second layer to generate final forecasts. Machine learning algorithms used in the M3 method include artificial neural networks (ANNs), support vector regression (SVR), gradient boosting machines (GBMs), and random forests.

## 6.2.2 Weather Distribution Modeling

In the literature, marginal distributions of weather variables are commonly modeled by unimodal distributions such as Beta and Gamma (Louie, 2010) or non-parametric distributions such as kernel density estimation (KDE) (Wang et al., 2018). However, unimodal distributions may not accurately quantify the variability of weather variables and non-parametric distributions are challenging to be solved analytically (Cui et al., 2018). Mixture distributions have been widely utilized in statistics to approximate multi-modal distributions. To accurately characterize the variability of weather variables, GMM is adopted in this study to model the weather variables. Three weather variables are considered in this study, which are global horizontal irradiance (GHI), wind speed (WS), and temperature (TEMP). The PDF of GMM is formulated as follows:

$$f_G(x|N_G, \omega_i, \mu_i, \sigma_i) = \sum_{i=1}^{N_G} \omega_i g_i(x|\mu_i, \sigma_i) \quad (6.1)$$

where  $N_G$  is the number of mixture components,  $U(\mu_i \in U)$  is an expected value vector,  $\Sigma(\sigma_i \in \Sigma)$  is a standard deviation vector, and  $\Omega(\omega_i \in \Omega)$  is a weight vector. Each component  $g(x; \mu_i, \sigma_i)$  follows a normal distribution, which is expressed as:

$$g_i(x|\mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad (6.2)$$

The GMM distribution has two constraints: (1) the integral of Eq. 6.1 equals unity, and (2) the summation of weight parameters equals unity as well, which are expressed as follows:

$$\int_{-\infty}^{+\infty} f_G(x|N_G, \omega_i, \mu_i, \sigma_i) dx = \int_{-\infty}^{+\infty} \sum_{i=1}^{N_G} \omega_i g_i(x|\mu_i, \sigma_i) dx = 1 \quad (6.3)$$

$$\sum_{i=1}^{N_G} \omega_i = 1 \quad (6.4)$$

The parameters of GMM are estimated by the expectation maximization (EM) algorithm. The goal of EM is to maximize the likelihood function with respect to parameters. More details about EM

can be found in (Hartley, 1958). The cumulative density function (CDF) ( $F_G$ ) corresponding to the estimated PDF is expressed as:

$$F_G(x|N_G, \omega_i, \mu_i, \sigma_i) = \sum_{i=1}^{N_G} \left[ \frac{\sqrt{\pi}}{2} \omega_i \sigma_i \operatorname{erf}\left(\frac{\mu_i - x}{\sigma_i}\right) \right] + C \quad (6.5)$$

where  $C$  is an integral constant, and  $\operatorname{erf}(\cdot)$  is a Gaussian error function.

### 6.2.3 Weather Scenario Generation

Once marginal distributions of historical actual and predicted weather variables are defined, the correlation among weather variables can be modeled through a multivariate joint distribution. The parameter  $x_{f,j}$  denotes the forecast value of the  $j$ th weather variable, and  $x_{a,j}$  denotes the actual value of the  $j$ th weather variable. We assume there are  $J$  weather variables. The parameter  $\vec{f}$  denotes the  $J$ -dimension vector of weather forecasts, i.e.,  $(x_{f,1}, \dots, x_{f,J})$ ;  $\vec{a}$  denotes the  $J$ -dimension vector of actual weather, i.e.,  $(x_{a,1}, \dots, x_{a,J})$ . The CDF and PDF are expressed as:

$$F(a, f) = P(X_{a,1} \leq x_{a,1}, \dots, X_{a,J} \leq x_{a,J}, X_{f,1} \leq x_{f,1}, \dots, X_{f,J} \leq x_{f,J}) \quad (6.6)$$

$$f(a, f) = \frac{\partial^{2J} F(a, f)}{\partial X_{a,1} \dots \partial X_{a,J} \partial X_{f,1} \dots \partial X_{f,J}} \quad (6.7)$$

Therefore, the weather scenario generation given predicted weather becomes sampling a multivariate distribution of the actual weather  $a$ , i.e.,  $(x_{a,1}, \dots, x_{a,J})$ , conditioning on the predicted weather  $f$ , i.e.,  $(x_{f,1}, \dots, x_{f,J})$ .

$$F(a|f) = P(X_{a,1} \leq x_{a,1}, \dots, X_{a,J} \leq x_{a,J} | X_{f,1} = x_{f,1}, \dots, X_{f,J} = x_{f,J}) \quad (6.8)$$

$$f(a|f) = \frac{\partial^J F(a|f)}{\partial X_{a,1} \dots \partial X_{a,J}} \quad (6.9)$$

In this study, Copula is used to model the inherent correlation among weather variables, and the conditional multivariate distribution is modeled based on the aforementioned weather marginal distributions. Given a multivariate joint distribution, it is challenging to directly generate samples

due to the high dimensionality of conditioning variables. In this study, Gibbs sampling is applied to sample the multidimensional random variables by sequentially sampling each component.

Gibbs sampling seeks to iteratively sample only one variable or a block of variables at a time from its distribution conditioned on the remaining variables (Tang et al., 2018). Therefore, Gibbs sampling converts sampling a multivariate distribution into sampling a set of conditional univariate distributions. The pseudocode of Gibbs sampling is illustrated in Algorithm 3. Note in this study, we use the forecasted weather as the initial input.

---

**Algorithm 3:** Gibbs sampling

---

**Result:** Weather scenarios

```

1 initialization:  $x_{a,j}^0 \leftarrow x_{f,1}, \forall j = 1, \dots, J;$ 
2 for scenario  $\eta = 1, \dots,$  do
3    $X_{a,1} | x_{a,2}^{(\eta-1)}, \dots, x_{a,J}^{(\eta-1)};$ 
4    $X_{a,2} | x_{a,1}^{(\eta)}, x_{a,3}^{(\eta-1)}, \dots, x_{a,J}^{(\eta-1)}, x_{f,1}, \dots, x_{f,J};$ 
5   ...
6    $X_{a,J} | x_{a,1}^{(\eta)}, \dots, x_{a,J-1}^{(\eta)}, x_{f,1}, \dots, x_{f,J}$ 
7 end

```

---

Copula is one of the most widely used methods for modeling the dependency among random variables. Based on Sklar's theory (Rüschendorf, 2009), any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula that describes the dependence structure between the variables. Therefore, the joint CDF in Eq. 6.6 can be written as:

$$F(a, f) = C(F(x_{a,1}), \dots, F(x_{a,J}), F(x_{f,1}), \dots, F(x_{f,J})) = C(S, T) \quad (6.10)$$

where,

$$S = F(x_{a,1}), \dots, F(x_{a,J}) \quad (6.11)$$

$$T = F(x_{f,1}), \dots, F(x_{f,J}) \quad (6.12)$$

$F(x_{a,j})$  and  $F(x_{f,j})$  denote the marginal CDFs of actual weather and marginal CDF of predicted weather, respectively.  $C(\cdot)$  is the Copula function.

Similarly, the joint PDF of actual weather and predicted weather can be expressed as:

$$f(a, f) = c(S, T) \cdot \prod_{j=1}^J (f(x_{a,j})f(x_{f,j})) \quad (6.13)$$

where,  $f(x_{a,j})$  and  $f(x_{f,j})$  are marginal PDFs of  $X_{a,j}$  and  $X_{f,j}$ , respectively, modeled by using the aforementioned GMM distribution based on historical actual and forecasting weather.  $c(\cdot)$  is the density of Copula. Then, the conditional univariate distribution of weather variable  $X_{a,j}$  can be deduced from the Bayesian formula, given by:

$$\begin{aligned} F(x_{a,j}|x_{a,1}, \dots, x_{a,j-1}, x_{a,j+1}, \dots, x_{a,J}, f) \\ = \frac{P(X_{a,j} \leq x_{a,j}, X_{a,p} = x_{a,p}, X_{f,q} = x_{f,q})}{P(X_{a,p} = x_{a,p}, X_{f,q} = x_{f,q})} \quad (6.14) \\ p = 1, \dots, j-1, j+1, \dots, J, \quad q = 1, \dots, J \end{aligned}$$

where the numerator can be written as:

$$\begin{aligned} P(X_{a,j} \leq x_{a,j}, X_{a,p} = x_{a,p}, X_{f,q} = x_{f,q}) &= \int_0^{x_{a,j}} f(a, f) dX_{a,j} \\ &= \int_0^{x_{a,j}} c(\cdot) \prod_{p=1}^J f(x_{a,p})f(x_{f,p}) dX_{a,j} \quad (6.15) \end{aligned}$$

and based on Eq. 6.13, the denominator can be written as:

$$P(X_{a,p} = x_{a,p}, X_{f,q} = x_{f,q}) = c'(\cdot) \frac{\prod_{p=1}^J (f(x_{a,p})f(x_{f,p}))}{f(x_{a,j})} \quad (6.16)$$

Note that in Eq. 6.16, we use  $c'(\cdot)$  instead of  $c(\cdot)$  since the dimension of Copula density here is  $2J - 1$  instead of  $2J$ . Overall, based on Eq. 6.15 and Eq. 6.16, the conditional univariate CDF can be expressed as:

$$\begin{aligned} F(x_{a,j}|x_{a,1}, \dots, x_{a,j-1}, x_{a,j+1}, \dots, x_{a,J}, f) \\ = \frac{\partial^{2J-1} C(\cdot)}{\partial X_{a,1} \dots \partial X_{a,j-1} \partial X_{a,j+1} \dots \partial X_{a,J} \partial X_{f,1} \dots \partial X_{f,J}} \cdot \frac{1}{c'(\cdot)} \quad (6.17) \end{aligned}$$

The conditional distribution of the actual weather variable given weather forecasts can be trained by using historical actual and forecasting weather data. With any given deterministic weather forecasts, the Gibbs sampling model and the trained conditional CDF in Eq. 6.17 are used together to generate a large number of weather scenarios. In this study, the Copula is selected through the “xvCopula” function from the **Copula** package (Yan et al., 2007) in R based on k-fold cross-validation.

#### **6.2.4 Probabilistic Solar Power Forecasting**

Once weather scenarios are given, a large number of solar power scenarios can be generated through a deterministic forecasting model. Assuming that we generate  $N$  weather scenarios for each hour, then we will have  $N$  deterministic solar power forecast values for each hour. Therefore, based on the empirical distribution function, the 1<sup>st</sup> to 99<sup>th</sup> percentiles of solar power can be calculated for each hour, thus generating probabilistic solar power forecasts. Though any deterministic forecasting model is capable of generating solar power forecasts, we adopted the M3 model mentioned in Section 6.2.1.

### **6.3 Case Studies and Results**

The developed wsp-SPF framework is evaluated at 7 solar farms in Texas that are selected from the ACTIVSg2000 system, i.e., a 2000-bus synthetic grid on the footprint of Texas (Birchfield et al., 2016). The data information at the 7 selected solar sites is briefly summarized in Table 6.1. To ensure the generality and diversity of data, some of the selected solar farms are closed to each other (e.g., C2 and C3; C5, C6, and C7), and some of the solar farms are geographically dispersed (e.e., C1 and C4). In addition, all the selected solar farms have different capacity which ranges from 1.05 MW to 230 MW. To match the solar power with corresponding weather information, the weather data is collected from the National Solar Radiation Database (NSRDB) (Sengupta et al., 2018). The NSRDB includes solar radiation and other meteorological information (e.g., wind

speed, air temperature, solar zenith angle) over the United States from 1998 to 2017 computed by the National Renewable Energy Laboratory’s (NREL’s) Physical Solar Model (PSM), with a 30 min temporal resolution. The solar power is generated based on NSRDB weather data using System Advisor Model (SAM) (Freeman et al., 2019). In this study, the duration of the collected data at the 7 selected solar sites spans four years from January 1<sup>st</sup> 2008 to December 31<sup>st</sup> 2011. Fig. 6.2 summaries the data partition of the case study. For all the 7 locations, the first 3/4 of the data is used as training data, in which the first 2/3 is used to train the deterministic weather forecasting model and the remaining 1/3 of the training data is used to train the conditional weather scenario model. The accuracy of the forecasts is evaluated by the remaining 1/4 of data. The number of weather scenarios generated from the conditional distribution is set as  $N_s=5,000$ . Though the developed wsp-SPF method is capable of generating forecasts at multiple forecasting horizons, only 1-h-ahead solar power forecasts are explored in this study.

Table 6.1: Data summary of the selected 7 solar sites

Site Name	Site ID	Lat.	Long.	Capacity (MW)	State
C1	3	29.58	-104.29	10	TX
C2	21	32.26	-101.41	1.5	TX
C3	22	32.25	-101.42	230	TX
C4	109	29.32	-100.38	29.7	TX
C5	286	30.55	-97.69	1.05	TX
C6	287	30.55	-97.69	22.5	TX
C7	288	30.54	-97.69	5.5	TX

Since the Copula-based weather scenario generation method is a two-step method. It is important to compare the performance of different deterministic forecasting models in the first step. Normalized indices of standard metrics like root mean squared error and mean absolute error, *i.e.*, NRMSE and NMAE, are adopted to evaluate the performance of deterministic forecasts. The 1HA deterministic forecasting errors by using the M3 deterministic forecasting model at the selected locations are summarized in Table 6.2. The persistence of cloudiness method (PS) (Cui et al., 2017) has been proved to be accurate in the shorter forecasting period. Therefore, to show the superiority of the M3 deterministic forecasting model, the PS method is adopted as the baseline. Overall,

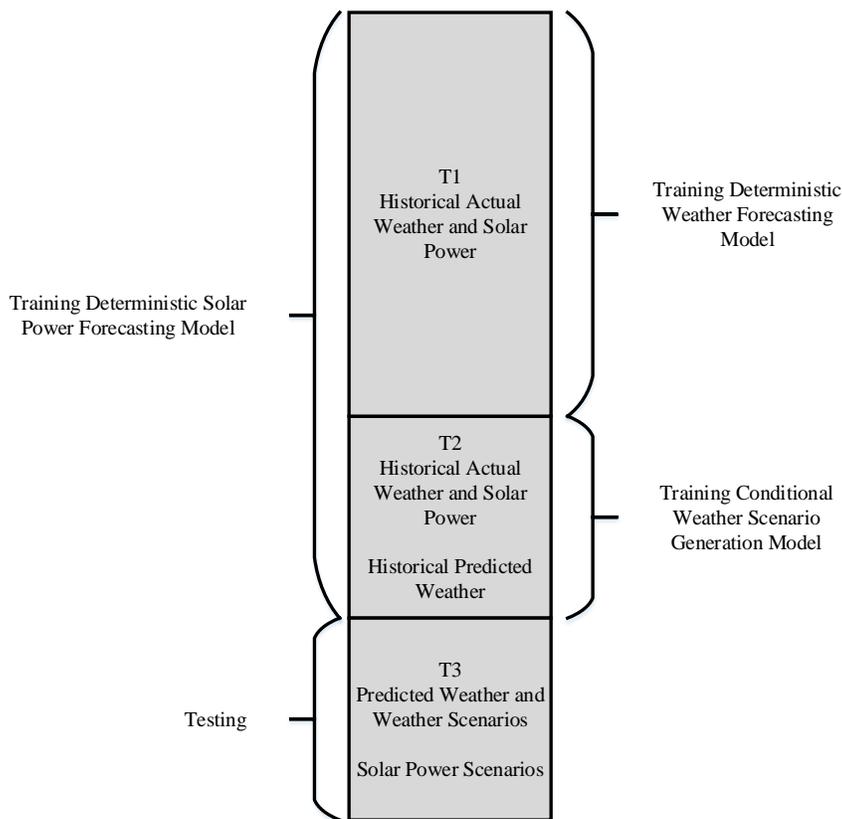


Figure 6.2: Data splitting for model training and testing

the accuracies of the M3 deterministic forecasts are better than those of persistence of cloudiness forecasts.

