



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

**QUANTIFICATION OF THE FORECASTING
UNCERTAINTY IN SMART GRID**

SONGJIAN CHAI

PhD

The Hong Kong Polytechnic University

2018

The Hong Kong Polytechnic University

Department of Electrical Engineering

**Quantification of the Forecasting Uncertainty in
Smart Grid**

SONGJIAN CHAI

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

Jun 2018

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

_____ Songjian CHAI _____ (Name of student)

*To my parents,
Xiaofeng Chai and Qing Guo,
and my wife,
Jie Shao*

Abstract

In the last two decades, the surging proliferation of renewable generations and the inception of competitive electricity markets worldwide have forced the decision makers to reconsider the planning, operation and trading mechanisms in modern power system. For renewable generations, such as wind power and solar photovoltaic power, the power output is characterized by variability and intermittence due the nature of chaotic weather conditions. While for electricity prices in competitive markets, the fundamental reasons behind are much more complex, the net load variability, system congestions, fuel prices and CO₂ allowances are always considered as the major contributors to the uncertainty of electricity price. All these factors drove the grid operators and energy traders to seek a powerful forecasting product to aid their decision-making processes.

Over the years, extensive works have been carried out on point (or deterministic) forecasts, which only gives one plausible estimate of the future. However, such forecasts are limited as they fail to inform the inevitable error information involved, which is fairly crucial for sagacious decision makings considering diversified uncertainties. This boosts the shift towards a more informative forecast tool under a probabilistic framework in recent years. In a nutshell, the uncertainty needs to be properly quantified as inputs fed into the specific applications of interest in one of the popular forms: quantiles, prediction intervals, PDF/CDF and scenarios.

This thesis concerns three types of them, i.e., prediction intervals, PDFs and scenarios, with respect to two vital forecasting tasks in Smart Grid, i.e., prognosis of

solar irradiance and market clearing prices. The research background and purpose are presented in Chapter 1. Chapter 2 gives a comprehensive review of the state-of-the-art techniques for the main forecasting activities in Smart Grid (e.g. wind power, solar photovoltaic power and electricity price). Subsequently, inspired by the fundamentals of information granules, a reliable prediction interval construction framework based on temporal granules is proposed for very short-term solar irradiance forecasts in Chapter 3. In Chapter 4, an effective density forecast approach based on ensemble extreme learning machines and a parametric post-processing technique is presented, which gives a full description of the underlying uncertainty involved in the day-ahead forecasts of Swedish market clearing prices. To further facilitate the generation of time trajectories, an efficient covariance structure determination method is developed to model the temporal dependency in the latter part of this chapter. Chapter 5 concludes the whole thesis and indicates the related aspects that can be enhanced and extended in the future.

List of Publications Arisen from the Thesis

Technical Papers in Refereed Journals

- [1] **S. Chai**, Z. Xu, and W. K. Wong, "Optimal Granule-Based PIs Construction for Solar Irradiance Forecast," *IEEE Transactions on Power Systems*, vol. 31, pp. 3332-3333, 2016.
- [2] **S. Chai**, Z. Xu and Y. Jia, "Conditional Density Forecast of Electricity Price based on Ensemble ELM and Logistic EMOS", has been accepted by *IEEE Transactions on Smart Grid* on 18th March, 2018, not published yet.
- [3] Y. Jia, C. S. Lai, Z. Xu, **S. Chai**, and K. P. Wong, "Adaptive partitioning approach to self-sustained smart grid," *IET Generation, Transmission & Distribution*, vol. 11, no. 2, pp. 485-494, 2017.
- [4] X. Xu, J. Zhao, Z. Xu, **S. Chai**, J. Li, and Y. Yu, "Stochastic Optimal TCSC Placement in Power System Considering High Wind Power Penetration," *IET Generation, Transmission & Distribution*, 2018.

Conference Papers in Refereed Proceedings

- [1] **C. Songjian**, X. Zhao, L. Loi Lei, and W. Kit Po, "An overview on wind power forecasting methods," in *Machine Learning and Cybernetics(ICMLC), 2015 International Conference on*, 2015, pp. 765-770.
- [2] **S. Chai**, Y. Jia, Z. Xu, and Z. Dong, "The Granule-Based Interval Forecast for Wind Speed," in *Proceedings of ELM-2015 Volume 2*, ed: Springer, 2016, pp. 263-272
- [3] **C. Songjian**, N. Ming, Z. Xu, L. Loi Lei, and K. P. Wong, "Nonparametric conditional interval forecasts for PV power generation considering the temporal dependence," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1-5

Acknowledgements

This work is accomplished at the Department of Electrical Engineering, The Hong Kong Polytechnic University (HKPU) during my Ph.D. candidature period from 2015 to 2018.

First and foremost, I would like to express my sincerest gratitude to my supervisor, Professor Zhao Xu, for his guidance, knowledge, support and patience throughout my entire graduate study. Back to 2014, when I was working as a technician in Beijing, Professor Zhao Xu, as the ex-supervisor in my Master career, offered me this valuable opportunity to pursue a higher level of degree. Since then, he continued to encourage and support me in both life and academy. It is his precious supervision that guides me to be engaged in the research of probabilistic forecast.

My heartfelt thanks go to my beloved wife, Ms Jie Shao, who accompanied me throughout my graduate career. With the continuous support and care from her, I was successful to get through a lot of agonizing times during the past years. In addition, I deeply thank my parents, Mr X.F.Chai and Ms Q.Guo, for their unconditional love, understanding, trust and encouragement despite the long distance between us.

The friendship of my lab-mates is much appreciated, especially for Mr M.Niu, Mr Y.W Jia and Mr J.Y Li, who gave me valuable advices and selfless help. Special thanks are given to Mr W.Can, who is currently the “Hundred Talents Program” Research Professor in Zhejiang University, for his expertise and insights on probabilistic prognosis, which is essential to the completion of this work.

Last but not the least, I would like to knowledge the support from Ph.D studentship awarded by The Hong Kong Polytechnic University.

Table of Contents

Abstract.....	iii
List of Publications Arisen from the Thesis	v
Acknowledgements.....	vii
Table of Contents	ix
List of Figures.....	xi
List of Tables.....	xiii
Abbreviations.....	xv
Chapter 1 Introduction.....	1
1.1 Background of Research.....	1
1.2 Motivation and Purpose.....	6
1.3 Scientific Contributions	9
1.4 Thesis Layout.....	11
Chapter 2 Literature Review.....	13
2.1 Introduction.....	13
2.2 Beyond Point forecasts: Probabilistic Forecasts.....	15
2.2.1 Why probabilistic forecasts? Benefits for different entities	15
2.2.2 Forms of probabilistic forecasts.....	19
2.2.3 Evaluation tools for probabilistic forecasts.....	23
2.3 A Review on State-of-the-art Probabilistic Forecasts for VRE Generations and Electricity Prices.....	33
2.3.1 Parametric methods.....	34
2.3.2 Nonparametric methods.....	39
2.4 Summary.....	46
Chapter 3 Granule-based Interval Forecast for Solar Irradiance.....	47
3.1 Introduction.....	47
3.2 Granular Computing	50
3.3 Experimental Data and Empirical Investigation	51
3.4 Extreme Learning Machine (ELM)	54
3.5 Uncertainty Analysis.....	57
3.5.1 Stochastic uncertainty.....	57
3.5.2 Knowledge uncertainty.....	58
3.6 Formulation of Granular Prediction Model on the basis of ELM.....	60
3.6.1 Granular GHI time series.....	61
3.6.2 Granular ELM-based NN.....	64