In this study, four benchmark models are selected for comparison, including one single probabilistic forecasting baseline model, and three other weather scenario generation-based models. The single benchmark model is quantile regression (QR). The three benchmark weather scenario generation models are fixed-date (FD), shifted-date (SD), and bootstrap (BS) methods, which have been used by Xie et al. in (Xie and Hong, 2016). Note that the same M3 deterministic forecasting model is used in the three weather scenario generation-based models.

1. FD: The fixed-date method assigns the weather profile of historical years to the current year.

The number of scenarios equals to the number of years the weather profiles is available.

Table 6.2: 1HA deterministic weather forecasting results by using M3 and PS

Model	Feature	Metric	Site						
			C1	C2	C3	C4	C5	C6	C7
M3	GHI	NMAE(%)	4.32	3.98	4.11	4.23	3.53	3.72	3.48
		NRMSE(%)	7.16	6.59	6.81	6.99	5.91	6.21	5.94
	WS	NMAE(%)	1.82	1.96	1.89	1.83	1.77	1.77	1.81
		NRMSE(%)	2.74	2.92	2.77	2.60	2.47	2.49	2.53
	TEMP	NMAE(%)	1.91	1.68	1.69	1.69	1.35	1.36	1.36
		NRMSE(%)	2.46	2.19	2.19	2.15	1.79	1.79	1.79
PS	GHI	NMAE(%)	6.79	6.63	6.63	6.62	6.59	6.60	6.57
		NRMSE(%)	10.84	10.79	10.79	10.95	11.00	11.01	10.94
	WS	NMAE(%)	3.10	3.61	3.61	3.34	3.31	3.31	3.31
		NRMSE(%)	4.60	5.27	5.27	4.85	4.91	4.91	4.91
	TEMP	NMAE(%)	2.57	2.23	2.23	2.31	2.12	2.12	2.11
		NRMSE(%)	3.40	3.07	3.07	3.13	2.94	2.94	2.94

Assume that  $m$  years weather profiles are available, we could generate  $m$  weather scenarios at each time step.

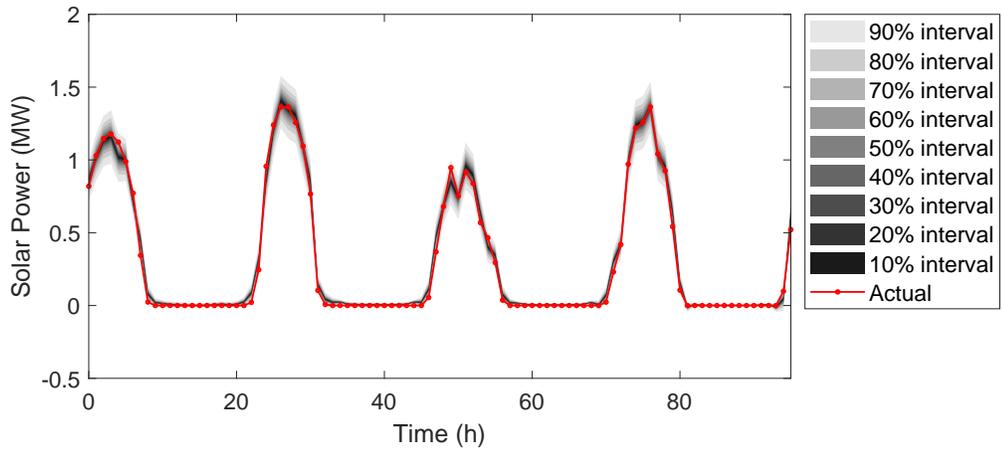
2. SD: The shifted-date method generally shifts the historical weather profile with a number of days. Then these shifted weather profiles are treated as weather scenarios of the current year. Assume that  $m$  years weather profiles are available, we could generate  $(2n + 1)m$  weather scenarios at each time step, where  $n$  is the number of days shifted forward or backward. The  $n$  is set to be 4 in this study.
3. BS: The bootstrap method divides the weather profile of each historical year into an equal length of blocks, and then randomly picks the blocks with replacement from any of the historical years to form a new temperature profile. In this study, we set the block length be 10. Overall, there are 37 blocks, where each of the first 36 blocks has a length of 10 and the 37th block has a length of 5.

The reasons for choosing these four baseline models are: (i) QR is a widely used method in probabilistic forecasting, which allows us to explore the forecasting enhancement by considering

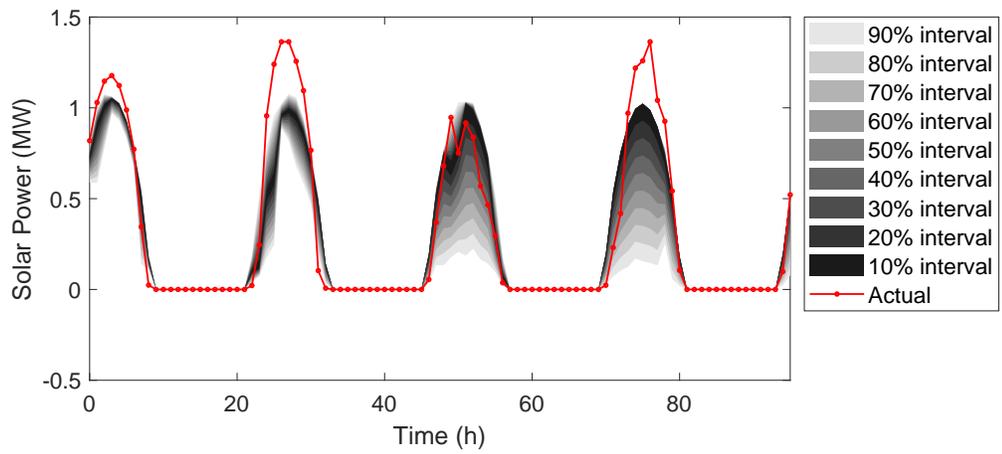
weather scenario generation; (ii) since a weather scenario generation model is included in the proposed wsp-SPF method, it is important to compare the accuracy of the proposed method with different weather scenario generation techniques. Note that the empirical probability distribution is adopted to calculate the quantile forecasts. However, for the FD and SD methods, the number of both weather and solar power scenarios is limited. As a result, some of the adjacent quantiles may share a same value.

Fig. 6.3(a) shows the 1HA probabilistic solar power forecasts of the C2 site from 2011-01-31 to 2011-02-03, generated from the proposed wsp-SPF model. It is observed that at the entire representative period, the solar power reasonably lies within the PIs. Figs. 6.3(b), 6.3(c), 6.4(d), and 6.4(e) show the probabilistic forecasts generated from the baseline M3-FD, M3-SD, M3-BS, and QR methods, respectively. It is seen that the PIs of the wsp-SPF model in Fig. 6.3(a) is narrower than the PIs with other weather scenario generation methods. This is due to that considering the correlation among weather variables has improved the weather scenario generation accuracy. In addition, the PIs of wsp-SPF are smoother than those of QR, which indicates stable and reliable probabilistic forecasts. Among the three weather scenario generation-based baseline methods, M3-BS outperforms M3-FD and M3-SD. It is mainly due to the limited number of scenarios in FD and SD models. It is also observed that the width of the PIs varies with the variability of solar power. For example, when the solar power fluctuates more frequently, the PI tends to be wider, and thereby the uncertainty in solar power forecasts is relatively higher.

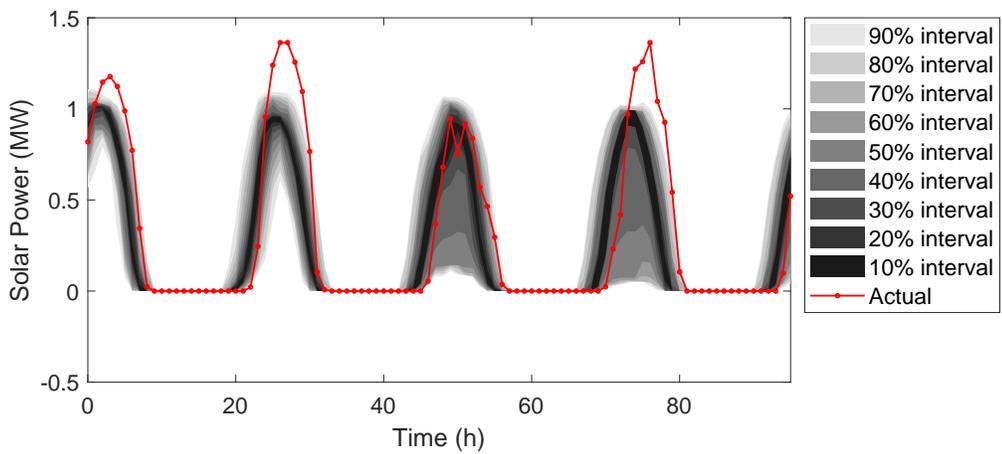
To show the effectiveness of the developed weather scenario generation based probabilistic forecasting framework, the normalized pinball loss (NPL) values of 1HA solar power forecasts from different models and their relative improvement (IP) with respect to wsp-SPF model are compared in Table 6.3. The sum of pinball loss is averaged over all quantiles from 1% to 99% and normalized by the solar power capacity. A lower pinball loss score indicates a better probabilistic forecast. Results show that the proposed wsp-SPF model gives the best performance. Moreover, the proposed wsp-SPF model has improved the pinball loss from 5.08% to 140.44% compared to



(a) wsp-SPF

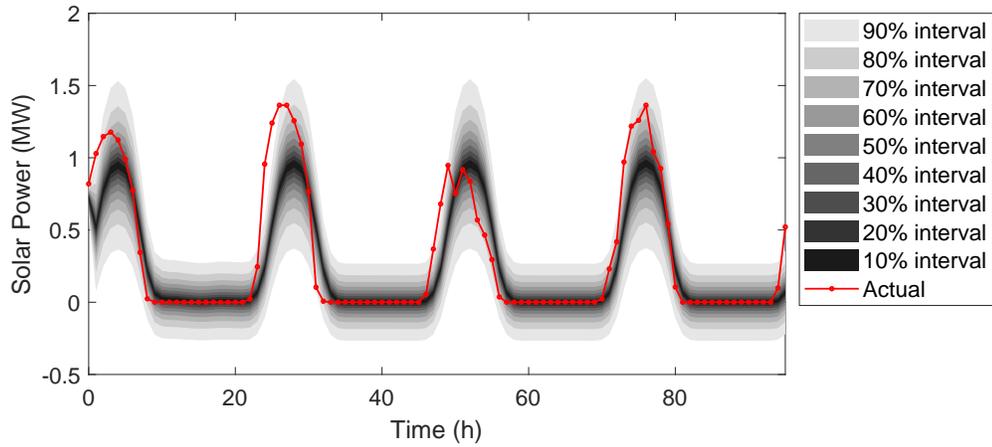


(b) M3-FD

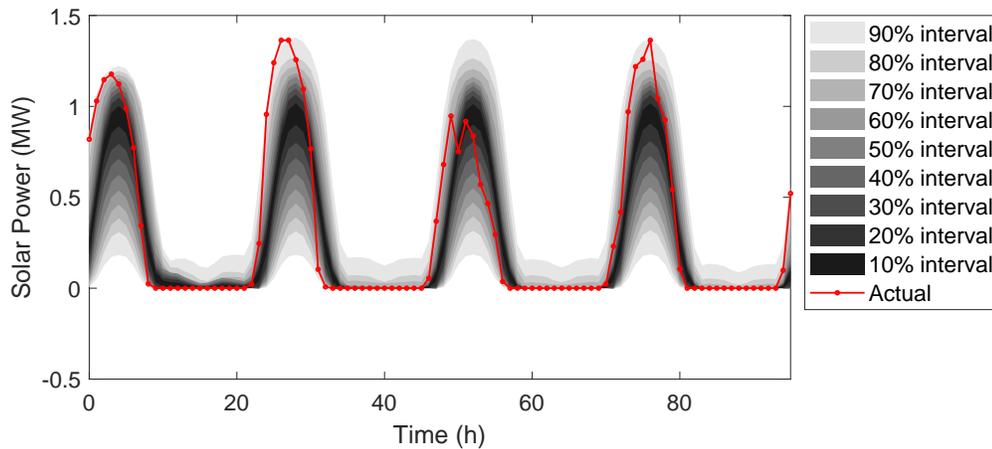


(c) M3-SD

Figure 6.3: 1HA probabilistic solar power forecasts



(d) M3-BS



(e) QR

Figure 6.3: (continued) 1HA probabilistic solar power forecasts

the four benchmark models, which validates the effectiveness of the proposed method. It is interesting to see the highest NPL improvement is from C5, which has the smallest solar capacity. This may be due to the solar farms with smaller capacity are more susceptible to weather conditions, thus the correlation between weather scenarios and PV power outputs becomes stronger. In contrast, for solar farm owns the biggest capacity like C3, the NPL improvement is intermediate. Therefore, there is no obvious linear relationship between NPL improvement brought from weather scenario generation and solar farm capacity. The improvement may be a joint result from different factors such as geographic location and capacity. Note that the wsp-SPF method has shown a better accu-

racy than M3-FD, M3-SD, and M3-BS, which indicates the improvement of the weather scenario generation by considering the correlation among weather variables. Furthermore, the wsp-SPF model outperforms QR, which shows the improvement by considering weather scenario generation modeling. In addition to pinball loss, two more standard metrics, i.e., reliability and sharpness, are also calculated to assess the performance of wsp-SPF.

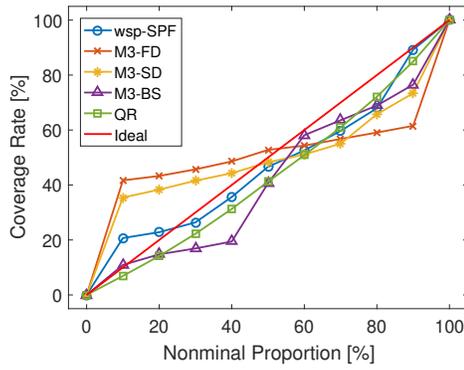
Table 6.3: Normalized pinball loss and relative improvement of different models

Model		Site						
		C1	C2	C3	C4	C5	C6	C7
NPL	wsp-SPF	<b>1.77</b>	<b>1.62</b>	<b>1.75</b>	<b>2.13</b>	<b>1.35</b>	<b>1.49</b>	<b>1.46</b>
	M3-FD	1.86	2.07	2.08	2.43	2.38	2.38	2.34
	M3-SD	3.08	3.11	3.11	2.99	3.12	3.11	3.11
	M3-BS	3.09	3.43	3.74	2.94	3.74	3.27	3.29
	QR	2.46	2.64	2.64	2.44	1.40	2.66	2.64
IP (%)	M3-FD	5.08	27.78	18.86	14.08	75.00	59.73	60.27
	M3-SD	74.01	91.97	77.71	40.37	128.68	108.72	113.01
	M3-BS	74.58	<i>111.72</i>	<i>113.71</i>	38.02	<i>140.44</i>	<i>119.46</i>	<i>125.34</i>
	QR	38.98	62.96	50.86	14.55	95.59	78.52	80.82

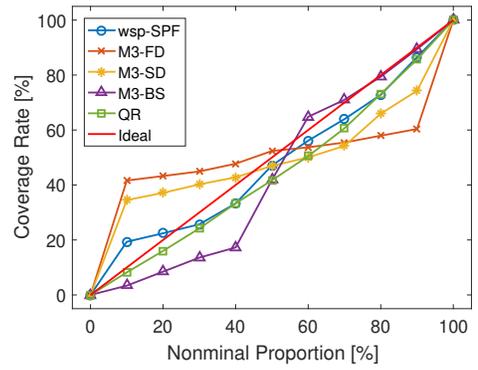
*Note:* The smallest normalized pinball loss value is in boldface. The highest relative improvement with respect to wsp-SPF model is in italic.

Fig. 6.4 shows the reliability plots of the probabilistic solar power forecasts with different forecasting models at the 7 sites. A forecast presents better reliability when the curve is closer to the diagonal. It is seen from Fig. 6.4 that overall the proposed wsp-SPF has better reliability performance than M3-FD, M3-SD, and M3-BS, indicating the enhancement resulted from correlation modeling between weather variables. In addition, the proposed wsp-SPF model has shown similar reliability to QR, while the PIs of wsp-SPF are narrower than those of QR, which secures accuracy without sacrificing reliability. In addition, note that the reliability at C5 is worse than those from other solar farms, and C3 has the best reliability. It is mainly because C5 has the smallest capacity, which is susceptible to weather variation. In contrast, C3 has the largest capacity, which is supposed to be more reliable and stable.

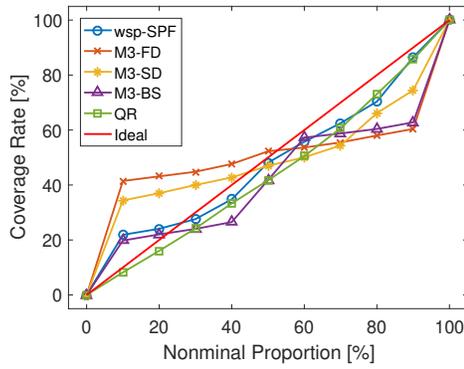
The sharpness plots of wsp-SPF and four baseline models at different sites are compared in Fig. 6.5. The expected interval size increases with increasing the nominal coverage rate, and



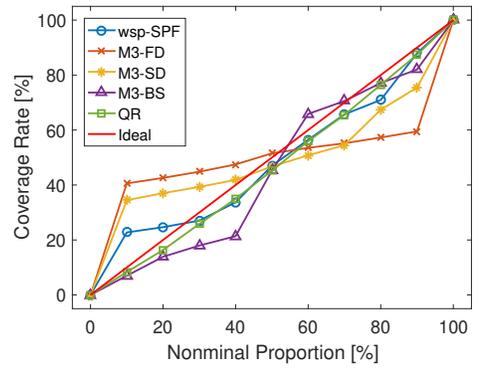
(a) C1



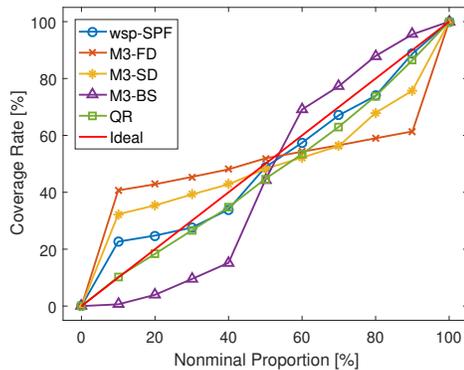
(b) C2



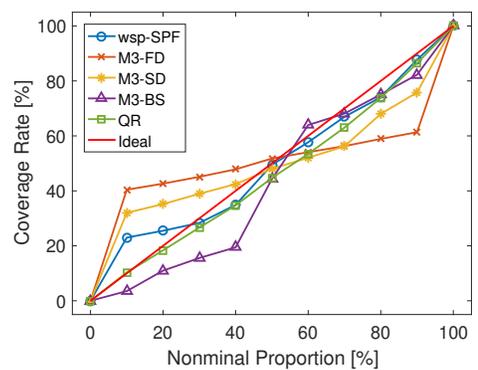
(c) C3



(d) C4

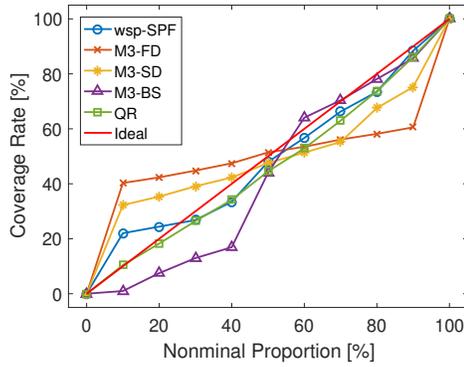


(e) C5



(f) C6

Figure 6.4: Reliability comparison of different models at different sites



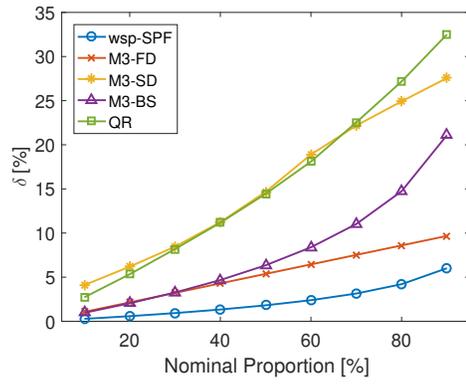
(g) C7

Figure 6.4: (continued) Reliability comparison of different models at different sites

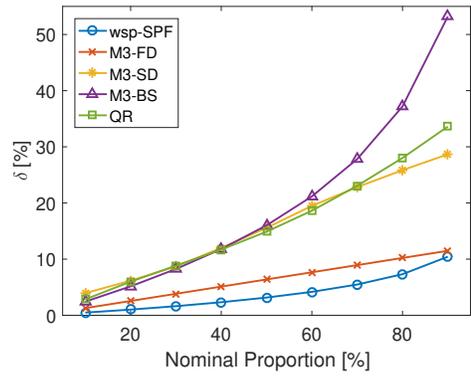
the sharpness of the proposed wsp-SPF model is significantly better than that of the four baseline models. Overall, the interval size of the proposed cp-AWPF model ranges from 1% to 10%, which indicates low sharpness. In addition, wsp-SPF has significantly better sharpness than M3-FD, M3-SD, and M3-BS at all sites, which validates the enhancement of correlation modeling in weather scenario generation. Since all the sites share low sharpness, it is hard to conclude the relationship between site capacity and sharpness. Moreover, it is seen that for solar farms in the same region (similar longitude and latitude), the sharpness is also similar. This characteristic may be used for solar resource assessment.

## 6.4 Conclusion

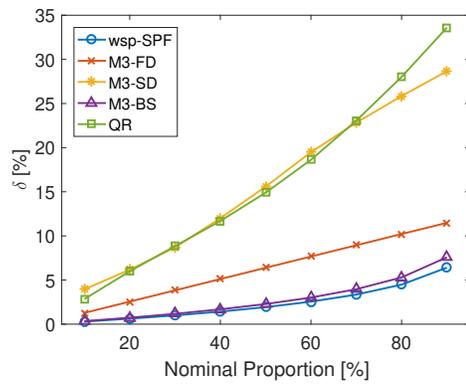
In this chapter, a weather scenario generation-based probabilistic solar power forecasting framework was developed. Gaussian mixture model was used to accurately model the weather marginal distribution. Copula was adopted to model the correlation among different weather variables through a high-dimensional joint distribution. Gibbs sampling was applied on the conditional CDF of the joint distribution to generate a large number of weather scenarios. Then, these weather scenarios are fed into a machine learning-based multi-model deterministic forecasting model to generate probabilistic solar power forecasts. Results at 7 selected solar farms showed that:



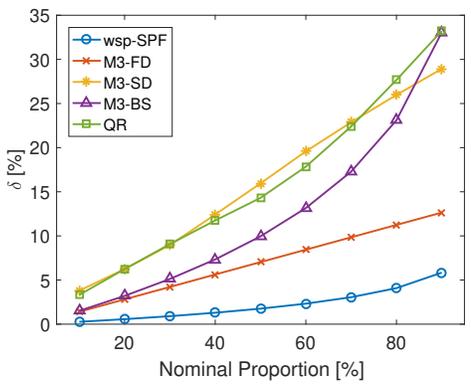
(a) C1



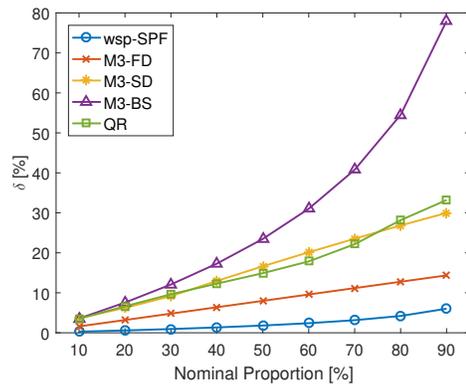
(b) C2



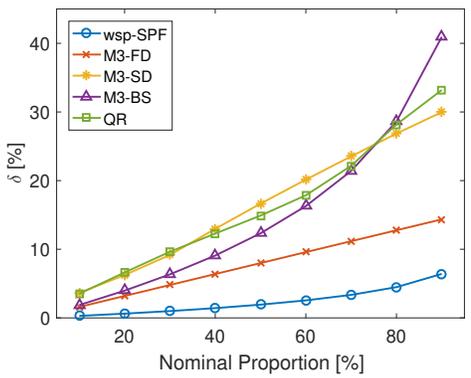
(c) C3



(d) C4

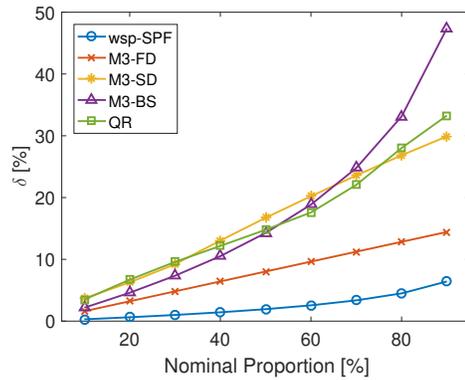


(e) C5



(f) C6

Figure 6.5: Sharpness comparison of different models at different sites



(g) C7

Figure 6.5: (continued) Sharpness comparison of different models at different sites

1. wsp-SPF could reduce the pinball loss score by up to 140% compared to four benchmark models.
2. The GMM model has shown better goodness-of-fit to weather distribution than single distribution models and KDE.
3. Considering correlation among weather variables could enhance weather scenario generation, thus providing better probabilistic forecasting accuracy.
4. The developed wsp-SPF framework is robust for solar farms at different locations with different capacities.
5. wsp-SPF has shown better sharpness than models without using weather scenario generation and models with other weather scenario generation methods (i.e., FD, SD, and BS). The reliability of wsp-SPF is close to the ideal diagonal, which indicates reasonable reliability.

## **CHAPTER 7**

### **UTILIZING PROBABILISTIC FORECASTING FOR ANOMALY DETECTION IN RENEWABLE ENERGY**

With the fast-growing trend of solar energy, accurate solar power forecasting plays an important role in grid integration of solar power and power system operations. Recently cyberattacks on solar power forecasting emerge with the increase of solar penetration, which may lead to substantial economic losses and power system reliability issues.

As suggested in chapter 2.1, probabilistic forecasting could be used for anomaly detection thus better assisting energy system cyber security. This chapter presents a novel probabilistic anomaly detection framework to effectively and accurately detect cyberattacks in solar power forecasting, which consists of three major components. First, a convolutional neural network is used to extract spatial correlations among solar farms at the bottom of the model. Second, a long short-term memory network is used to capture the temporal dependencies within solar power data and generate deterministic solar power forecasts, which is then converted to probabilistic forecasts through pinball loss optimization. Finally, the probabilistic solar power forecasts are used for anomaly detection. The effectiveness of the proposed framework is validated by using solar farms from the Texas 2000-bus synthetic grid system with a variety of data integrity attacks. Numerical results of case studies at the sites show that the developed probabilistic anomaly detection methodology could effectively detect data integrity attacks in solar power forecasting with a relatively high accuracy.

#### **7.1 Motivation and Objective**

Machine learning and deep learning based deterministic forecasting models have been widely adopted in the literature to assist the aforementioned three types of anomaly detection methods due to their strong abilities of modeling the non-linearity within data (Xin et al., 2018). Trained

forecasting models are tested on input data and to compare the predicted values with the measurements. A model trained through health data will be able to predict the expected behavior based on history observations. However, there will be a mismatched behavior between the prediction and observation if the data is under cyberattacks. For example, Wang et al. (Wang et al., 2019) presented a power consumption anomaly detection method based on long short-term memory (LSTM) point forecasts and error patterns. In (Wang and Ahn, 2020), a real-time anomaly detection method was proposed for electric load based on an integration of support vector machine (SVM), k-nearest neighbors (KNN), and a cross-entropy loss function. However, most of the existing anomaly detection methods are deterministic and thus insufficient to characterize the uncertainties in cyberattacks. Probabilistic approaches that provide quantitative uncertainty information associated with cyberattacks are therefore expected to better assist power system operations and renewable energy integration. Probabilistic forecasting usually takes the form of prediction intervals (PIs) or scenarios (Sun et al., 2020a). The PIs of different confidence levels are generally used to assist decision making, which could naturally be used as thresholds to label anomaly events. If a prediction falls outside a desired PI, such prediction may be labeled as anomaly. Probabilistic forecasting methods can be classified into parametric methods and non-parametric methods (Sun et al., 2020b). However, none of these two types of probabilistic forecasting methods has been applied to data integrity anomaly detection in solar power forecasting.

Overall, several challenges present in existing methods for energy anomaly detection: (i) most of the existing anomaly detection methods focus on load forecasting and electricity price forecasting, and little work has been done on solar power forecasting and wind power forecasting; (ii) most of the anomaly detection models are deterministic, which could not well describe the uncertainties in solar power; (iii) the spatial-temporal correlation between different targets (e.g., solar farms, wind farms, or load centers) is not considered, which may not be reliable towards sophisticated designed cyberattacks; (iv) most anomaly detection methods could not detect multiple solar farms simultaneously.

## 7.2 Methodology

In this chapter, we plan to utilize deep learning-based probabilistic forecasting methods for anomaly detection for solar farms under cyberattacks. The cyberattacks on solar power forecasting may lead to substantial economic losses and power system reliability issues. Therefore, an accurate anomaly detection method is need to assist solar energy integration and system reliability.

The anomaly detection is formulated as a binary classification problem through probabilistic solar power forecasts. The details of the two cornerstone components in deterministic forecasting, i.e., CNN and LSTM, are first introduced. Then, a pinball loss optimization-based probabilistic solar power forecasting method is described. At last, the overall anomaly detection framework for solar power forecasting under data integrity attack is presented.