3.6.3	PI-skill-oriented training strategy	65
3.7	Experimental Results	71
3.7.1	Benchmarking methods	71
3.7.2	Comparative studies and discussion	77
3.8	Summary	88
Chapter 4	Conditional Density Forecast of Electricity Price based on Ensemble ELM and Logistic EMOS	93
4.1	Introduction.....	93
4.2	The Data.....	97
4.3	Quantification of Uncertainty in Swedish MCP Forecasting.....	99
4.3.1	Feature Selection.....	100
4.3.2	Ensemble Forecasts on the basis of ELM Networks	105
4.3.3	Logistic Distribution based EMOS	107
4.3.4	Parameter Estimation	109
4.4	Temporal Dependency Modeling.....	110
4.5	Results.....	113
4.5.1	Determination of ensembles' dimensionality and EMOS's training periods.....	114
4.5.2	Discussion of the predictive density via ELC-EMOS	115
4.5.3	Comparison with Gaussian EMOS model	118
4.5.4	Comparison with benchmarks of density forecasts	121
4.5.5	Day-ahead scenarios generation and evaluation	130
4.6	Summary	136
Chapter 5	Conclusions and Future Scope.....	137
5.1	Conclusions.....	137
5.2	Future Scope	140
Appendix	B-1
Reference	B-5

List of Figures

Fig. 1.1 Global cumulative capacity of wind and solar PV power, 2006 - 2016	2
Fig. 1.2 Decision-making activities with respect to different forecasting horizons	3
Fig. 2.1 τ -th quantile of the predictive CDF for variable X	20
Fig. 2.2 Schematic of quantile score (QS)	26
Fig. 2.3 Schematic of CRPS	28
Fig. 3.1 Time series of 1-min GHI, measured at King's Park Meteorological Station on 28th Mar 2018 in Hong Kong. A snapshot downloaded from HKO website [134]	53
Fig. 3.2 2D plot of GHI throughout the year of 2012	53
Fig. 3.3 From numeric time series to granular time series.....	62
Fig. 3.4 Architecture of GELM.....	65
Fig. 3.5 ACE Scatters of different approaches over 10 months in 2012 in terms of different PINCs.....	85
Fig. 3.6 Snapshot of 90% PIs associated with the 10-min average GHI observed in 10 successive days in November 2012.....	87
Fig. 4.1 Four bidding areas (SE1, SE2, SE3, SE4) in Sweden with corresponding MCP at certain hour on 23rd Apr 2018.....	98
Fig. 4.2 Schematic of proposed forecasting strategy	99
Fig. 4.3 Timeline of Nord Pool Elspot.....	100
Fig. 4.4 Autocorrelation coefficients of Swedish MCP in 2013. Grey lines mean data that are not available at forecast.....	103
Fig. 4.5 Correlation coefficients between MCP and its most recent price over 24 hours in 2013.....	103
Fig. 4.6 Intraday profiles of average MCP in terms of different weekdays in 2013.	104
Fig. 4.7 Typical generalization performance with respect to different ELM structures.	106
Fig. 4.8 (a) The potential ensemble predictions and real MCPs on 17th Nov 2015. (b) Histogram of potential ELM's ensemble errors and the corresponding	

Normal and Logistic fits.....	108
Fig. 4.9 Schematic diagram of inverse transformation for a realization in predictive margins.	111
Fig. 4.10 Forecasting skill over period from 1st Jul 2014 to 31st Jan 2015 in terms of different training days and dimensionalities of ensembles.....	115
Fig. 4.11 3D visualization of predictive densities of Swedish MCPs on 28th Jul 2016	116
Fig. 4.12 Time-adaptive model fitting strategy.....	116
Fig. 4.13 MCP series with prediction intervals in weeks with different price patterns.	117
Fig. 4.14 Quantile reliability diagram for the Normal and Logistic models.....	121
Fig. 4.15 Average CRPS in terms of 24 hours	128
Fig. 4.16 Average ES (EUR/MWh) and VS-0.5 ([EUR/MWh] ^{0.5}) of scenarios obtained by different approaches. Blue bar and red bar represent ES and VS-0.5, respectively	133
Fig. 4.17 Empirical correlation matrix of the transformed gaussian random variables of the observed probabilistic forecasts for 1st Nov 2015. Highest to lowest correlations are represented by yellow to blue colors.	134
Fig. 4.18 20 representative day-ahead scenarios for different daily pattern.	135

List of Tables

Table 3.1 Skill of proposed GELM model against benchmarks in terms of 90% PI over different months in 2012.	80
Table 3.2 Skill of proposed GELM model against benchmarks in terms of 95% PI over different months in 2012.	81
Table 3.3 Skill of proposed GELM model against benchmarks in terms of 99% PI over different months in 2012.	82
Table 3.4 Means and standard deviations of ACEs in terms of different PINCs over 10 months in 2012.	83
Table 3.5 Skills of different granularities and different granulation methods evaluated in November 2012.	88
Table 4.1 Selected features for Swedish MCP.	104
Table 4.2 Average CRPS (EUR/MWh) of Normal- and Logistic- EMOS Models in the Context of Different Weekly Scenarios.	118
Table 4.3 Test Statistics and Probability of DM Test of Equal Forecast Performance for Comparison of Logistic-EMOS and Gaussian-EMOS.	120
Table 4.4 Average CRPS (EUR/MWh) against benchmarking models on a monthly Basis. Smallest CRPS values are marked in bold.	127
Table 4.5 Test Statistics and Probability of DM Test of Equal Forecast Performance for Comparison of ELC-EMOS and Benchmarks.	127
Table 4.6 Testing for Quantile Reliability by Averaging APDs over Quantiles From 0.01 to 0.99 with 0.1 Increment.	128
Table A.1 Skills of 90% PI obtained by BS-QR model with past 10-min GHI averages (BS-QR-V1) versus that with past 1-min raw data (BS-QR-V2).	B-1
Table A.2 Skills of 90% PI obtained by BELM model with past 10-min GHI averages (BELM-V1) versus that with past 1-min raw data (BELM-V2).	B-1
Table A.3 Skills of 90% PI obtained by DIF model with past 10-min GHI averages (DIF-V1) versus that with past 1-min raw data (DIF-V2).	B-2
Table B.1 Input variable candidates for Swedish MCP.	B-2

Abbreviations

VRE	Variable Renewable Energy
PV	Photovoltaic
MCP	Market Clearing Price
CI	Confidence Interval
PI	Prediction Interval
HKO	Hong Kong Observatory
CDF	Cumulative Distribution Function
PDF	Probability Density Function
PICP	Prediction Interval Coverage Probability
PINC	Prediction Interval Nominal Coverage
ACE	Average Coverage Error
APD	Absolute Probabilistic Deviation
CRPS	Continuous Ranked Probability Score
PINAW	Normalized Average Width of Prediction Intervals
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARMAX	Autoregressive Moving Average with Exogenous Variables
GARCH	Generalized Autoregressive Conditional Heteroskedastic
IHMAR	Iterated Hsieh-Manski Estimator based Autoregressive
SNAR	Smoothed Nonparametric Maximum Likelihood Estimator Based Autoregressive
QRA	Quantile Regression Averaging
GAMLSS	Generalized Additive Models for Location, Scale and Shape
NN	Neural Network
ELM	Extreme Learning Machine
SLFN	Single Hidden-layer Feedforward Neural Network
MLE	Maximum Likelihood Estimation
CWC	Coverage-width-based Criterion
NWP	Numerical Weather Prediction
AI	Artificial Intelligence
GHI	Global Horizontal Irradiance
DNI	Direct Normal Irradiance
SVM	Support Vector Machine
KDE	Kernel Density Estimation
QR	Quantile Regression
GBM	Gradient Boosted Machines
GHI	Global Horizontal Irradiance
IG	Information Granule
GELM	Granular Extreme Learning Machine
EMOS	Ensemble Model Output Statistics

DIF	Direct Interval Forecast
PSO	Particle Swarm Optimization
FCM	Fuzzy C-means Clustering
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
LASSO	Least Absolute Shrinkage and Selection Operator
LOLP	Loss of Load Probability
LOLE	Loss of Load Expectation
EENS	Expected Energy Not Served
COPT	Capacity Outage Probability Table
CVaR	Conditional Value at Risk
MAE	Mean Absolute Error
ES	Energy Score
VS	Variogram Score
IGN	Ignorance Score

Chapter 1 Introduction

1.1 Background of Research

In the last decade, the world has noticed an explosion of renewable generations, particularly in 2016, the renewable power generating capacity saw its largest annual increase ever than before, with 161 gigawatts (GW) of capacity added, to nearly 2017 GW at the end of 2016 [1]. Among various types of renewable generations, such as wind power, photovoltaic (PV) power, hydropower, biomass, biofuel, geothermal energy and so forth [2], the cumulative global capacity of the variable renewable generations, such as wind power and PV power, accounted for 39% of the total renewable capacity and experienced a tremendous increment over the last decade, which can be seen in Fig. 1.1, from 80 GW in 2006 to 790 GW in 2016 [1]. The top countries for highest penetration levels of wind power and solar PV power are Denmark and Honduras, respectively. In Denmark, the wind power has met 37.6% of electricity demand in the whole country, and in Honduras, 9.8% of the total electricity consumption came from the PV power [1].

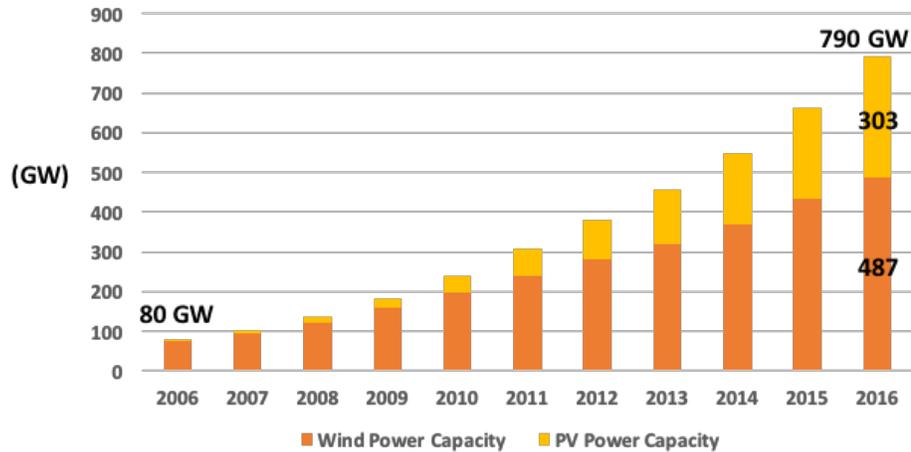


Fig. 1.1 Global cumulative capacity of wind and solar PV power, 2006 - 2016

However, compared to the dispatchable power sources, the weather-dependent nature gives rise to large uncertainty and variability of the wind power and PV power output, both issues can lead to unexpected consequences, such as the fluctuations of voltage and frequency, increasing demand for ancillary services and economic loss for energy traders. It has been pointed out that there is distinct difference between the terms “variability” and “uncertainty” [3]. “Variability” refers to the change of generation output due to fluctuations of wind or sun, while “uncertainty” describes the unpredictability the timing and magnitude of the changes in generation output. In this sense, application of accurate forecasting tools is indispensable to reduce the uncertainty of variable renewable energy (VRE) generation, so that its variability can be more precisely accommodated [3].

The forecasting tools for VRE generations are tailored to different end-using cases and forecasting horizons, as illustrated in Fig. 1.2, where very short-term and short-term forecasting achieve high popularity in Smart Grid.

Very short-term forecasting is normally used for power smooth, real-time dispatch, storage control and real-time energy trading (e.g., 5 minutes for Australian electricity market clearing). Short-term forecasting refers to the forecasts performed up to 24-72 hours in advance. Such forecasts are particularly crucial for the decision-making activities relating to system operation (e.g., unit commitment, economic dispatch, reserve setting) and day-ahead energy bidding. Medium-term forecasting is useful for the maintenance scheduling of power grid and VRE plants. Long-term forecasting can be applied for the siting and planning of VRE plants. On the other hand, the purposes for different end users may be distinct. A forecast tuned for an individual energy producer might be designed to maximize its profits, whereas a regional forecast made by a system operator might aim to minimize risk and to secure the system reliability and security [4].

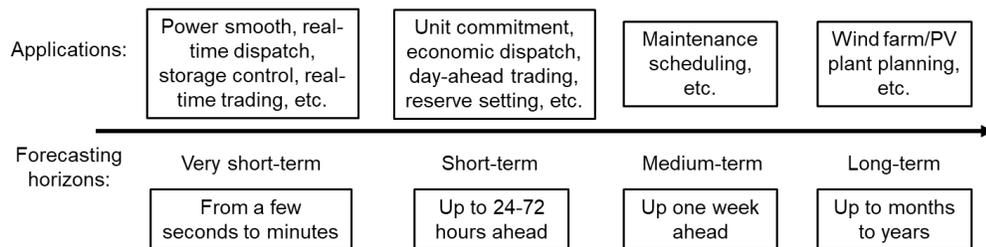


Fig. 1.2 Decision-making activities with respect to different forecasting horizons

In the meanwhile, over the last two decades, the electricity markets have experienced a significant restructuring and deregulation worldwide [5]. Even China, as the world's largest electricity producer, embarks on breaking through the past monopoly of electricity market up to 30 more years, and attempts to build up spot markets nationally by 2020 [6]. In deregulated market, the

electricity became a commodity that can be traded between producers and buyers through free competitions [5]. As a result, the electricity price was no longer regulated by certain departments as a fixed tariff but was determined by the offers provided by the producers and customers, which varies over the market horizon. Market participants acting on the energy exchange require accurate electricity price forecasts to maximize their profits [7]. For instance, power producers must optimize the use of their production portfolio by pricing and bidding their available production capacity into the market. On the other hand, demand-side participants look for feasible options to avoid the high electricity prices during peak hours. However, due to the distinct properties of electricity, that is the supply and demand should be matched instantaneously and it cannot be stored in an efficient and cost-effective manner [8], the electricity price exhibits high volatility. Moreover, the surging integration of renewable energy into grid and limited capacity of transmission lines can even give rise to extreme prices, this made the forecasts of electricity price more challengeable and unpredictable than that of load series, which indicates strong seasonal patterns.

Over the past years, the bulk of forecasting activities concerning VRE generation and electricity price is carried out in the deterministic framework [9-11], which provides a single value for a certain lead time. However, this single value is only one representative of the multiple likely outcomes, and subject to evitable errors. Sometimes, the level of errors is acceptable within the risk threshold of the related applications, but sometimes it can be

significant to cause unexpected consequences. It is thus desired to inform the error information in the forecasting results in some reasonable and understandable manner, that can be made use of by decision makers to hedge against the risks and make optimal decisions. For example, to compensate the uncertainties of net load, the system operators have to procure a certain amount of reserve in advance, which is directly linked to the VRE generation forecasting errors [12]. Excessive reserve would lead to an economic loss and an inadequate amount of reserve could result in a potential reliability issue. Therefore, better knowledge of the associated error information with the point estimate in advance can significantly assist the grid operators to make a justified tradeoff between economics and risk management.

To account for the uncertainty information, several prevailing operational solutions have been developed lately to facilitate the end-users' problems, such as stochastic programming [13-18] and robust optimization [19-24], where the quantitative uncertainty information is normally served as input variables. Then, the key challenge translates to seek a proper tool to quantify the uncertainty information. Probabilistic forecast is a promising way to fulfil this task, which is favored by a great number of grid operators and market participants nowadays, as it gives a more informative description than point forecast with respect to the future uncertainty. In probabilistic forecast, the potential likely outcomes are given with different probabilities, end-users can easily know the occurring likelihood of certain event and the range of all possible outcomes. Incorporating probabilistic forecast into decision-making

processes is believed to benefit more for the end-users [25-30]. Therefore, in recent years, a surging interest has shifted to the research of probabilistic forecasts, particularly in the field of wind power. Generally, these studies can be categorized as quantiles [31], probabilistic intervals [32], predictive probability density function (PDF) [33] and scenarios [34].