In this chapter, we seek to identify the anomaly behavior in solar power forecasting. The overview of data integrity attacks is depicted in Fig. 7.1. Different from other related works,

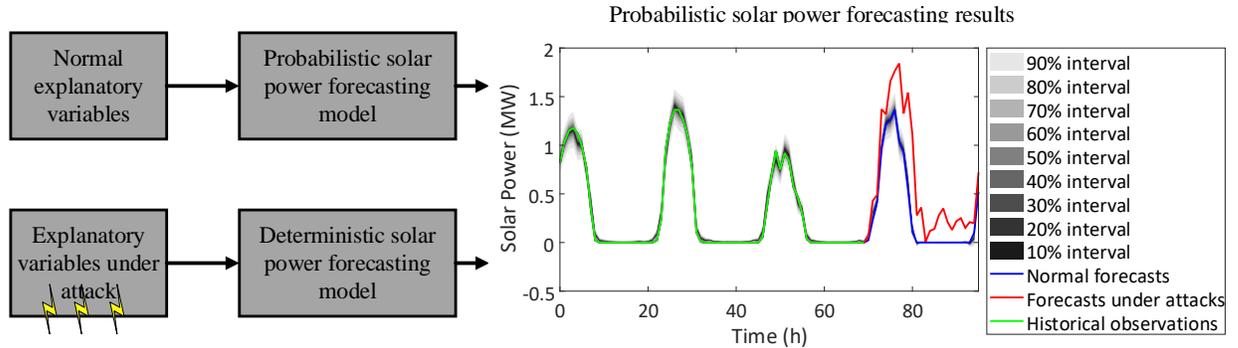


Figure 7.1: Overview of data integrity attacks in solar power forecasting

our proposed method focuses on utilizing probabilistic forecasting to detect anomaly behaviors. Moreover, the spatial-temporal correlation among different solar farms is considered, which not only enhances the probabilistic solar power forecasting accuracy, but also allows us to cross-check the possible malicious correlation changes within different solar farms.

Studies have shown that geographically dispersed solar farms are correlated to a certain extent (Zhang et al., 2016). In addition, power produced from the same solar farm at different times is

typically temporally correlated (Prasad et al., 2017). It would be interesting to explore the impacts of spatial-temporal correlation on the performance of anomaly detection in solar power forecasting. Instead of detecting the anomaly behavior in one solar farm, in this study, we detect multiple solar farms simultaneously. To model multiple solar farms, we assume the target solar farms over a geographic region are non-uniformly distributed as an array with  $P$  rows and  $Q$  columns as illustrated in Fig. 7.2. The solar power time-series at site  $(1,Q)$  is represented with the red curve.

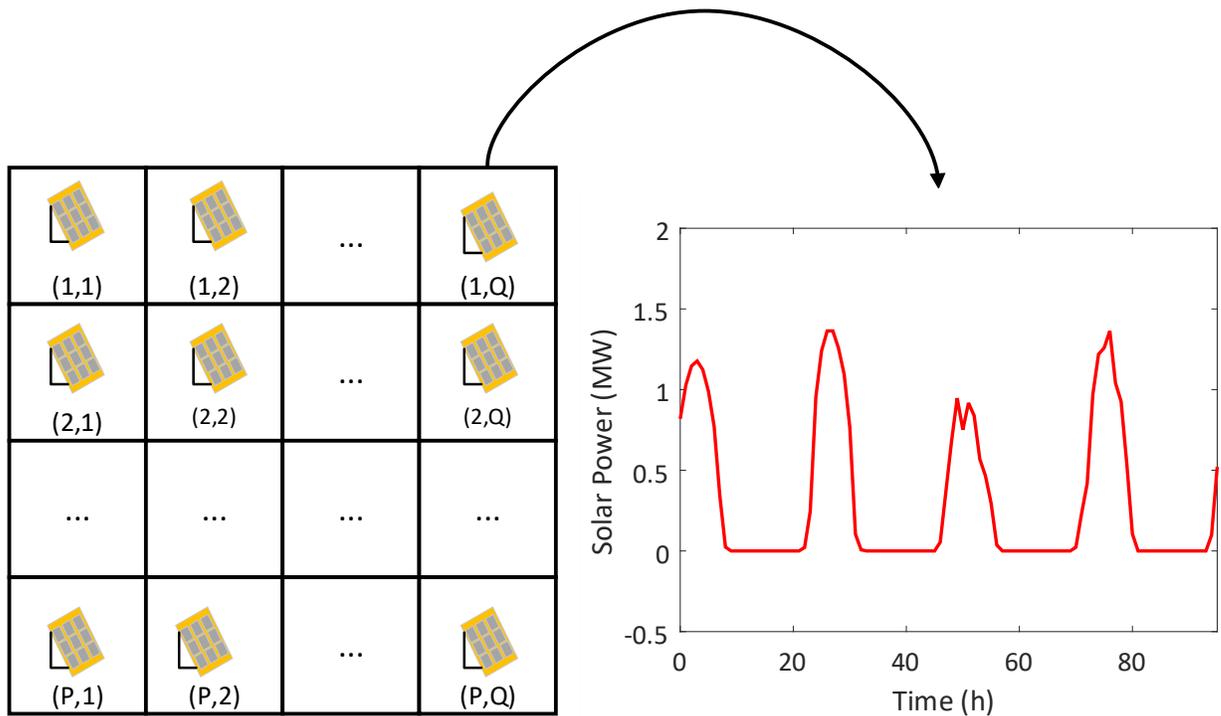


Figure 7.2: Spatial solar power sequence

Therefore, the solar power time-series of multiple solar farms at time  $t$  can be represented by a matrix. Let  $y(p,q)_t$  denotes the solar power at  $(p,q)$ , the solar power spatial matrix (SPSM) of all

the solar farms in the array at time  $t$  could be represented as:

$$y_t = \begin{bmatrix} y(1,1)_t & y(1,2)_t & \dots & y(1,Q)_t \\ y(2,1)_t & y(2,2)_t & \dots & y(2,Q)_t \\ \vdots & \vdots & & \vdots \\ y(P,1)_t & y(P,2)_t & \dots & y(P,Q)_t \end{bmatrix} \quad (7.1)$$

This SPSM contains the spatial correlation of all the solar farms at time  $t$ . To model the temporal correlation, similarly, the spatial-temporal sequence describing the solar power of the array could be constructed by organizing the SPSM in chronological order, as illustrated in Fig. 7.3. To this end, the outputs of the solar power forecasting model become an array instead of a single value.

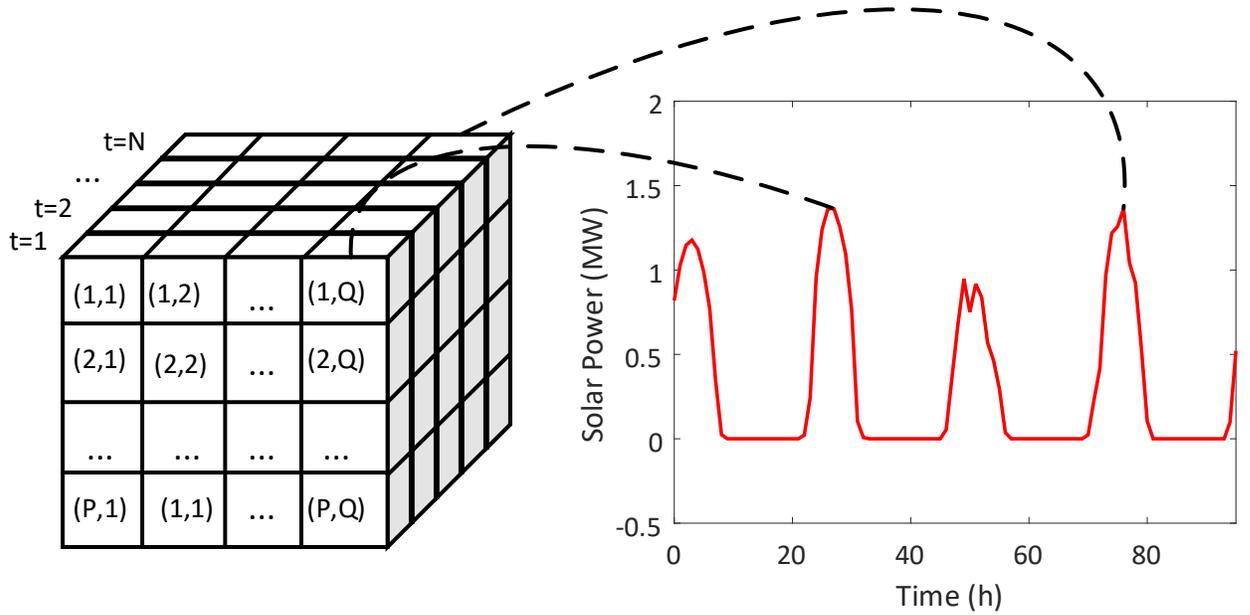


Figure 7.3: Spatial-temporal solar power sequence

### 7.2.1 CNN-based Feature Extraction

CNN is one of the most widely used deep learning models for extracting hierarchical spatially invariant features (Selvin et al., 2017). CNN usually consists of convolutional layers, pooling layers, and fully-connected layers. At the convolutional layers, a convolution operation is applied to the

input feature maps and then these feature maps are transformed through a non-linear activation function. Through the convolution operation, most revealing features of the input can be extracted. The convolution operation process can be expressed as:

$$x_m^o = \Phi\left(\sum_{n \in D_m} x_n^{o-1} * k_{mn}^o + b_m^o\right) \quad (7.2)$$

where  $x_m^o$  denotes the  $m$ th feature map of the  $o$ th layer,  $*$  denotes the convolution operation,  $D_m$  is the inputs pool,  $b_m^o$  denotes the bias,  $k_{mn}$  is the kernel of the  $o$ th layer, and  $\Phi(\cdot)$  is the activation function.

The goal of pooling layer is to progressively reduce the spatial size of the output feature maps from the convolutional layer through a down sampling function, which is expressed as:

$$x_m^o = d(\cdot)(x_m^{o-1}) \quad (7.3)$$

where  $d(\cdot)$  denotes the down sampling function. The feature map is downsampled in such a way that the maximum feature response within a given sample size is retained.

The fully-connected layer connects every neuron in one layer to every neuron in another layer, thus combining all local features into global features to form the output:

$$x^o = \Phi(\omega^o x^{o-1} + b^o) \quad (7.4)$$

where  $\omega^o$  and  $b^o$  are the weight matrix and bias vector of the  $o$ th layer, respectively.

## 7.2.2 LSTM Network

LSTM is a special recurrent neural network (RNN) architecture for time series modeling and forecasting, which has the capability of learning and memorizing long-term dependencies within the time-series data. A standard RNN has one hidden layer, which could only trace back to few time steps due to the vanishing gradient effect (Bouktif et al., 2018). To better capture the long-term

dependencies, LSTM introduces different gates which could regulate the gradient flow of the network. Following the work of (Olah, 2015), the inner structure of the LSTM unit is illustrated in the bottom right block of Fig. 7.4 and described in Eq. 7.5.

$$\begin{aligned}
i_t &= \sigma(x_t W_{ix} + h_{t-1} W_{ih} + c_{t-1} W_{ic} + b_i) \\
f_t &= \sigma(x_t W_{fx} + h_{t-1} W_{fh} + c_{t-1} W_{fc} + b_f) \\
c_t &= c_{t-1} f_t + i_t \cdot \tanh(x_t W_{xc} + h_{t-1} W_{ch} + b_c) \\
o_t &= \sigma(x_t W_{ox} + h_{t-1} W_{oh} + b_o) \\
h_t &= o_t \cdot \tanh(c_t)
\end{aligned} \tag{7.5}$$

where  $i_{(\cdot)}$ ,  $f_{(\cdot)}$ , and  $o_{(\cdot)}$  are the input gate, forget gate, and output gate, respectively.  $\sigma$  denotes the sigmoid activation function,  $h_t$  is the state at  $t$ ,  $x_t$  denotes input,  $o_t$  is the cell output, and  $c_t$  is the memory state. LSTM updates its hidden state  $c_t$  by using the current input  $x_t$  and the previous state  $c_{t-1}$ . The final state  $h_t$  is determined by  $c_t$  and  $o_t$ . The weights are optimized by minimizing the difference between the LSTM outputs and training samples. In this study, the input vector of the multi-input LSTM can be expressed as:

$$x_t = [y_{t-1}, \Phi_t] \tag{7.6}$$

where  $\Phi_t$  denotes the feature vector of the time step  $t$ , and  $y_t$  denotes the observation at time step  $t$ .

### **The Overall Probabilistic Anomaly Detection Framework**

The overall framework of the proposed probabilistic anomaly detection methodology is illustrated in Fig. 7.4. It consists of three major steps:

1. Step 1 (gray blocks): Feed historical data into an CNN-LSTM-based forecasting machine to predict solar power.

2. Step 2 (orange blocks): Convert point forecasts to prediction intervals (PIs) using a parametric probabilistic forecasting method based on designated predictive distribution shapes and pinball loss optimization.
3. Step 3 (green blocks): Detect anomalies based on the threshold confidence and evaluate the performance.

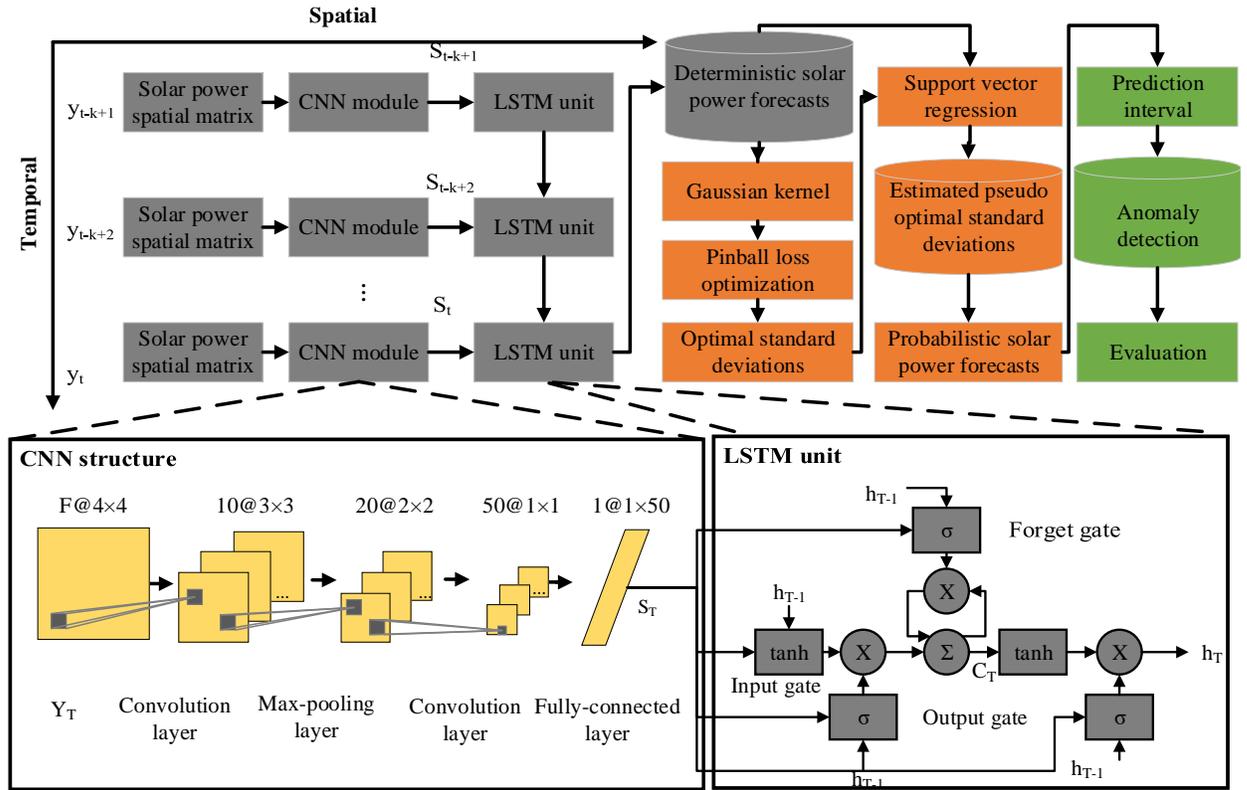


Figure 7.4: Overall framework of CNN-LSTM-based probabilistic anomaly detection

Followed the work by (Zhu et al., 2019), the CNN-LSTM architecture is built with historical SPSMs data in an end-to-end manner. A batch of SPSM data is first convoluted to spatial feature maps through the CNN configuration of CNN-LSTM. In this way, the spatial feature maps, such as distance, solar zenith angle, etc., could be relatively easier capsuled in the parameters of the CNN. As shown in the bottom left box of Fig. 7.4, the CNN configuration has a total of four layers (two convolutional layers, one max-pooling layer, and a fully-connected layer). The filter numbers of the

first and second convolution layer are 10 and 50, respectively. The topology and hyperparameters of the designed CNN architecture are determined by grid search optimization with 10-fold cross validation. The designed CNN is expected to extract high-level abstract spatial features from the SPSMs. The extracted abstract CNN features are then fed into the LSTM configuration. Although LSTM with one hidden layer might be able to capture the temporal patterns in the solar power data, increasing the depth of the architecture by increasing hidden layers can further enhance the performance (Graves et al., 2013). Therefore, two LSTM layers are adopted to operate the hidden state at different timescales and learn temporal correlations that lie within the solar power data. As shown in Fig. 7.4, the first LSTM layer extracts temporal features in the form of hidden state from the previous convolutional feature maps. Then, the second LSTM layer outputs the predicted SPSMs. The numbers of neurons in the first and second LSTM layers are 50 and 100, respectively. Similar to CNN, the hyperparameter optimization of the LSTM are performed by the grid search with 10-fold cross validation on top of the optimal architecture. To avoid overfitting, 10% of neurons are randomly dropped in the LSTM layers.