1.2 Motivation and Purpose

At present, the probabilistic forecast for wind power has reached a relatively mature stage, considerable efforts have been spent by academics and practitioners in the last decade, to develop a variety of approaches or models concerning wind power [35]. Whereas the research on the forecast of solar energy and electricity price in the probabilistic regime still situates at a relative infancy. As mentioned-above, as increased PV generations, either at system level or custom level (behind-the-meter), are integrated into the grid, as well as the gradual reforming towards competitive electricity markets worldwide, it also necessitates advanced models to inform of the future uncertainty of PV power and electricity price for different related entities. Although the mechanism of wind power forecasting can be applicable to these two quantities, some unique properties [7, 10] make the prognosis of them even more challengeable. This motivates us to make a comprehensive investigation for PV energy and electricity price, respectively.

This thesis aims at proposing effective solutions for analyzing, modeling and forecasting the potential uncertainties involved in the forecasting process for PV energy and electricity price. Specifically, three complementary

solutions are developed, including very short-term prediction interval (PI) construction for solar irradiance, day-ahead conditional density forecast for market clearing price (MCP) and temporal scenario generation for MCP. The solutions can be applied to the specific operational applications depending on the problem formulations.

Very short-term PI construction for solar irradiance: Instead of directly predicting the solar PV power, we concentrate on the forecast of solar irradiance, as it is the primary source of PV generation. Besides, we believe that solar irradiance forecasting provides more utility than PV power output forecasting. For example, prior knowledge of the amount of solar irradiance over time under the local environmental conditions is a key input for choosing the optimal location, technology and size of a solar energy project. For the short-term respective, the solar irradiance can be converted into the solar power by a series of equations [10], offering the guidance for dispatch and operation in the near future. On the other hand, the expansive deployment of micro grids makes the very short-term (from a few seconds to minutes) forecast for the solar irradiance of particular importance, since the online dispatch order is normally made within a short period (e.g. 15 minutes). In light of above, the uncertainty quantification for the short-term solar irradiance forecasting is indispensable. In this work, a granule-based nonparametric PI constructing method is proposed. The irradiance data used is measured at King's Park Meteorological Station, provided by Hong Kong Observatory (HKO). The developed PI of solar irradiance allows the micro-grid operators to make the

optimal dispatch strategy for other generation units involved, reducing the risks associated with the forecasting errors.

Day-ahead conditional density forecast for MCP: The ‘electricity price’ or ‘spot price’ appeared in previous surveys typically refers to MCP, which is determined by the centralized bidding in a day-ahead market. The bids and offers are submitted by the agents for the delivery of electricity during each hour (or a shorter load period) of the next day before a certain market closing time [9]. Great success has been achieved in the domain of deterministic MCP forecast during last three decades [9]. However, the forecasting errors are impossible to eliminate, even with highly accurate models. As a growing number of practitioners realize that the limitations of point forecasts, probabilistic forecast of MCP becomes popular in recent few years. As Amjady et. al [36] remark, high-quality MCP probabilistic forecasts would help utilities to submit effective bids with low risk. In this thesis, to give the end-users a full picture of all potential future outcomes, we build up a hybrid model for the entire distribution estimation, instead of merely a single or a set of PIs/quantiles.

Temporal scenario generation for MCP: The current forecasting activities for MCP, like discussed above, are mostly carried out for a certain look-ahead time individually, while ignoring the interdependence in the look-ahead horizons. However, some problem formulations (e.g., stochastic optimization) for trading electricity require a portfolio of price trajectories over the successive lead times as inputs [37, 38], rather than several independent

marginal distributions. In this sense, it is necessary to develop effective approaches to construct possible realizations of the stochastic prices. In this work, by taking advantages of the well-calibrated marginal distribution derived from the proposed density forecasting model for MCP, we further apply the Gaussian Copula [39] to capture the essential temporal interdependence among these marginals. To avoid arbitrary hypothesis and complicated computations, a highly efficient determination process for the covariance structure is developed.

1.3 Scientific Contributions

To achieve the original solutions presented in the last section, this thesis investigates into some novel approaches to fulfil these solutions. The primary scientific contributions are summarized as follows:

1. The paradigm of information granules (IGs) is firstly incorporated into the construction of PIs. IGs are treated as collections of entities (say numeric readings) that are grouped together because of their similarity, functional closeness or any other criterion that captures a feature of indistinguishability [40]. It has been widely used in the field of control and decision-making. Considering the traditional artificial intelligence (AI) based PI construction models all depend on crisp (single numeric) inputs and interval outputs [41-43], which is not consistent in model establishment. Based on the conceptualization of IG, granulating both model inputs and model parameters seems to be a more straightforward and rational way to yield PIs, since the stochastic uncertainty involved in the

original data and knowledge uncertainty resulting from the misspecification of model parameters can be well captured by the proper IGs, respectively. Hence, in this thesis, a generic granular neural network (NN) design scheme for PI construction is proposed. The granular mapping is built upon the extreme learning machine (ELM) network due to its excellent generalization capabilities and fast learning speed, through endowing granular outputs with proper scoring rules, the PIs with different confidence levels can be yielded. Numeric experiments are carried out using the global horizontal irradiance (GHI) measurements provided by HKO, results show that the proposed granular ELM (GELM) is able to produce highly reliable and sharp PIs for very short-term GHI prediction.

2. A hybrid model based on ensemble ELM and Logistic Ensemble Model Output Statistics (EMOS) is developed to construct the conditional predictive density for Swedish MCP. Density is more quantitatively informative than quantiles or PIs, particularly at the distribution tails. The instable property of ELM network motivates us to quantify the model uncertainty by combining multiple predictions via a bunch of ELM-NN networks. Logistic distribution is employed to model the predictive distribution as it is more robust to the outliers than Gaussian distribution. By further utilizing the Continuous Ranked Probability Score (CRPS) oriented EMOS to establish the relationship between distribution parameters and ensemble predictions, the MCP density can be forecasted in a time-adaptive way. Comprehensive skill verifications, including CRPS,

quantile reliability and Diebold-Mariano (DM) test, for the predictive density are carried out against benchmarks. Results show that the developed hybrid model can yield predictive density with superior quality.

3. To model the aggregated uncertainties within contiguous lead times, a highly efficient determination process for the covariance structure is developed. Gaussian Copula theory [39] allows us to model the essential interdependence in the lead times once marginal distributions are obtained. The most crucial task is to determine a sound covariance structure for the multivariate gaussian variable. Without any hypothetical parameterization or enumerative computations in traditional approaches, the covariance structure is simply modeled by the observed probabilistic forecasts in the similar weekdays. Through verification via two complementary scores, energy score and variogram score, this dependency modeling method proves to yield most skillful trajectories with comparison to benchmarks.

1.4 Thesis Layout

The remainder of the thesis is organized as follows,

Chapter 2 provides a brief survey of the state-of-the-art works related to this study. Besides, the fundamentals and practical values of probabilistic prognosis are discussed with several examples in Smart Grid. In the last part of this chapter, popular types of probabilistic forecasts are presented along with the corresponding proper evaluation criteria.

Chapter 3 proposes a granule-based nonparametric PI construction model for GHI. The theoretical background of granular computing is presented at first.

The GHI data used for numeric study and its empirical investigation are described later on. ELM as the basis model is also reviewed in this chapter. It is followed by the uncertainty analysis in a modeling system. Setup of the proposed GELM and the associated PI-skill-oriented learning scheme is presented afterwards. Lastly, the evaluation is carried out against five benchmarks, showing the effectiveness of the proposed GELM model.

In Chapter 4, a density constructing model using ensemble ELM and Logistic EMOS is developed for day-ahead MCP forecast. Appropriate input features are selected firstly from a family of candidates for each ensemble member. Then, the establishment of Logistic EMOS and its parameters' tuning strategy in terms of CRPS is presented. The proposed modeling process for the temporal dependence structure of the next-day predictive margins is also described in this chapter. In the end of this chapter, the evaluations are performed with regard to the derived independent predictive densities and the time trajectories, respectively.

Eventually, Chapter 5 concludes this thesis, and provides perspectives for future work.

Chapter 2 Literature Review

2.1 Introduction

With the surging integration of VRE into power system, the grid operators and market participants are facing new challenges arising from the undispachable nature of VRE and the volatility of MCPs. All entities are seeking for the competitive forecast products to aid their decision-making processes. The prevailing practice for them is to produce a point estimate for a certain lead time, which is closely related to the forecasting quality, inferior quality yields large forecasting error that could cause severe consequences for the related applications. Therefore, the decision makers come to realize that simply knowing one single outcome is not enough, the associated uncertainty should be properly accounted for in the complex decision-making processes [44]. This leads to the emergence of probabilistic forecast, which aims to give a quantitative description of the uncertainty information.

Among a series of crucial forecasting activities in Smart Grid, wind power probabilistic forecasting has reached a mature stage, this might be due to the expansive integration of wind farms into power grid in the early 21th century, particularly in Denmark, where it accounts for a large share of the total electricity supply, from 12.1% in 2000 to 49.2% in 2015 [45]. On the other hand, until recent few years, a dramatic increase of PV generations was seen throughout the world, as shown in Fig. 1.1. Hence, the investigation of PV power probabilistic forecasting is in the early stage. In addition, the

'immaturity' of PV power forecasting might results from its small-scale and geographical spreading characteristics. The PV power producer and independent system operator (ISO) are more inclined to know the approximating aggregated generations from a PV plant or a large control zone for a certain time. The individual small-scale PV generation is hard to predict as it is highly influenced by the unique weather condition and cloud motion in the specific region. On the contrary, the aggregated PV production is much more predictable, since the intermittent variations in the outputs of a PV plant or a zone can be cancelled out by each individual component. Consequently, in spite of the rapid development of PV generations worldwide, vast number of works still focus on the point forecast of aggregated PV power output.

Similarly, although the competitive electricity market has been incepted over years, the probabilistic forecasting of electricity price still gains limited interest among academics and practitioners. This situation continues up to year 2014, when the Global Energy Forecasting Competition (GEFCom2014) was hold. The price track in this competition attracts 287 contestants worldwide, and this makes the beginning of the era of probabilistic electricity price forecast [46]. Nevertheless, the competition is not the fundamental reason, rather the effect of increased interest in probabilistic price forecasting.

In this chapter, we first discuss the key role of probabilistic forecast in today's decision-making problems in power system, with the practical values tailored to different applications. Then, four typical representations of probabilistic forecasts are introduced, followed by the review of popular

evaluation metrics with regard to all representations. In Section 2.3, a comprehensive overview of state-of-the-art approaches of probabilistic forecast is given in terms of parametric and nonparametric methods. Conclusion is made in Section 2.4.

2.2 Beyond Point forecasts: Probabilistic Forecasts

2.2.1 Why probabilistic forecasts? Benefits for different entities

In today's electricity market, with more variable and distributed resources flooded into the grid, decision makers are facing miscellaneous uncertainties than ever before. As a result, it seems less rational to solve the decision-making problems in a deterministic framework, analysing and translating these uncertainties into readily understandable forms would substantially aid to make a sound decision. Representing the uncertainty information (e.g., prediction errors) in the form of probabilistic forecasts has attracted great attentions of market participants, particularly for VRE power producers and system operators [35, 47]. VRE power producers pursue the maximization of their profits, whereas the system operators are responsible for the stability of power system on a continuous basis in a cost-effective way. The related decision-making processes include energy bidding [24, 35, 47, 48], operating reserve allocation [25, 26, 49, 50], unit commitment and economic dispatch [19, 27, 28, 51, 52].

(1) VRE power producers

The competitive electricity market requires the power producers to submit the bidding prices and amount of energy for the next operational day, usually ahead of 12-36 hours. The imbalance cost is a major influenced factor of VRE power producers' revenue, which is directly linked to the quality of VRE forecast. The maximal revenue could not be obtained unless point forecast of VRE are perfect and error-free [35]. Nonetheless, this can be achieved by managing the forecasting uncertainty information of VRE via some advanced mathematical optimization tools.

Literature [35] provides an optimal bidding strategy for wind power producers based on stochastic optimization [18], where the uncertainty of wind power production is modeled as a predictive distribution. A closed-form solution for the optimal bidding amount is eventually derived as a function of certain quantile of the predictive distribution. Through simulating in a multi-MW wind farm in Dutch electricity market, it is proved that the bidding strategy derived from wind power probabilistic predictions can reduce the regulation costs by 39% as compared to that resorting to some advanced point forecast approaches [47].

In [24], both forecasting uncertainty of wind power and MCP are taken into account and represented by 95%-level prediction intervals, respectively. Then, a robust formulation is developed considering both day-ahead market and balancing market to hedge the risk for wind power producers. In this way, the resulting revenue can be guaranteed to be maximal under the worst uncertainty scenario.

Robust optimization [21] has its merits on less computational burden and relaxed requirements for predictive distributions. However, it is often argued by the conservative nature and not as well linked to fundamental axioms of rational decisions under uncertainty, as it plans for worst-case scenarios rather than an expected utility metric [44].

(2) System operators

One of the major task for system operators is to define the appropriate amount of operating reserves to be allocated to maintain a specified level of system security. In other words, enough reserve should be procured to restore the balance of supply and demand in the event of unexpected deviations from the forecasting system operation state [53]. In the history, the system uncertainties mainly come from two sources: the possibility of large generators failing and the load forecasting errors. Hence, the reserve amount can be set by simple rules based on the size of largest generator or some fractions of the total load [54]. However, as substantially increasing VRE generations integrated into grid, rules-of-thumb techniques are evidently too conservative and uneconomical to define the reserve levels, sophisticated quantification and management approaches are required to cater for the ever-changing uncertainties. In principle, the optimal level is defined by specific risk indices, such as loss of load probability (LOLP), loss of load expectation (LOLE) and expected energy not served (EENS), within the cost constraints in response to unscheduled deviations in generation and demand [53].

Literatures [25, 49] develop a decision tool to aid the day-ahead setting of operating reserves. The density forecasts of wind power and load, as well as a capacity outage probability table (COPT) and an outage replacement rate are used as inputs, the reserve level is finally determined by a risk/reserve curve and a risk/reserve cost curve according to the decision-makers' preferences. This tool has been installed in the Portuguese Transmission System Operator (TSO), providing suggested reserve allocations during day-ahead and intra-day market sessions [49].

Literature [55] considers the uncertainty in an urban micro-grid arising from PV generations and load, the net demand forecasting errors are modeled through Gaussian distribution. The impact of net demand uncertainties to the system reliability are assessed under two risk indices: LOLP and EENS. Similar to the previously presented work, risk/reserve curve is obtained for each hour of the operational day, which guides the optimal setting of reserves for next day according to different risk indices.

In literature [50], the uncertainties of wind power forecasts, load forecasts and power plant outages are modeled by scenarios, where the scenarios of wind production and load are generated together to account for their mutual correlation. Then, two approaches are framed to determine the reserve requirements in DK1 area of Nord Pool. The first one depends a chosen value of LOLP to deliver a set level of security, regardless of the cost. The second method considers both risk (the expected LOLP) and the associated costs, where the costs are defined as the sum of the expected costs of allocating and

deploying reserve and the expected costs of shedding load. Through minimizing the conditional value at risk (CVaR) at a given risk-parameter, the corresponding solution can be obtained. The numerical study shows the CVaR-based method is more likely to yield a less expensive schedule as compared to the Danish TSO's simple schedule, while procuring adequate levels of reserves.