Though the CNN-LSTM model is composed of two kinds of network architectures, i.e., the CNN and LSTM, it can be jointly trained with one loss function. Note that, the loss function in this task is constructed by the differences between the predicted SPSMs and their corresponding observations, instead of those between the single values in solar power forecasting for a single site.

Once deterministic solar power forecasts are generated, a Gaussian distribution is adopted to model the solar power predictive distribution. The Gaussian distribution is characterized by a mean value ( $\mu$ ) and a standard deviation ( $\sigma$ ). The mean value is approximated by the deterministic point forecast and the standard deviation  $\sigma$  is calculated by minimizing the pinball loss of the quantile function at each time step. The pinball loss value of a certain quantile  $L_m$  is expressed as:

$$L_{m,t}(q_{m,t}, y_t) = \begin{cases} \left(1 - \frac{m}{100}\right) \times (q_{m,t} - y_t), & y_t < q_{m,t} \\ \frac{m}{100} \times (y_t - q_{m,t}), & y_t \geq q_{m,t} \end{cases} \quad (7.7)$$

where  $y_t$  represents the  $t$ th observation,  $m$  represents a quantile percentage from 1 to 99, and  $q_m$  represents the predicted quantile. For a given  $m$  percentage, the quantile  $q_m$  represents the value of a random variable whose cumulative distribution function (CDF) is  $m$  percentage. Pinball loss is one of the most widely used metrics for evaluating probabilistic forecasts (Sun et al., 2019). Smaller pinball loss values indicate better probabilistic forecasting. The process of probabilistic anomaly detection is described as follows.

1. Parameterize the quantile in terms of  $\mu$  and  $\sigma$ , where  $\mu$  is assumed to be the point forecast.

The  $m$ th quantile of the  $t$ th point forecast,  $q_{m,t}$ , is expressed as:

$$q_{m,t} = F^{-1}\left(\frac{m}{100}, \hat{y}_t, \sigma_t\right) \quad (7.8)$$

where  $\hat{y}_t$  and  $y_t$  are deterministic forecasts and observations, respectively.  $F^{-1}(\cdot)$  is the inverse CDF function. The corresponding pinball loss is expressed in Eq. 7.7.

2. Calculate the unknown parameter  $\sigma$  at each time step by minimizing the averaged sum of pinball loss through a genetic algorithm (GA) (Koza and Koza, 1992):

$$\begin{aligned} \sigma_t^* &= \arg \min_{\sigma_t} \frac{1}{N_m} \sum_{m=1}^{N_m} L_{m,t}(\sigma_t, y_t, \hat{y}_t, m) \\ \text{subject to} \quad &\sigma_l \leq \sigma_t \leq \sigma_u \end{aligned} \quad (7.9)$$

where  $\sigma_t^*$  is the optimal standard deviation of the  $t$ th time step;  $N_m = 99$  is the number of quantiles;  $\sigma_l$  and  $\sigma_u$  are the lower and upper bound of  $\sigma$ , which are set as 0.01 and 50, respectively.

3. A support vector regression (SVR) surrogate model (Sun et al., 2020a) is used to fit the point forecast and  $\sigma^*$  in the training stage, which is used to generate unknown pseudo optimal standard deviations,  $\hat{\sigma}^*$ , in the forecasting stage.
4. During probabilistic forecasting, both the deterministic forecasts, i.e.,  $\mu$ , generated by CNN-LSTM and the estimated pseudo optimal standard deviation  $\hat{\sigma}^*$  generated by SVR are used to determine the prediction intervals (PIs) (Feng et al., 2019).

5. For each time step, the prediction falls into a certain PI, which is used to estimate the likelihood of it being anomaly (outliers). The deterministic prediction decides the best estimate of next step solar power, while the probabilistic prediction quantifies the uncertainty of all possible observations. A larger PI denotes further deviation from its nominal value. In this study, we assume that an anomaly is spotted whenever the prediction falls out of the 70% PI. This detection threshold can be further tuned through a sensitivity study, which is beyond the scope of this study.

### **7.3 Case Studies and Results**

This section applies the developed probabilistic anomaly detection method to 16 wind farms in Texas, and analyze the results based on different types of data integrity attacks.

#### **7.3.1 Data Summary**

The developed probabilistic anomaly detection framework is evaluated at 16 solar farms in Texas that are selected from the ACTIVSg2000 system, i.e., a 2000-bus synthetic grid on the footprint of Texas (Birchfield et al., 2016). These 16 non-uniformly distributed solar farms are represented by a  $4 \times 4$  array based on their related locations. All the selected solar farms have different capacities which range from 1.05 MW to 230 MW. In this study, the duration of the collected data at the 16 selected solar sites spans four years from January 1<sup>st</sup> 2008 to December 31<sup>st</sup> 2011. For all the selected solar farms, the first 3/4 of the data is used as training data, in which the first 2/3 is used to train the deterministic forecasting model and the remaining 1/3 of the training data is used to train the probabilistic forecasting model. The accuracy of the forecasts is evaluated by the remaining 1/4 of data. The forecast horizon in this study is 1 hour. Note that the data integrity attacks are only applied to the test data, i.e., the test SPSMs.

### 7.3.2 Data Integrity Attack Templates

Four kinds of data integrity attacks in the literature are implemented in this study: scaling attack (SA), correlated attack (CA), replay/swap attack (RSA), and random attack (RA). We assume that these data integrity attacks only apply to the test data, i.e., the forecasting model is healthy.

1. SA: scaling attacks modify the sensor measurement down or up through a scaling attack parameter  $\lambda_s$ .

$$\hat{x}_t = \begin{cases} x_t & \text{for } t \notin \tau_a \\ (1+\lambda_s) \times x_t & \text{for } t \in \tau_a \end{cases} \quad (7.10)$$

where  $\tau_a$  denotes the attack period,  $x_t$  denotes the test tampered with cyberattacks. In this study,  $\lambda_s$  is 0.1 and the scaling attacks are applied to all test periods.

2. CA: correlated attacks replace the measurements with historical data of the same period (i.e., historical data from the same time of a day/same day of a week). In this way, the tampered data will share the same correlation structure. In this study, historical data are separated by day of a week, and we randomly replace the test data with historical data which shares the same period.
3. RSA: replay/swap attacks assume that the attackers possess the test data of a nearby solar farm, which shares similar weather condition. In this study, a nearby solar farm is selected based on the minimum geographical distance. If the capacity of the nearby solar farm is different from the target farm, the solar power is scaled to the same capacity level.
4. RA: random attacks add a random number to the test dataset. This random number could be generated by a uniform random function.

$$\hat{x}_t = \begin{cases} x_t & \text{for } t \notin \tau_a \\ x_t + \lambda_{RA} \times rand(t) & \text{for } t \in \tau_a \end{cases} \quad (7.11)$$

where  $rand(t)$  is a random number generator and  $\lambda_{RA}$  is a scale factor and normally defined as a half of the maximum solar power forecast value (Cui et al., 2019).

### 7.3.3 CNN-LSTM Deterministic Forecasting Results

Since the probabilistic forecasting method used for anomaly detection is a two-step method, it is important to compare the performance of different deterministic forecasting models in the first step. For a single site  $(p, q)$ , standard metrics like root mean squared error and mean absolute error, *i.e.*, RMSE and MAE, are usually adopted to evaluate the performance of deterministic forecasts. They are defined by:

$$MAE(p, q) = \frac{1}{T} \sum_{t=1}^T |\hat{y}(p, q)_t - y(p, q)_t| \quad (7.12)$$

$$RMSE(p, q) = \sqrt{\frac{\sum_{t=1}^T (\hat{y}(p, q)_t - y(p, q)_t)^2}{T}} \quad (7.13)$$

where  $\hat{y}_t$  is the forecasted solar power,  $y_t$  is the solar power observation, and  $T$  is the sample size. However, the solar power forecasts in this study are matrices instead of single values. Therefore, two modified indices, *i.e.*, RMSE for array (RMSE-A) and MAE for array (MAE-A) are adopted to evaluate the overall performance of the model. They are defined by:

$$MAE-A = \frac{1}{P \times Q} \sum_{p=1}^P \sum_{q=1}^Q MAE(p, q) \quad (7.14)$$

$$RMSE-A = \sqrt{\frac{1}{P \times Q} \sum_{p=1}^P \sum_{q=1}^Q RMSE(p, q)^2} \quad (7.15)$$

A smaller MAE-A or RMSE-A indicates better forecasting performance. The 1HA deterministic forecasting errors by using the CNN-LSTM deterministic forecasting model at the selected locations are summarized in Table 7.1. The persistence of cloudiness method (PS) (Cui et al., 2017) has been proved to be accurate in shorter forecasting horizons. Therefore, to show the superiority

of the CNN-LSTM deterministic forecasting model, the PS method is adopted as a baseline. Overall, the accuracies of the CNN-LSTM deterministic forecasts are better than those of PS forecasts under both normal and abnormal circumstances.

Table 7.1: 1HA deterministic solar power forecasting results by using CNN-LSTM and PS

Model	Metric	Scenario				
		SA	CA	RSA	RA	w/o attack
CNN-LSTM	MAE-A	2.33	2.65	2.76	2.81	1.33
	RMSE-A	4.98	5.70	5.73	5.68	4.81
PS	MAE-A	2.52	2.74	2.98	3.09	1.52
	RMSE-A	5.05	5.71	5.82	5.90	5.03

### 7.3.4 Benchmarks and Comparison Settings

In this study, three anomaly detection benchmark models are selected for comparison, including two probabilistic forecasting based models and one deterministic model. The deterministic anomaly detection model is based on error distribution clustering, i.e., the one that doesn't belong to any cluster is marked as anomaly behavior. To guarantee a fair comparison, CNN-LSTM with same parameter settings are adopted as the predictor. It is named as CNN-LSTM-D for comparison in the results analysis. The two probabilistic forecasting benchmark models are quantile regression (QR), and the same CNN-LSTM-based probabilistic forecasting model but without considering spatial-temporal correlation (CNN-LSTM-w/o), i.e., each solar farm is forecasted separately.

The reasons for choosing these three baseline models are: (i) it is important to understand the advantages of probabilistic anomaly detection methods compared with state-of-the-art deterministic methods; (ii) QR is a widely used non-parametric method in probabilistic forecasting, which allows us to explore the effectiveness of non-parametric methods; (iii) since a spatial-temporal model is included in the proposed CNN-LSTM method, it is important to understand the advantages by considering spatial-temporal correlation. Note that the empirical probability distribution is adopted to calculate the quantile forecasts.

### 7.3.5 Probabilistic Forecasting Results

Once probabilistic solar power forecasts are generated, with the estimated empirical predictive PDF of solar power, the quantiles  $q_1, q_2, \dots, q_{99}$  can be calculated. Pinball loss is a widely used metric to evaluate the overall performance of probabilistic forecasts, which is defined by:

$$PL_{p,q} = \begin{cases} \left(1 - \frac{m}{100}\right) \times (q_{m,t} - y_t), & y_t < q_{m,t} \\ \frac{m}{100} \times (y_t - q_{m,t}), & y_t \geq q_{m,t} \end{cases} \quad (7.16)$$

where  $q_{m,t}$  represents the  $m$ th quantile at time  $t$ . However, to evaluate probabilistic forecasts of matrices instead of single values, a modified pinball loss, i.e., pinball loss for array (PL-A) is adopted. For a solar farm located at  $(p, q)$ , PL-A is defined by:

$$PL-A = \frac{1}{P \times Q} \sum_{p=1}^P \sum_{q=1}^Q PL_{p,q} \quad (7.17)$$

To show the effectiveness of the developed probabilistic forecasting framework, the normalized PL-A (NPL-A) of 1HA solar power forecasts of different models under both normal and abnormal circumstances are compared in Table 7.2. The sum of PL-A is averaged over all quantiles from 1% to 99% and normalized by the solar farm capacity. A lower NPL-A score indicates a better probabilistic forecast. Results show that the proposed CNN-LSTM-Gaussian model gives the best performance compared to the two benchmark probabilistic forecasting models, which shows the improvement brought from spatial-temporal correlation modeling and pinball loss optimization. For the cases under data integrity attacks, the NPL-A is much larger than those of normal conditions, it also explains why a reliable and accurate anomaly detection is needed.

### 7.3.6 Probabilistic Anomaly Detection Results

This section evaluates the performance of the proposed probabilistic anomaly detection method. Performance skill scores are calculated based on a contingency table as shown in Table 7.3. True positive (TP) denotes the number of detected attacks; false negative (FN), i.e., type II error, denotes

Table 7.2: Normalized NPL-A under different attack templates

Model	Site				
	SA	CA	RSA	RA	w/o attack
CNN-LSTM-Gaussian	<b>2.68</b>	<b>2.73</b>	<b>2.96</b>	<b>3.13</b>	<b>1.23</b>
CNN-LSTM-w/o	2.90	3.12	3.69	3.99	2.30
QR	2.73	3.03	3.43	3.67	2.12

*Note:* The smallest NPL-A value is in boldface.

the number of cases where missed detection of attacks; false positive (FP), i.e., false alarm or type I error, denotes the number of cases where normal data is treated as attacks; true negative (TN) denotes the number of normal data is correctly identified.  $N_s$  is the total number of test samples. Among these indexes, FP can cause false alarms, which may add redundant work to system operators, while FN missed by the detection model may cause losses to market end users.