A stochastic formulation is developed in [27] to minimize the total costs of unit commitment, including expected production costs, the expected cost of unserved energy and reserve curtailment, and start-up costs. Seeing that the unit commitment problem has a strong time-dependency component in its definition (ramping, shut-down/start-up decisions), the uncertainty in wind power prognosis is modeled by the predictive scenarios with temporal interdependence of the forecasting errors instead of the independent distribution. Through numerical experiments, the stochastic unit commitment relying on the scenario representation of uncertainty is demonstrated to have advantages over deterministic approaches that mimic the classical models.

Note that the beneficiaries of probabilistic forecasts are not limited to the energy producers and system operators, more operational problems (e.g., demand response, EV charging) in demand side are needed to exploit to accommodate various types of probabilistic forecasts.

2.2.2 Forms of probabilistic forecasts

Generally, the probabilistic forecasts applied in current decision-making cases of power system can be categorized into four forms: quantiles,

probabilistic intervals, PDF/cumulative distribution function (CDF) and scenarios. To a certain extent, they can be transformed mutually.

2.2.2.1 Quantiles

Quantiles refer to the points in a distribution that relate to the rank order of values in that distribution. In the context of quantile forecast, assuming that \hat{F}_t is the predictive CDF for variable X , as illustrated in Fig. 2.1, the forecasted quantile $\hat{q}_t^{(\tau)}$ with proportion $\tau \in [0,1]$ of this CDF is defined as the value y_t such that $\mathbb{P}(X \leq y_t) = \tau$. In general, the forecasted quantiles can be obtained by using the quantile regression models [56] or directly extracting from the predictive CDF.

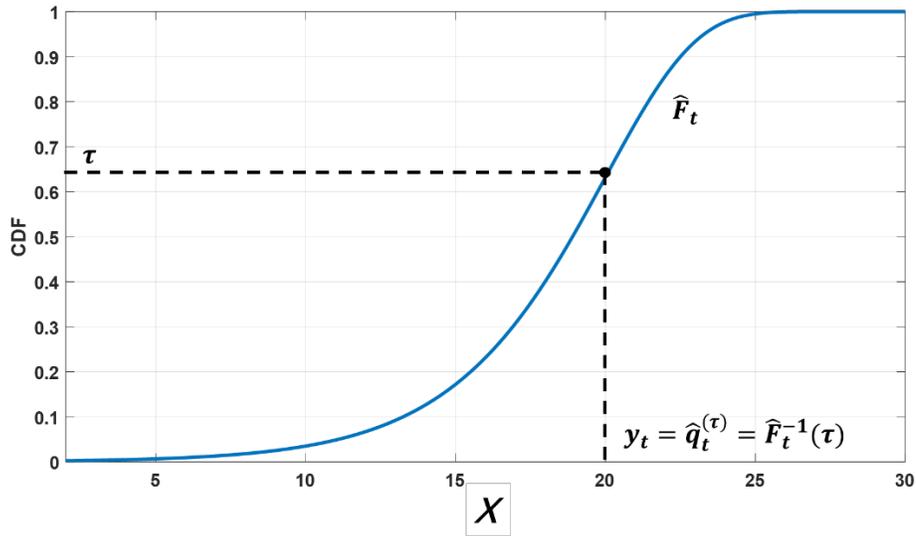


Fig. 2.1 τ -th quantile of the predictive CDF for variable X

2.2.2.2 Probabilistic intervals

In the statistical analysis of power system, probabilistic interval commonly refers to “confidence interval (CI)” and “prediction interval (PI)”,

which are totally different in conception and are always used erroneously by researchers [57]. A confidence interval is a range for the expectation of certain random variable. Suppose one randomly sample from a dataset of MCP and calculate the 90% CI is from 20 EUR/MWh to 40 EUR/MWh, it means that there is 90% confidence that the mean of the entire MCP dataset falls within this range. When it comes to prediction, the CI describes a range that is likely to contain the mean response of a well-established forecast engine. On the other hand, PI is a range associated with a random variable yet to be observed, within a specified probability of the random variable lying within the range. For example, a 90% PI informs that one can be 90% confident that the future measurement can fall into this range. PI can be constructed either in a direct manner or by two different quantiles. For instance, a 90% PI can be derived by 5%-quantile and 95%-quantile as its lower and upper bound, such PIs centered on the median of the predictive distribution are also called central prediction intervals [47].

2.2.2.3 PDF and CDF

The predictive PDF and CDF aim to give the full quantitative description of the uncertainty, from which one can be informed of any densities and quantiles throughout the entire distribution. Most importantly, such representations are capable of describing the extreme situations in the forecasts by distribution tails. To obtain the predictive PDF and CDF, one can simply rely on the parametric ways, that is, calculating the statistical parameters of the specified predictive distribution. Alternatively, they can be derived using a set

of forecasted individual quantiles along with some interpolation techniques. The latter one has its merits on the free-assumption of distribution shape, and thus being able to give a more rational description of the uncertainty information, since the time- and location-varying uncertainty does not always follow a specific distribution in practice. Nonetheless, it suffers from high computational burden and requiring additional modeling of the distribution tails.

2.2.2.4 Scenarios

Although PDF and CDF provide considerable information about the inherent uncertainty in forecasting processes, they fail to inform about interdependence structure at different locations and/or lead times. The reason is that marginal PDF and CDF in the probabilistic forecast context are generated on per-site, per lead time basis.

In this sense, if operational problems are spatially and/or temporally coupled, the predictive marginal distributions would be only suboptimal inputs to decision-making. Stochastic unit-commitment is an example for an operational problem which requires information about uncontrollable generation spread over a control area as well as the development of forecast errors through successive lead times. This drives us to seek a way to model the interdependence in all look-ahead times and all locations of interest. Copula [39] is powerful tool to allow us to specify the dependency between random variables independently of their marginal distributions, and thus establishing the joint distribution. By sampling from this joint distribution, a series of

trajectories can be generated, which are further fed into the multivariate stochastic process of interest as inputs.

2.2.3 Evaluation tools for probabilistic forecasts

Over the years, a number of metrics has been developed to evaluate the performance of probabilistic forecasts. In essence, these evaluation tools rely on the comparison between the observed (actual) values and the probabilistic forecasts over a certain period. In a broad view, they are designed based on two statistical criteria: reliability and sharpness [58-60].

Gneiting et al. state that ‘probabilistic forecasting aims to maximize the sharpness of the predictive distribution, subject to reliability’. Reliability (also called calibration or unbiasedness) refers to the statistical consistency between the distributional forecasts and the observations. Taking a PI estimation for example, the resultant PI is said to be reliable (well-calibrated or unbiased) if a 90% PI covers 90% of the observations over the calibration period. Sharpness, on the other hand, indicates how tightly the predictive distribution covers the actual one. In the following, we will discuss the most-commonly used evaluation metrics in terms of different forms of probabilistic forecasts.

2.2.3.1 Reliability

(1) Interval reliability

Reliability is always deemed as the foremost attribute of the probabilistic forecasting quality [42]. It is widely used in the evaluation of PI and predicted quantiles, which measures the deviation of predicted proportion from the

nominal proportion. The smaller the deviation is, the higher reliability the PIs or predicted quantiles possesses.

For the PIs, by the definition of reliability, the empirical coverage rate (also called PI coverage probability, PICP) should match the nominal coverage rate (also called PI nominal coverage, PINC): $\mathbb{P}(y_t \in [\hat{L}_t, \hat{U}_t]) = (1 - \alpha)$. For instance, the 90% PIs (i.e., with $\alpha = 10\%$) should yield the nominal coverage of 90%. To obtain PICP, we typically focus on an indicator I_t , implying whether the observation lies into the PI or not.

$$I_t = \begin{cases} 1 & \text{if } y_t \in [\hat{L}_t, \hat{U}_t] \\ 0 & \text{if } y_t \notin [\hat{L}_t, \hat{U}_t] \end{cases} \quad (2.1)$$

where y_t is the observation at time t , \hat{L}_t and \hat{U}_t are the lower bound and upper bound of the PI at time t , respectively. Through evaluating over a calibrated period of length T , the PICP can be calculated as

$$\text{PICP} = \frac{1}{T} \sum_{t=1}^T I_t \quad (2.2)$$

To measure the deviation from PINC, the average coverage error (ACE = PICP - PINC) is introduced. Apparently, to yield a PI with high reliability, the absolute value of ACE should be as close to zero as possible [42].

(2) Quantile reliability

The predictive individual quantiles can also be assessed in the context of reliability. Likewise, an indicator $C_t^{(\tau)}$ is introduced to identify that if the observation y_t is lower than the predicted τ -th quantile $\hat{q}_t^{(\tau)}$ or not:

$$C_t^{(\tau)} = \begin{cases} 1 & \text{if } y_t \leq \hat{q}_t^{(\tau)} \\ 0 & \text{if } y_t > \hat{q}_t^{(\tau)} \end{cases} \quad (2.3)$$

Thus, we can write the predicted proportion $\hat{\tau}$ by averaging the summation of $C_t^{(\tau)}$ over the calibrated period:

$$\hat{\tau} = \frac{1}{T} \sum_{t=1}^T C_t^{(\tau)} \quad (2.4)$$

Similarly, $\hat{\tau}$ should be kept as close as possible to its nominal proportion τ to guarantee a high reliability of the derived quantiles. Hence, we introduce the absolute probabilistic deviation (APD = $|\tau - \hat{\tau}|$) to measure their difference [61].

2.2.3.2 Sharpness

Sharpness determines the concentration of the predictive distributions. Diagnostic tools regarding sharpness involve the box plot and average width of PI. The latter is more straightforward once the lower and upper bound of PI are known, which is usually given as [62]:

$$\text{PINAW} = \frac{1}{TR} \sum_{t=1}^T (\hat{U}_t - \hat{L}_t) \quad (2.5)$$

where R is the maximum range of the targets. Narrower PIs or predictive distributions would be more intuitively appealing to decision-makers. Unlike reliability, which is a joint property of the predictions and observations, sharpness only concerns the predictions [46]. However, it is more justified to

verify sharpness along with reliability through skill score for probabilistic forecasts [60].

2.2.3.3 Skill score

Skill score provides a summary measurement for reliability and sharpness simultaneously.

(1) Quantile score

The overall performance of quantiles is often evaluated by the quantile score (QS), on the basis of the most-popular pinball loss function [63], which is expressed by an asymmetric piecewise linear function as:

$$QS(\hat{q}_t^{(\tau)}, y_t, \tau) = \begin{cases} (1-\tau)(\hat{q}_t^{(\tau)} - y_t) & \text{if } y_t < \hat{q}_t^{(\tau)} \\ \tau(y_t - \hat{q}_t^{(\tau)}) & \text{if } y_t \geq \hat{q}_t^{(\tau)} \end{cases} \quad (2.6)$$

Fig. 2.2 gives the shape of quantile score. As can be seen, this loss function is always positive, the further away from the target y_t , the larger the value of quantile score. The slope is used to reflect the desired imbalance in the quantile forecast [63].

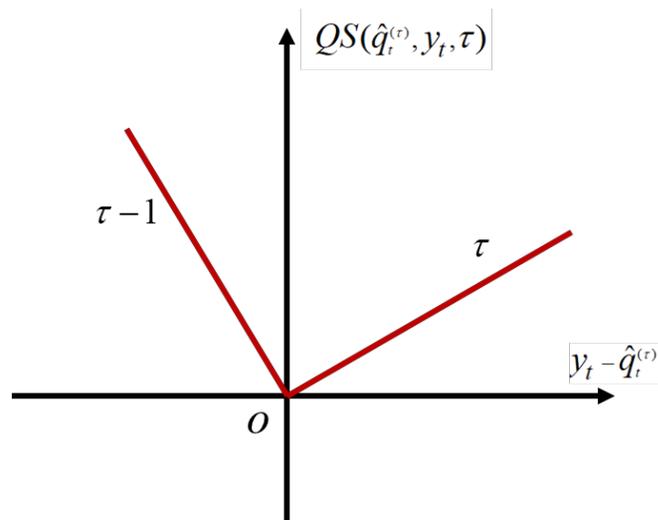


Fig. 2.2 Schematic of quantile score (QS)

In practice, this score is usually computed over a set of forecasted quantiles instead of a single one over the calibrated period, thus yielding the average quantile score as:

$$\overline{QS} = \frac{1}{M \cdot T} \sum_{t=1}^T \sum_{m=1}^M QS(\hat{q}_t^{(r)}, y_t, \tau_m) \quad (2.7)$$

where M is the total number of quantiles verified. A lower score indicates a better quantile forecast.

(2) Interval score

When evaluating the overall skill of PIs, the interval score (also called Winkler score) is extensively used [58]. For a central $(1 - \alpha) \times 100\%$ PI, whose lower and upper bound are given by the predictive quantile at level $\alpha/2$ and $(1 - \alpha/2)$, its interval score is defined as:

$$IS(\hat{L}_t, \hat{U}_t, y_t, \alpha) = \begin{cases} -2\alpha(\hat{U}_t - \hat{L}_t) - 4(\hat{L}_t - y_t) & \text{if } y_t < \hat{L}_t \\ -2\alpha(\hat{U}_t - \hat{L}_t) & \text{if } y_t \in [\hat{L}_t, \hat{U}_t] \\ -2\alpha(\hat{U}_t - \hat{L}_t) - 4(y_t - \hat{U}_t) & \text{if } y_t > \hat{U}_t \end{cases} \quad (2.8)$$

This score is also negative-oriented. It rewards the forecaster for a narrow PI and gives a penalty if the PI does not include the observation. When multiple non-crossing PIs, with nominal coverage rate from $(1 - \alpha_1)$ to $(1 - \alpha_M)$, are examined, we adopt the average interval score as:

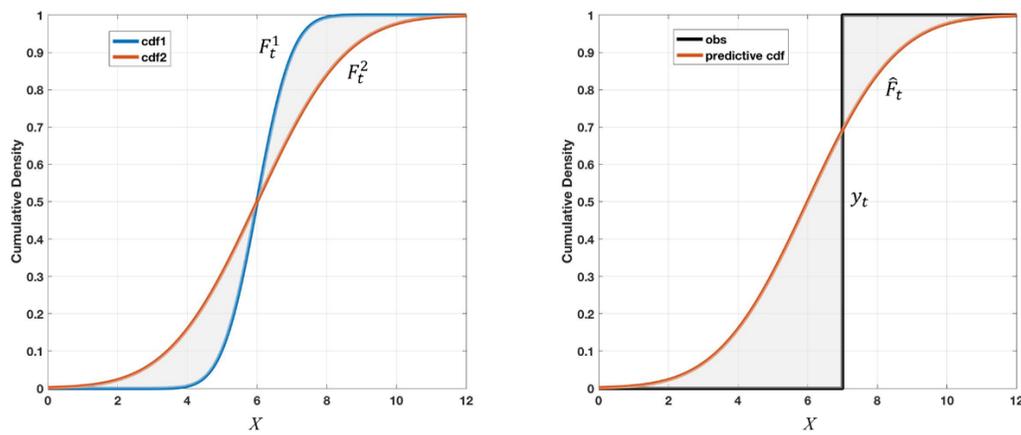
$$\overline{IS} = \frac{1}{M \cdot T} \sum_{t=1}^T \sum_{m=1}^M IS(\hat{L}_t, \hat{U}_t, y_t, \alpha_m) \quad (2.9)$$

(3) Continuous ranked probability score (CRPS)

Unlike quantile score and interval score, that rely on the evaluations of aggregate individual quantiles or intervals, the CRPS is able to assess the distributions as a whole [64]. CRPS is fundamentally related to rank probability score by comparing two full distributions [58, 64], which is defined as:

$$CRPS(F_t^1, F_t^2) = \int_{-\infty}^{\infty} (F_t^1(x) - F_t^2(x))^2 dx \quad (2.10)$$

where F_t^1 and F_t^2 are two distributions of variable X . This equation can be schematically interpreted by Fig. 2.3 (a), where the total area between two CDFs are the resultant CRPS value.