Table 7.3: Contingency table of attack detection

	Attack (Yes)	Attack (No)	Total
Detected (Yes)	TP (hit)	FP (miss)	TP+FP
Detected (No)	FN (miss)	TN (hit)	FN+TN
Total	TP+FN	FP+TN	$N_s=TP+FP+FN+TN$

We calculated the true positive rate (TPR), false positive rate (FPR), and F-1 score of the anomaly detection results. Mathematical expressions of the three metrics are expressed as:

$$TPR = \frac{TP}{TP + FN} \quad (7.18)$$

$$FPR = \frac{FP}{FP + TN} \quad (7.19)$$

$$F-1 = \frac{2TP}{2TP + FP + FN} \quad (7.20)$$

where TPR measures the proportion of actual attacks that are correctly identified, FPR measures the portion of normal data mistakenly categorized as attacks, and the F-1 score is the harmonic mean

of the precision and recall. For the TPR and F-1 score metrics, a value approaching 1.0 indicates better performance; while for the FPR metric, a value closer to 0 indicates better performance.

The evaluation metrics of different models are compared and summarized in Table 7.4. Overall, the proposed CNN-LSTM-Gaussian anomaly detection method has a higher TPR, F-1 Score, and lower FPR compared with other anomaly detection methods, which shows the effectiveness of the proposed probabilistic anomaly detection algorithm. Note also that the models of CNN-LSTM-Gaussian perform better than the CNN-LSTM-w/o, which indicates that the optimization can help achieve better detection performance in parametric methods. In addition, it is shown that the scores of SA, CA, RSA are better than that of RA. It is mainly due to the larger solar power magnitude change under RA. Moreover, the false alarm rate of QR is relatively higher than that of other detection methods. It is mainly because the PIs of the QR method is generally wider, and anomaly behaviors may be mistakenly easier treated as normal behavior. It is worth mention that the CNN-LSTM-D deterministic method performs similar to the proposed method. It also explains why the deterministic anomaly detection method dominates the anomaly detection literature due to its simplicity and relatively good performance.

Table 7.4: Probabilistic Anomaly Detection Results

Method	Metrics	Attack Scenario			
		SA	CA	RSA	RA
CNN-LSTM-Gaussian	TPR	<b>0.91</b>	<b>0.82</b>	<b>0.94</b>	<b>0.89</b>
	FPR	<b>0.21</b>	<b>0.23</b>	<b>0.30</b>	<b>0.41</b>
	F-1 score	<b>0.82</b>	<b>0.86</b>	<b>0.78</b>	<b>0.71</b>
CNN-LSTM-w/o	TPR	0.89	0.81	0.88	0.84
	FPR	0.24	0.27	0.33	0.38
	F-1 score	0.80	0.82	0.76	0.70
QR	TPR	0.80	0.77	0.72	0.60
	FPR	0.39	0.48	0.36	0.40
	F-1 score	0.65	0.68	0.58	0.71
CNN-LSTM-D	TPR	0.90	0.80	0.89	0.86
	FPR	0.21	0.24	0.30	0.41
	F-1 score	0.81	0.85	0.76	0.70

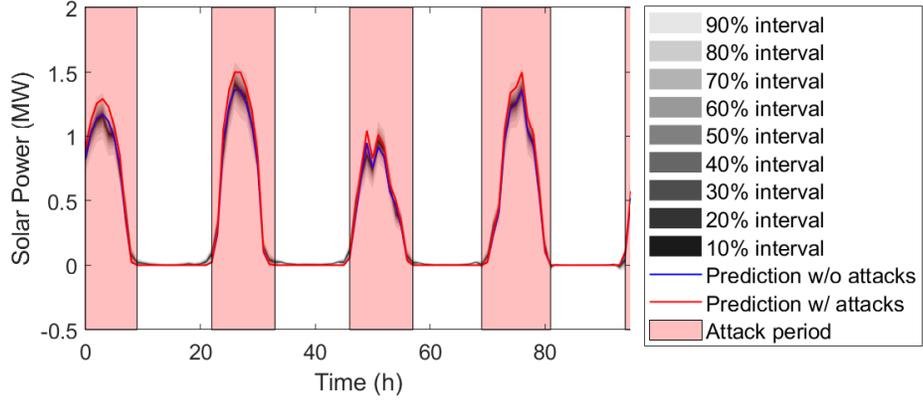
*Note:* The best TPR, FPR, and F-1 score among different models are marked in boldface.

To better visualize the probabilistic anomaly detection results, the PIs of a selected time period under different types of data integrity attacks are illustrated in Fig. 7.5.

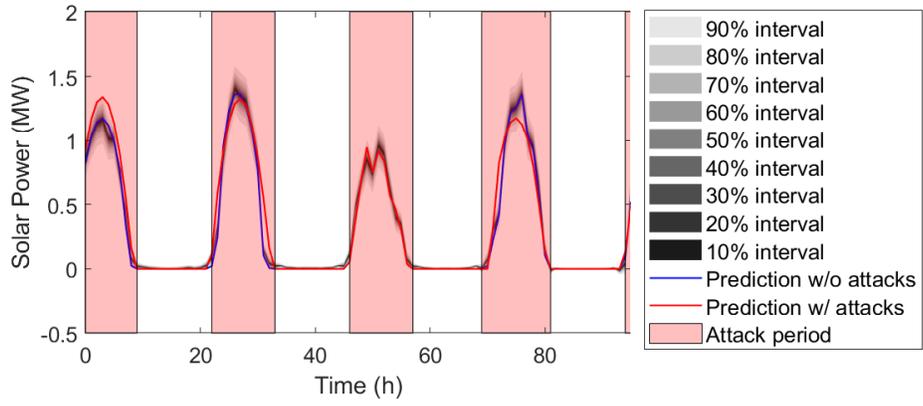
It is observed that at most part of the period without attacks, the solar power reasonably lies within the PIs. When the prediction in the attack period falls out of the 70% PI, it is defined as a truth positive detection. In contrast, a failed detection happens when the prediction falls within the 70% CI during the attack period. It is seen from Fig. 7.5 that the magnitude change of solar power under RA and RSA is higher than that of CA and SA. It is mainly because in SA and CA, spatially correlated solar farms receive similar amount of solar irradiance and experience similar weather effects. However, under both normal and attack scenarios, the high detection accuracy validates the robustness of the proposed method. In addition, the variability of RA is higher than that of other attacks, thereby the uncertainty under RA is relatively higher. Note that the designed data integrity attack frequency and magnitude may largely affect the detection rate. In addition, the selected threshold of PIs for anomaly detection also affects the detection accuracy. Therefore, it is hard to conclude which kind of data integrity attack is relatively easier to be detected by the probabilistic method. It is important to emphasize that the objective of this study is not to design new advisory models or beat state-of-the-art detection algorithms, but to explore the possibility of utilizing probabilistic forecasting for anomaly detection in solar power forecasting.

## **7.4 Conclusion**

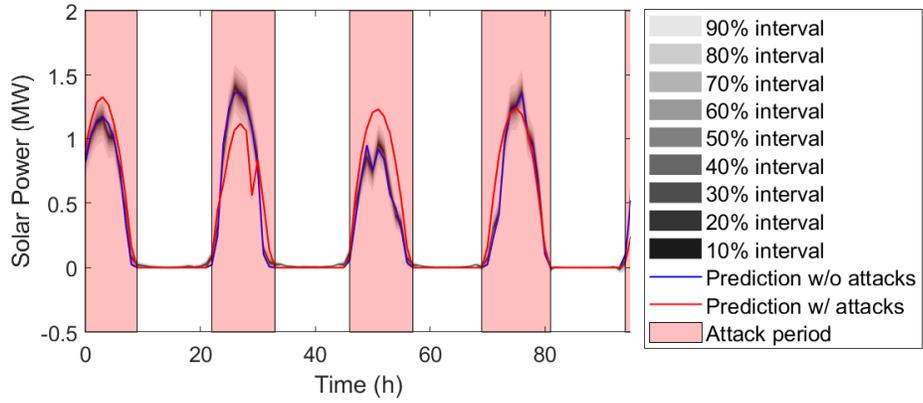
This chapter developed a deep learning-based probabilistic anomaly detection method in solar power forecasting. Results of the case study under different attack scenarios showed that the developed probabilistic anomaly detection method was able to effectively detect different types of cyberattack with high accuracy. The proposed probabilistic anomaly detection framework could be applied in the decision-making process of real-time unit commitment and economic dispatch and to inform the system operator of abnormal events. In addition, this proposed probabilistic anomaly detection framework could also be used to promote the integration of solar power in



(a) SA

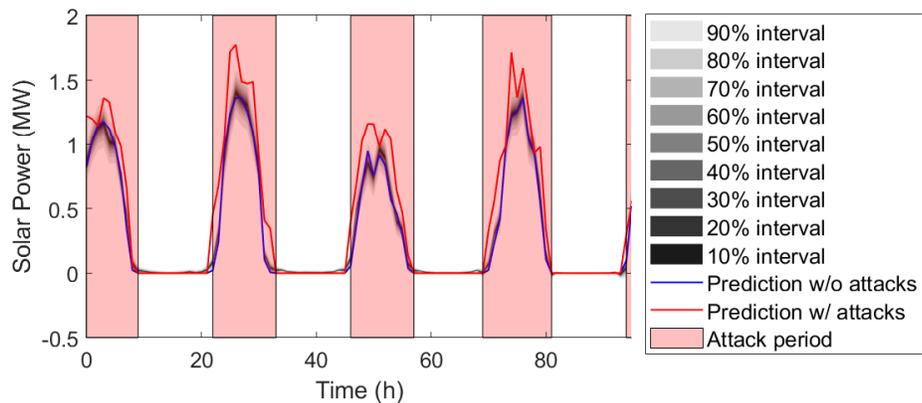


(b) CA



(c) RSA

Figure 7.5: Probabilistic anomaly detection on different attack templates



(d) RA

Figure 7.5: (continued) Probabilistic anomaly detection on different attack templates

power systems by providing a reliable ancillary services. Potential future work will (i) extend the probabilistic anomaly detection model to wind power and load, and (ii) explore the vulnerability of different probabilistic forecasting models against different data integrity attacks.

## CHAPTER 8

### CONCLUSION

Accurate renewable energy forecasts benefit reliable and economic power system operations. However, most current forecasting methods are deterministic, which might not be sufficient to characterize the inherent uncertainty of renewable energy. In summary, the use of probabilistic renewable energy forecasts is still young. The probabilistic renewable energy forecasts are only used in a primitive way in most systems, and there is not a systematic way to integrate them into system operation and scheduling routines. However, operators start to realize the importance of probabilistic renewable energy forecasts in shaping more cost-effective, stable and reliable power systems. In addition, the spatial-temporal correlation among renewable energy farms are not well studied in the literature.

This dissertation presents a formal study of probabilistic renewable energy forecasting by considering spatial-temporal correlation. The study dissects the probabilistic renewable energy forecasting problem into three thrusts. The first two thrusts improved probabilistic renewable energy forecasting accuracy and the third thrust applied the developed probabilistic renewable energy forecasting method to anomaly detection.

In the first two thrusts, we developed several data-driven probabilistic renewable energy forecasting methodologies to enhance the short-term probabilistic wind and solar forecasting with spatial-temporal modeling. In the third thrust, a probabilistic anomaly detection method based on spatial-temporal correlation was proposed against cyberattacks in renewable energy forecasting. The successful development of this research accomplished five major achievements:

- A two-step probabilistic wind forecasting approach based on predictive distribution optimization was developed. The proposed methodology was able to improve the pinball loss by up to 35%.
- A multi-distribution ensemble probabilistic wind power forecasting method was developed. The proposed methodology was able to improve the pinball loss by up to 20.5%

- An aggregated probabilistic wind power forecasting approach was developed based on spatial-temporal correlation among different wind farms. The proposed methodology was proved to enhance the probabilistic wind power forecasting.
- A weather scenario generation-based probabilistic solar power forecasting framework was developed by considering correlations between weather variables. The proposed methodology could improve short-term probabilistic solar power forecasting accuracy.
- The advanced deep learning network, i.e., CNN-LSTM, was utilized to provide probabilistic anomaly detection by modeling spatial-temporal correlation among solar farms. The developed CNN-LSTM network provided high anomaly detection rate and low false alarms.

The proposed methodologies provided more accurate and robust forecasts, compared to the state-of-the-art probabilistic forecasting models. Moreover, the proposed methodologies showed high detection rate and low false alarm rate on anomaly detection problems. Successful deployments of these methodologies are highly potential to mitigate the intrinsic uncertainty in renewable energy integration to power systems. Moreover, with the development of artificial intelligence techniques, large amounts of data are being collected by smart grid devices, such as the advanced metering infrastructures (AMIs) and the phasor measurement units (PMUs), which facilitate the deep applications of AI techniques together with renewable energy forecasting, and thus better assisting power systems.

Beyond these, although probabilistic renewable energy forecasting is adopted in some markets, most of them use it to improve situational awareness and it rarely plays any important role in the decision-making process. By giving a thorough review on the potential methods and latest trends in the literature to promote the integration of probabilistic renewable energy forecasting, we have the following findings. First, many probabilistic methods are proposed in the literature to handle various forms of uncertainties in renewable energy. Naturally, probabilistic renewable energy forecasts can also be integrated into the decision making process of real-world power markets. Each

method has its own advantages and disadvantages and the correct tool should be selected accordingly. For example, the spatial and temporal correlations across the renewable energy productions are important in the modeling process and it may have adverse impact on power system economy and reliability to neglect such correlations. The scenario based methods, such as the MC method and stochastic optimization, are better at capturing such correlations, while they tend to be more computationally demanding and often require scenario reduction. In contrast, robust optimization requires less computation time and can be applied to large scale power systems. However, the solution tends to be conservative. In addition, while probabilistic forecasts can take a variety of forms, such as PDFs, prediction intervals or quantiles, they must be converted into the correct form to be used as input of a specific method. For example, scenarios are used in the Monte Carlo simulation and stochastic optimization while uncertainty sets must be constructed for robust optimization. Therefore, most of the probabilistic renewable energy forecasting methods have their own dedicated area for the representation of model uncertainties. We find that parametric methods are more frequently used in most studies since they are the most fundamental form, and a distribution with closed form is friendly to computation. However, in the literature, non-parametric probabilistic renewable energy forecasting methods, particularly quantiles, dominate the studies of probabilistic renewable energy forecasting. Therefore, there is a clear gap between probabilistic renewable energy forecasting and the use of probabilistic renewable energy forecasting. Particularly, future studies should bridge the gap by placing more emphasis on the use of parametric probabilistic renewable energy forecasting, even the predominant non-parametric methods typically have better performance. To this end, ensemble methods and hybrid methods, which take the advantages of both parametric methods and non-parametric methods, might be more suitable.

It is also worthwhile to mention that most of the existing anomaly detection methods focus on load forecasting and electricity price forecasting, and little work has been done on renewable energy forecasting. One possible reason is the lack of public available well-labelled renewable energy dataset under cyberattacks. Another reason might be that releasing a real-world cyberattack scenario in renewable energy forecasting to public may help attackers design sophisticated enough

attack templates to bypass the detection algorithms. To bridge this gap and promote the development of anomaly detection methods in renewable energy forecasting, simulated dataset under cyberattack in renewable energy field should be paid with more attention.

Potential further extensions on this topic include: (i) quantify vulnerability of different probabilistic forecasting models, (ii) perform probabilistic forecasting under extreme weather conditions, and (iii) explore probabilistic behind-the-meter solar forecasting.

## REFERENCES

- Agoua, X. G., R. Girard, and G. Kariniotakis (2018). Probabilistic model for spatio-temporal photovoltaic power forecasting. *IEEE Transactions on Sustainable Energy*, 1–1.
- Arbelaitz, O., I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition* 46(1), 243–256.
- Baxevani, A. and A. Lenzi (2018). Very short-term spatio-temporal wind power prediction using a censored gaussian field. *Stochastic Environmental Research and Risk Assessment* 32(4), 931–948.
- Bessa, R. J., V. Miranda, A. Botterud, Z. Zhou, and J. Wang (2012). Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy* 40(1), 29–39.
- Birchfield, A. B., T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye (2016). Grid structural characteristics as validation criteria for synthetic networks. *IEEE Transactions on power systems* 32(4), 3258–3265.
- Bludszuweit, H., J. A. Domínguez-Navarro, and A. Llombart (2008). Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems* 23(3), 983–991.
- Botterud, A., Z. Zhi, J. Wang, R. J. Bessa, H. Keko, J. Mendes, J. Sumaili, and V. Miranda (2011). Use of wind power forecasting in operational decisions. Technical report, Argonne National Laboratory (ANL).
- Bouktif, S., A. Fiaz, A. Ouni, and M. Serhani (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies* 11(7), 1636.
- Bracale, A. and P. De Falco (2015). An advanced bayesian method for short-term probabilistic forecasting of the generation of wind power. *Energies* 8(9), 10293–10314.
- Breiman, L. (1996). Bagging predictors machine learning, 24, 123–140.
- Breidl, K., T. Turkington, and M. Stowasser (2015). Simulating daily precipitation and temperature: a weather generation framework for assessing hydrometeorological hazards. *Meteorological Applications* 22(3), 334–347.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review* 78(1), 1–3.
- Bruninx, K. and E. Delarue (2014). A statistical description of the error on wind power forecasts for probabilistic reserve sizing. *IEEE transactions on sustainable energy* 5(3), 995–1002.

- Burnham, K. P. and D. R. Anderson (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.
- Chandola, V., A. Banerjee, and V. Kumar (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)* 41(3), 1–58.
- Cibulka, L., M. Brown, L. Miller, and A. Meier (2012). User requirements and research needs for renewable generation forecasting tools that will meet the need of the caiso and utilities for 2020. *A White Paper Report Prepared by CIEE*.
- Cui, M., C. Feng, Z. Wang, J. Zhang, Q. Wang, A. Florita, V. Krishnan, and B.-M. Hodge (2017). Probabilistic wind power ramp forecasting based on a scenario generation method. In *IEEE Power Energy Society General Meeting, Chicago, IL*.
- Cui, M., V. Krishnan, B.-M. Hodge, and J. Zhang (2018). A copula-based conditional probabilistic forecast model for wind power ramps. *IEEE Transactions on Smart Grid* 10(4), 3870–3882.
- Cui, M., J. Wang, and M. Yue (2019). Machine learning-based anomaly detection for load forecasting under cyberattacks. *IEEE Transactions on Smart Grid* 10(5), 5724–5734.
- Cui, M., J. Zhang, B.-M. Hodge, S. Lu, and H. F. Hamann (2017). A methodology for quantifying reliability benefits from improved solar power forecasting in multi-timescale power system operations. *IEEE Transactions on Smart Grid* 9(6), 6897–6908.
- Delignette-Muller, M. L., C. Dutang, et al. (2015). fitdistrplus: An r package for fitting distributions. *Journal of Statistical Software* 64(4), 1–34.
- Dowell, J. and P. Pinson (2016). Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid* 7(2), 763–770.
- Draxl, C., A. Clifton, B.-M. Hodge, and J. McCaa (2015). The wind integration national dataset (wind) toolkit. *Applied Energy* 151, 355–366.
- El-Dakkak, O., S. Feng, M. Wahbah, T. H. EL-Fouly, and B. Zahawi (2019). Combinatorial method for bandwidth selection in wind speed kernel density estimation. *IET Renewable Power Generation* 13(10), 1670–1680.
- Fang, X., B.-M. Hodge, E. Du, N. Zhang, and F. Li (2018). Modelling wind power spatial-temporal correlation in multi-interval optimal power flow: A sparse correlation matrix approach. *Applied energy* 230, 531–539.
- Fatemi, S. A., A. Kuh, and M. Fripp (2018). Parametric methods for probabilistic forecasting of solar irradiance. *Renewable Energy* 129, 666 – 676.
- Feng, C., M. Cui, B.-M. Hodge, and J. Zhang (2017). A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Applied Energy* 190, 1245–1257.

- Feng, C., M. Sun, and J. Zhang (2019). Reinforced deterministic and probabilistic load forecasting via q-learning dynamic model selection. *IEEE Transactions on Smart Grid*.
- Feng, C. and J. Zhang (2018). Short-term load forecasting with different aggregation strategies. In *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers.
- Feng, C. and J. Zhang (2019). Reinforcement learning based dynamic model selection for short-term load forecasting. *arXiv preprint arXiv:1811.01846*.
- Forbes, C., M. Evans, N. Hastings, and B. Peacock (2011). *Statistical distributions*. John Wiley & Sons.
- Freedman, J. M., J. Manobianco, J. Schroeder, B. Ancell, K. Brewster, S. Basu, V. Banunarayanan, B.-M. Hodge, and I. Flores (2014). The wind forecast improvement project (wfip): A public/private partnership for improving short term wind energy forecasts and quantifying the benefits of utility operations. the southern study area, final report. Technical report, AWS Truepower, LLC, Albany, NY (United States).
- Freeman, J., N. DiOrio, N. Blair, D. Guittet, P. Gilman, and S. Janzou (2019). Improvement and validation of the system advisor model. Technical report, National Renewable Energy Laboratory.
- Gallego-Castillo, C., R. Bessa, L. Cavalcante, and O. Lopez-Garcia (2016). On-line quantile regression in the rkhs (reproducing kernel hilbert space) for operational probabilistic forecasting of wind power. *Energy 113*, 355–365.
- Gamble, C., B. Hodge, P. Edwards, E. K. Chartan, J. Duckworth, A. Shrirang, A. Ayers, Z. Zhou, S.-i. Yim, J. Zhang, et al. (2018). Windview. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States); Argonne . . .
- Giebel, G., R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl (2011). The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, ANEMOS. plus.
- Graves, A., A.-r. Mohamed, and G. Hinton (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 6645–6649. IEEE.
- Haben, S. and G. Giasemidis (2016). A hybrid model of kernel density estimation and quantile regression for gefcom2014 probabilistic load forecasting. *International Journal of Forecasting 32*(3), 1017–1022.
- Hall, A. R. (2005). *Generalized method of moments*. Oxford University Press.

- Hamann, H. F. (2017). A multi-scale, multi-model, machine-learning solar forecasting technology. Technical report, IBM, Yorktown Heights, NY (United States). Thomas J. Watson Research Center.
- Hansen, C. H., C. J. Doolan, and K. L. Hansen (2017). *Wind Farm Noise: Measurement, Assessment, and Control*. John Wiley & Sons.
- Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics* 14(2), 174–194.
- Henderson, M., W. Henson, J. Norden, W. Coste, R. Zavadil, R. Piwko, G. Jordan, G. Hinkle, and N. Miller (2011). Iso new england wind integration study. In *2011 IEEE Power and Energy Society General Meeting*, pp. 1–6. IEEE.
- Hong, T. and S. Fan (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting* 32(3), 914–938.
- IRENA (2019a). Renewable power generation costs in 2018. Technical report, International Renewable Energy Agency, Abu Dhabi.
- IRENA (2019b). Renewable power generation costs in 2018. Technical report, International Renewable Energy Agency, Abu Dhabi.
- Jascourt, S., C. Cassidy, E. Wertz, and T. Hartman (2019). Probabilistic forecasts for solar uncertainty management and mitigation for exceptional reliability in grid operations (summer-go).
- Jin, B. (2008). Fast bayesian approach for parameter estimation. *International Journal for Numerical Methods in Engineering* 76(2), 230–252.
- Jørgensen, J. U. and C. Möhrlen (2008). Aeso wind power forecasting pilot project. Technical report, Technical report, Ebberup, Denmark.
- Juban, J., N. Siebert, and G. N. Kariniotakis (2007). Probabilistic short-term wind power forecasting for the optimal management of wind generation. In *Power Tech, 2007 IEEE Lausanne*, pp. 683–688. IEEE.
- Kantar, Y. M. (2015). Generalized least squares and weighted least squares estimation methods for distributional parameters. *REVSTAT—Stat. J* 13, 263–282.
- Khodayar, M. and J. Wang (2018). Spatio-temporal graph deep neural network for short-term wind speed forecasting. *IEEE Transactions on Sustainable Energy*, 1–1.
- Kim, D. and J. Hur (2018). Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method. *Energy* 157, 211–226.

- Koza, J. R. and J. R. Koza (1992). *Genetic programming: on the programming of computers by means of natural selection*, Volume 1. MIT press.
- Krishna, V. B., G. A. Weaver, and W. H. Sanders (2015). Pca-based method for detecting integrity attacks on advanced metering infrastructure. In *International Conference on Quantitative Evaluation of Systems*, pp. 70–85. Springer.
- Landry, M., T. P. Erlinger, D. Patschke, and C. Varrichio (2016). Probabilistic gradient boosting machines for gefcom2014 wind forecasting. *International Journal of Forecasting* 32(3), 1061–1066.
- Lange, M. (2005). On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors. *Journal of solar energy engineering* 127(2), 177–184.
- Lee, D. and R. Baldick (2016). Probabilistic wind power forecasting based on the laplace distribution and golden search. In *Transmission and Distribution Conference and Exposition (T&D), 2016 IEEE/PES*, pp. 1–5. IEEE.
- Lefèvre, S., C. Sun, R. Bajcsy, and C. Laugier (2014). Comparison of parametric and non-parametric approaches for vehicle speed prediction. In *American Control Conference (ACC), 2014*, pp. 3494–3499. IEEE.
- Lenzi, A., I. Steinsland, and P. Pinson (2018). Benefits of spatiotemporal modeling for short-term wind power forecasting at both individual and aggregated levels. *Environmetrics* 29(3), e2493.
- Li, P., X. Guan, and J. Wu (2015). Aggregated wind power generation probabilistic forecasting based on particle filter. *Energy conversion and management* 96, 579–587.
- Liang, G., S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong (2016). The 2015 ukraine blackout: Implications for false data injection attacks. *IEEE Transactions on Power Systems* 32(4), 3317–3318.
- Lin, Y., M. Yang, C. Wan, J. Wang, and Y. Song (2018). A multi-model combination approach for probabilistic wind power forecasting. *IEEE Transactions on Sustainable Energy* 10(1), 226–237.
- Liu, B., J. Nowotarski, T. Hong, and R. Weron (2015). Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Transactions on Smart Grid* 8(2), 730–737.
- Liu, H., H.-q. Tian, and Y.-f. Li (2012). Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction. *Applied Energy* 98, 415–424.
- Liu, Y., P. Ning, and M. K. Reiter (2011). False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)* 14(1), 1–33.

- Lojowska, A., D. Kurowicka, G. Papaefthymiou, and L. van der Sluis (2010). Advantages of arma-garch wind speed time series modeling. In *Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on*, pp. 83–88. IEEE.
- Lore, K. G., D. M. Shila, and L. Ren (2018). Detecting data integrity attacks on correlated solar farms using multi-layer data driven algorithm. In *2018 IEEE Conference on Communications and Network Security (CNS)*, pp. 1–9. IEEE.
- Louie, H. (2010). Characterizing and modeling aggregate wind plant power output in large systems. In *Power and Energy Society General Meeting, 2010 IEEE*, pp. 1–8. IEEE.
- Malvaldi, A., S. Weiss, D. Infield, J. Browell, P. Leahy, and A. M. Foley (2017). A spatial and temporal correlation analysis of aggregate wind power in an ideally interconnected europe. *Wind Energy* 20(8), 1315–1329.
- Mandal, P., S. T. S. Madhira, J. Meng, R. L. Pineda, et al. (2012). Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. *Procedia Computer Science* 12, 332–337.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* 11(2), 431–441.
- Meitz, M. and P. Saikkonen (2011). Parameter estimation in nonlinear ar-garch models. *Econometric Theory* 27(6), 1236–1278.
- Menemenlis, N., M. Huneault, and A. Robitaille (2012). Computation of dynamic operating balancing reserve for wind power integration for the time-horizon 1–48 hours. *IEEE Transactions on Sustainable Energy* 3(4), 692–702.
- Mohan, N., K. Soman, and S. S. Kumar (2018). A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. *Applied energy* 232, 229–244.
- Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of applied Meteorology* 12(4), 595–600.
- Olah, C. (2015). Understanding lstm networks.
- Opitz, D. and R. Maclin (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research* 11, 169–198.
- Ordiano, J. Á. G., W. Doneit, S. Waczowicz, L. Gröll, R. Mikut, and V. Hagenmeyer (2017). Nearest-neighbor based non-parametric probabilistic forecasting with applications in photovoltaic systems. *arXiv preprint arXiv:1701.06463*.

- Pelikan, E., K. Eben, J. Resler, P. Juruš, P. Krč, M. Brabec, T. Brabec, and P. Musilek (2010). Wind power forecasting by an empirical model using nwp outputs. In *Environment and Electrical Engineering (EEEIC), 2010 9th International Conference on*, pp. 45–48. IEEE.
- Pinson, P. (2012). Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61(4), 555–576.
- Pinson, P., H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl (2009). From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 12(1), 51–62.
- Pinson, P., H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis (2007). Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* 10(6), 497–516.
- Prasad, A. A., R. A. Taylor, and M. Kay (2017). Assessment of solar and wind resource synergy in australia. *Applied Energy* 190, 354–367.
- Qin, Y., K. Li, Z. Liang, B. Lee, F. Zhang, Y. Gu, L. Zhang, F. Wu, and D. Rodriguez (2019). Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. *Applied Energy* 236, 262–272.
- Ren, Y., X. Qiu, and P. N. Suganthan (2014). Empirical mode decomposition based adaboost-backpropagation neural network method for wind speed forecasting. In *Computational Intelligence in Ensemble Learning (CIEL), 2014 IEEE Symposium on*, pp. 1–6. IEEE.
- Ren, Y., P. Suganthan, and N. Srikanth (2015). Ensemble methods for wind and solar power forecasting—a state-of-the-art review. *Renewable and Sustainable Energy Reviews* 50, 82–91.
- Rüschendorf, L. (2009). On the distributional transform, sklar’s theorem, and the empirical copula process. *Journal of Statistical Planning and Inference* 139(11), 3921–3927.
- Schapire, R. E. (1990). The strength of weak learnability. *Machine learning* 5(2), 197–227.
- Scrucca, L. et al. (2013). Ga: a package for genetic algorithms in r. *Journal of Statistical Software* 53(4), 1–37.
- Selvin, S., R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman (2017). Stock price prediction using lstm, rnn and cnn-sliding window model. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1643–1647. IEEE.
- Sengupta, M., Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby (2018). The national solar radiation data base (nsrdb). *Renewable and Sustainable Energy Reviews* 89, 51–60.

- Sheather, S. J. and M. C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)* 53(3), 683–690.
- Shin, J. Y., H.-H. Kwon, J.-H. Lee, and T.-W. Kim (2020). Probabilistic long-term hydrological drought forecast using bayesian networks and drought propagation. *Meteorological Applications* 27(1), e1827.
- Shrestha, D. L. and D. P. Solomatine (2006). Experiments with adaboost. rt, an improved boosting scheme for regression. *Neural computation* 18(7), 1678–1710.
- Sobhani, M., T. Hong, and C. Martin (2020). Temperature anomaly detection for electric load forecasting. *International Journal of Forecasting* 36(2), 324–333.
- Steinwart, I., A. Christmann, et al. (2011). Estimating conditional quantiles with the help of the pinball loss. *Bernoulli* 17(1), 211–225.
- Sun, M., C. Feng, E. K. Chartan, B.-M. Hodge, and J. Zhang (2019). A two-step short-term probabilistic wind forecasting methodology based on predictive distribution optimization. *Applied Energy* 238, 1497–1505.
- Sun, M., C. Feng, and J. Zhang (2019). Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation. *Applied Energy* 256, 113842.
- Sun, M., C. Feng, and J. Zhang (2020a). Multi-distribution ensemble probabilistic wind power forecasting. *Renewable Energy* 148, 135–149.
- Sun, M., C. Feng, and J. Zhang (2020b). Probabilistic solar power forecasting based on weather scenario generation. *Applied Energy* 266, 114823.
- Sun, M., C. Feng, J. Zhang, E. K. Chartan, and B.-M. Hodge (2018). Probabilistic short-term wind forecasting based on pinball loss optimization. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–6. IEEE.
- Takeuchi, I., K. Nomura, and T. Kanamori (2009). Nonparametric conditional density estimation using piecewise-linear solution path of kernel quantile regression. *Neural Computation* 21(2), 533–559.
- Tang, C., Y. Wang, J. Xu, Y. Sun, and B. Zhang (2018). Efficient scenario generation of multiple renewable power plants considering spatial and temporal correlations. *Applied Energy* 221, 348–357.
- Tastu, J., P. Pinson, E. Kotwa, H. Madsen, and H. A. Nielsen (2011). Spatio-temporal analysis and modeling of short-term wind power forecast errors. *Wind Energy* 14(1), 43–60.

- Tewari, S., C. J. Geyer, and N. Mohan (2011). A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets. *IEEE Transactions on Power Systems* 26(4), 2031–2039.
- Wan, C., Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong (2014). Optimal prediction intervals of wind power generation. *IEEE Transactions on Power Systems* 29(3), 1166–1174.
- Wang, C., Z. Liang, J. Liang, Q. Teng, X. Dong, and Z. Wang (2018). Modeling the temporal correlation of hourly day-ahead short-term wind power forecast error for optimal sizing energy storage system. *International Journal of Electrical Power & Energy Systems* 98, 373–381.
- Wang, H.-z., G.-q. Li, G.-b. Wang, J.-c. Peng, H. Jiang, and Y.-t. Liu (2017). Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied energy* 188, 56–70.
- Wang, P. and M. Govindarasu (2019). Cyber-physical anomaly detection for power grid with machine learning. In *Industrial Control Systems Security and Resiliency*, pp. 31–49. Springer.
- Wang, X. and S.-H. Ahn (2020). Real-time prediction and anomaly detection of electrical load in a residential community. *Applied Energy* 259, 114145.
- Wang, X., T. Zhao, H. Liu, and R. He (2019). Power consumption predicting and anomaly detection based on long short-term memory neural network. In *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, pp. 487–491. IEEE.
- Wang, Y., D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang (2019). Probabilistic individual load forecasting using pinball loss guided lstm. *Applied Energy* 235, 10–20.
- Wang, Y., Q. Hu, D. Meng, and P. Zhu (2017). Deterministic and probabilistic wind power forecasting using a variational bayesian-based adaptive robust multi-kernel regression model. *Applied Energy* 208, 1097–1112.
- Wang, Y., N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, and C. Kang (2018). Combining probabilistic load forecasts. *IEEE Transactions on Smart Grid* 10(4), 3664–3674.
- Wang, Z., W. Wang, C. Liu, Z. Wang, and Y. Hou (2018). Probabilistic forecast for multiple wind farms based on regular vine copulas. *IEEE Transactions on Power Systems* 33(1), 578–589.
- Wilczak, J., C. Finley, J. Freedman, J. Cline, L. Bianco, J. Olson, I. Djalalova, L. Sheridan, M. Ahlstrom, J. Manobianco, et al. (2015). The wind forecast improvement project (wfip): A public–private partnership addressing wind energy forecast needs. *Bulletin of the American Meteorological Society* 96(10), 1699–1718.
- Wolf, M. and D. Serpanos (2020). False data injection attacks. In *Safe and Secure Cyber-Physical Systems and Internet-of-Things Systems*, pp. 73–83. Springer.

- Wu, J., B. Zhang, and K. Wang (2012). Application of adaboost-based bp neural network for short-term wind speed forecast. *Power System Technology* 36(9), 221–225.
- Xie, J. and T. Hong (2016). Temperature scenario generation for probabilistic load forecasting. *IEEE Transactions on Smart Grid* 9(3), 1680–1687.
- Xie, L., Y. Gu, X. Zhu, and M. G. Genton (2014). Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Transactions on Smart Grid* 5(1), 511–520.
- Xin, Y., L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang (2018). Machine learning and deep learning methods for cybersecurity. *IEEE Access* 6, 35365–35381.
- Yan, J. et al. (2007). Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software* 21(4), 1–21.
- Yue, M. (2017). An integrated anomaly detection method for load forecasting data under cyberattacks. In *2017 IEEE Power & Energy Society General Meeting*, pp. 1–5. IEEE.
- Yue, M., T. Hong, and J. Wang (2019). Descriptive analytics-based anomaly detection for cyber-secure load forecasting. *IEEE Transactions on Smart Grid* 10(6), 5964–5974.
- Zhang, B., P. Dehghanian, and M. Kezunovic (2016). Spatial-temporal solar power forecast through use of gaussian conditional random fields. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5. IEEE.
- Zhang, G., Y. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H. H.-C. Iu (2014). An advanced approach for construction of optimal wind power prediction intervals. *IEEE transactions on power systems* 30(5), 2706–2715.
- Zhang, J., C. Draxl, T. Hopson, L. Delle Monache, E. Vanvyve, and B.-M. Hodge (2015). Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. *Applied Energy* 156, 528–541.
- Zhang, N., C. Kang, C. Singh, and Q. Xia (2016). Copula based dependent discrete convolution for power system uncertainty analysis. *IEEE Transactions on Power Systems* 31(6), 5204–5205.
- Zhang, N., C. Kang, Q. Xu, C. Jiang, Z. Chen, and J. Liu (2013). Modelling and simulating the spatio-temporal correlations of clustered wind power using copula. *Journal of Electrical Engineering & Technology* 8(6), 1615–1625.
- Zhang, Y. and J. Wang (2015). Gefcom2014 probabilistic solar power forecasting based on k-nearest neighbor and kernel density estimator. In *Power & Energy Society General Meeting, 2015 IEEE*, pp. 1–5. IEEE.

- Zhang, Y. and J. Wang (2018). A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without direct access to off-site information. *IEEE Transactions on Power Systems* 33(5), 5714–5726.
- Zhang, Y., J. Wang, and X. Luo (2015). Probabilistic wind power forecasting based on logarithmic transformation and boundary kernel. *Energy conversion and management* 96, 440–451.
- Zhang, Y., J. Wang, and X. Wang (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews* 32, 255–270.
- Zhang, Z.-S., Y.-Z. Sun, D. W. Gao, J. Lin, and L. Cheng (2013). A versatile probability distribution model for wind power forecast errors and its application in economic dispatch. *IEEE transactions on power systems* 28(3), 3114–3125.
- Zhao, T., Y. Zhang, and H. Chen (2018). Spatio-temporal load forecasting considering aggregation features of electricity cells and uncertainties in input variables. *Journal of Electrical Engineering & Technology* 13(1), 38–50.
- Zhao, Y., L. Ye, P. Pinson, Y. Tang, and P. Lu (2018). Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. *IEEE Transactions on Power Systems* 33(5), 5029–5040.
- Zheng, K., Q. Chen, Y. Wang, C. Kang, and Q. Xia (2018). A novel combined data-driven approach for electricity theft detection. *IEEE Transactions on Industrial Informatics* 15(3), 1809–1819.
- Zheng, R., J. Gu, Z. Jin, H. Peng, and Y. Zhu (2019). Load forecasting under data corruption based on anomaly detection and combined robust regression. *International Transactions on Electrical Energy Systems*, e12103.
- Zhu, Q., J. Chen, D. Shi, L. Zhu, X. Bai, X. Duan, and Y. Liu (2019). Learning temporal and spatial correlations jointly: A unified framework for wind speed prediction. *IEEE Transactions on Sustainable Energy* 11(1), 509–523.

## **BIOGRAPHICAL SKETCH**

Mucun Sun received his Bachelor of Engineering degree in Electrical Engineering from Huazhong University of Science & Technology, Wuhan in June 2014, and his Master of Science degree in Electrical Engineering from New York State University at Buffalo in March 2016. His research is focused on machine learning/deep learning-based probabilistic renewable energy forecasting, anomaly detection, and power system operation.

## CURRICULUM VITAE

# Mucun Sun

### Contact Information:

Department of Mechanical Engineering  
The University of Texas at Dallas  
800 W. Campbell Rd.  
Richardson, TX 75080-3021, U.S.A.

Website: mucun.tech  
Email: mucun.sun@utdallas.edu

### Educational History:

B.S., Electrical Engineering, Huazhong University of Science and Technology, 2014  
M.S., Electrical Engineering, University at Buffalo, 2016  
Ph.D., Mechanical Engineering, The University of Texas at Dallas, 2020

### Employment History:

Research Intern, United Technologies Research Center, Aug 2019 – Jan 2020  
Contractor, Oncor Electric Delivery LLC, May 2019 – Aug 2019  
Software Engineer Intern, National Renewable Energy Laboratory (NREL), May 2018 – Aug 2018  
Data Analyst Intern, State Grid of China, July 2014 – Aug 2014

### Journal Publications:

**Mucun Sun**, Jie Zhang. “Conditional Aggregated Probabilistic Load Forecasting based on Spatial-temporal Correlation.” (under review).

**Mucun Sun**, Jie Zhang. “Deep Learning-Based Probabilistic Anomaly Detection for Solar Farms under Cyberattacks.” (under review).

**Mucun Sun**, Cong Feng, Jie Zhang. “Probabilistic Solar Power Forecasting Based on Weather Scenario Generation.” *Applied Energy*, Vol. 266, 2020, pp. 114823.

**Mucun Sun**, Cong Feng, Jie Zhang. “Multi-Distribution Ensemble Probabilistic Wind Power Forecasting.” *Renewable Energy*, Vol. 148, 2020, pp. 135-149.

**Mucun Sun**, Cong Feng, Erol Kevin Chartan, Bri-Mathias Hodge, Jie Zhang. “A Two-Step Short-Term Probabilistic Wind Forecasting Methodology Based on Predictive Distribution Optimization.” *Applied Energy*, Vol. 238, 2019, pp. 1497-1505.

**Mucun Sun**, Cong Feng, Jie Zhang, “Conditional Aggregated Probabilistic Wind Power Forecasting Based on Spatio-temporal Correlation.” *Applied Energy*, Vol. 256, 2019, pp. 113842.

Cong Feng, **Mucun Sun**, Jie Zhang, “Reinforced Deterministic and Probabilistic Load Forecasting via Q-Learning Dynamic Model Selection.” *IEEE Transactions on Smart Grid*, 2019 Vol. 11, Issue 2, 2020, pp. 1377-1386.

Cong Feng, **Mucun Sun**, Mingjian Cui, Erol Kevin Chartan, Bri-Mathias Hodge, Jie Zhang. “Characterizing forecastability of wind sites in the United States.” *Renewable Energy*, Vol 133, 2019, pp. 1352-1365.

Qingjun Huang, **Mucun Sun**, Xudong Zou, Liangchun Liu, Wei Xiong. “Design of the Motor Simulating System Based Power Electrical Load.” *Automation of Electric Power System*, 38.3 (2013): 71-77.

## **Conference Publications**

**Mucun Sun**, Lingyu Ren, Nai-yuan Chiang, “Data-Driven Probabilistic Anomaly Detection for Electricity Market under Cyber Attacks.” submitted to IEEE Probabilistic Methods Applied to Power Systems (PMAPS2020).

Cong Feng, **Mucun Sun**, Jie Zhang, Kate Doubleday, Bri-Mathias Hodge, Peng Du, “A Data-driven Method for Adaptive Reserve Requirements Estimation via Probabilistic Net Load Forecasting.” IEEE Power&Energy Society General Meeting (PES2020).

**Mucun Sun**, Cong Feng, Jie Zhang, “Factoring Behind-the-Meter Solar into Load Forecasting: Case Studies under Extreme Weather.” IEEE Innovative Smart Grid Technologies Conference (ISGT 2020), Washington D.C., February 17-20, 2020.

**Mucun Sun**, Cong Feng, Jie Zhang, “Aggregated Probabilistic Wind Power Forecasting Based on Spatio-Temporal Correlation.” IEEE Power&Energy Society General Meeting, Atlanta, Aug 4-8, 2019.

**Mucun Sun**, Cong Feng, Jie Zhang, Erol Kevin Chartan, Bri-Mathias Hodge, “Probabilistic Short-term Wind Forecasting Based on Pinball Loss Optimization.” Probabilistic Methods Applied to Power Systems Conference (PMAPS), Boise, Idaho, June 24-28, 2018.

**Mucun Sun**, Chih-Lun Chang, Jie Zhang, Ali Mehmani, Patricia J. Culligan, “Break-even analysis of battery energy storage in buildings considering time-of-use rates.” 2018 IEEE Annual Green Technologies Conference, Austin, TX, April 4-6, 2018. **(Best Paper Award)**

Kaige Zhu, Souma Chowdhury, **Mucun Sun**, Jie Zhang, “Grid Optimization of Shared Energy Storage Among Wind Farms Based On Wind Forecasting.” IEEE PES Transmission&Distribution Conference&Exposition, Denver, CO, April 16-19, 2018.

Qingjun Huang, **Mucun Sun**, Zou Xudong, Tong Li, Xiong Wei, Chen Jianqing, “A Reverse Current Tracking Based LVRT Strategy for Doubly Fed Induction Generator (DFIG)” The 39th Annual Conference of the IEEE Industrial Electronics Society 2013.

## **Book and Book Chapters**

Book Chapter: **Mucun Sun**, Jie Zhang, “Data-driven Anomaly Detection in Modern Power Systems,” Book Title: “Security for Cyber-Physical Systems: Vulnerability and Impact,” Springer, 2020. (in press)

Book Chapter: Cong Feng, **Mucun Sun**, Morteza Dabbaghjamanesh, Jie Zhang, “Advanced Machine Learning Applications to the Modern Power Systems,” Book Title: “New Technologies for Power System Operation and Analysis,” Elsevier, 2019. (in press)

### **Talks and Presentations:**

“Aggregated Probabilistic Wind Power Forecasting Based on Spatio-Temporal Correlation.” The University of Texas at Dallas Spring 2019 Poster Competition, Richardson, TX, April 2019.

“Probabilistic Short-term Wind Forecasting Based on Pinball Loss Optimization.” The 39th International Symposium on Forecasting, Boulder, CO, June 2018.

“Grid Optimization of Shared Energy Storage Among Wind Farms Based On Wind Forecasting.” WindSTAR IAB meeting, Dallas, TX, Nov 2016

### **Professional Services:**

**Journal Reviewer:** IEEE Transactions on Smart Grid, IEEE Systems Journal, Applied Energy, Renewable Energy, IEEE Transactions on Industry Applications, IEEE Transactions on Cyber-Physical Systems

**Conference Reviewer:** IEEE GreenTech 2019