(a) Area between two distinct CDFs

(b) Area between predictive CDF and a crisp value

Fig. 2.3 Schematic of CRPS

When it comes to the probabilistic forecasts for a desired crisp value (e.g., y_t), Eq. (2.10) can be transformed into a special form, by replacing one CDF with a single step-function (step from 0 to 1) at the observed value, resulting in:

$$CRPS(\hat{F}_t, y_t) = \int_{-\infty}^{\infty} \left(\hat{F}_t(x) - \mathbf{1}(x - y_t) \right)^2 dx \quad (2.11)$$

where $\mathbf{1}$ is the Heaviside step function, which takes the value 0 when $x < y_t$ and the value 1 otherwise. This calculation process is illustrated in Fig. 2.3 (b), where the predictive CDF is denoted by the red line, and the observation is represented by a step-function in black at y_t , CRPS is thereby calculated by the area filled with grey between them. Apparently, a smaller area (lower CRPS value) indicates a better predictive skill. The average CRPS can thus be calculated by examining all calibrated points as:

$$\overline{CRPS} = \frac{1}{T} \sum_{t=1}^T CRPS(\hat{F}_t, y_t) \quad (2.12)$$

CRPS is more appealing than other skill scores mostly on its generalized definition, that is, it does not require the introduction of a number predefined classes (e.g., intervals in the interval score) on which results may depend [46]. It is applicable to all existing probabilistic forecasts.

In the context of parametric forecasts, the CRPS can be written as analytic forms corresponding to different pre-defined distributions [64, 65]. Due to the simplicity of implementation, in such cases, the CRPS is always adopted as the optimum score to estimate the regression parameters. However, it is hampered in the nonparametric situation, where the forecasts are represented by a set of discrete values or quantiles, posing great numerical challenges. Literature [66] provides useful guidance on calculating the CRPS discretely. The CRPS is decomposed into absolute differences (first component) and spread (second

component; which measures the lack of sharpness), as expressed in the following:

$$CRPS(\hat{F}_t, y_t) = \mathbb{E}_{\hat{F}_t} |X - y_t| - \frac{1}{2} \mathbb{E}_{\hat{F}_t} |X - X'| \quad (2.13)$$

where X and X' are independent copies sampled from the predictive distribution \hat{F}_t . Moreover, the CRPS can be derived through examining a finite number of individual quantiles (e.g., $\tau = 0.01, \dots, 0.99$) on the basis of pinball loss function (2.6) [65, 67]:

$$\begin{aligned} CRPS(\hat{F}_t, y_t) \\ = 2 \int_0^{\hat{F}_t(y_t)} \tau (y_t - \hat{q}_t^{(\tau)}) d\tau - 2 \int_{\hat{F}_t(y_t)}^1 (1 - \tau) (y_t - \hat{q}_t^{(\tau)}) d\tau \end{aligned} \quad (2.14)$$

we can further write its compact form as:

$$CRPS(\hat{F}_t, y_t) = 2 \int_0^1 (\mathbf{1}(y_t < \hat{q}_t^{(\tau)}) - \tau) (\hat{q}_t^{(\tau)} - y_t) d\tau \quad (2.15)$$

In this manner, without any smoothing acts with respect to the estimated quantiles, the CRPS can be properly calculated in the case that numerous quantiles are provided.

(4) Multivariate verification

The above discussed metrics are all limited to univariate, when the probabilistic forecasts are generated in the form of temporal or spatial trajectories, a proper multivariate verification is needed.

Energy score (ES) a multivariate generalization of CRPS by extending the univariate in (2.13) to a multivariate vector. For a given set of estimated scenarios $\mathbf{s}_t^{(j)} \in \mathbb{R}^{1 \times H}$ with dimension H , issued at time t , ES is expressed as

$$\text{ES}_t = \frac{1}{J} \sum_{j=1}^J \|\mathbf{y}_t - \mathbf{s}_t^{(j)}\|_2 - \frac{1}{2J^2} \sum_{i=1}^J \sum_{j=1}^J \|\mathbf{s}_t^{(i)} - \mathbf{s}_t^{(j)}\|_2 \quad (2.16)$$

where J is the total number of scenarios, $\|\cdot\|_2$ is the multi-dimensional l^2 norm and $\mathbf{y}_t \in \mathbb{R}^{1 \times H}$ is the observed vector. ES is effective in detecting the erroneous linear trend corresponding to the forecasts simulated. However, it is always limited by the discrimination issues, that is, it is insufficient to identify the incorrectly specified correlations between the components of the multivariate quantity [68, 69].

Variogram score (VS) is a powerful verification tool to address the above-mentioned problem by considering the pairwise differences of the components of the multivariate quantity, its l -th order expression is given by

$$\text{VS}_t^l = \sum_{m=1}^H \sum_{n=1}^H w_{m,n} \left(|y_{t,m} - y_{t,n}|^l - \frac{1}{J} \sum_{j=1}^J |s_{t,m}^{(j)} - s_{t,n}^{(j)}|^l \right) \quad (2.17)$$

where $w_{m,n}$ are nonnegative weights, indicating the significance of all pairwise differences of observations and forecasts. $y_{t,m}$ and $y_{t,n}$ are the m -th and n -th component of observed vector \mathbf{y}_t , respectively. $s_{t,m}^{(j)}$ and $s_{t,n}^{(j)}$ are j -th realizations of components $\mathbf{s}_{t,m} \in \mathbb{R}^{J \times 1}$ and $\mathbf{s}_{t,n} \in \mathbb{R}^{J \times 1}$, respectively. This score measures the dissimilarity between approximations of the variograms of order l of observations and forecasts over all pairs of components of the quantity [69]. Both ES and VS are negative-oriented.

Note that in the context of multivariate verification, there is no single score rule in existence that can serve all purposes [69], therefore, utilization of

different scores based on different desired properties are strongly advocated before making a justified conclusion.

2.3 A Review on State-of-the-art Probabilistic Forecasts for VRE Generations and Electricity Prices

In contrast to the bulk of research available on probabilistic forecasting of wind power [35] and its expansive applications in electricity market nowadays, one can hardly find the practices of PV power probabilistic forecasting. Generally, despite that the PV power has its unique characteristics, such as higher ramp rates due to lack of inertia and movements of clouds, distributed location or small-scale (behind-the-meter) generations [70], the forecast of solar PV energy can basically benefit from the developments in wind power due to the inherent weather-dependent nature they both share.

The probabilistic forecasting of electricity prices has become of particular interests until last few years, yet it is an undeveloped topic. Other than the wind power forecasting, where its close relationship to meteorological predictions are well established and commonly accepted [46], the influential factors of electricity prices are even more complicated to identify. The potential drivers of the price volatility are reported to comprise, but not limited to: demand, transmission congestion, generation outage, market participants behaviors, etc. These factors, and the uncertainties associated with them, are hard to incorporate into the price forecasting model [46].

Therefore, in this section, an overview of the current research activities is presented chiefly with regard to probabilistic forecast for wind power and electricity price. The approaches can be broadly classified into two categories: parametric and nonparametric.

2.3.1 Parametric methods

In parametric models, the shape of predictive distribution needs to be pre-defined, which is usually described as the analytic form with respect to the distribution parameters (e.g. location, scale). For example, the predictive distribution of wind power forecasting errors is often assumed as Gaussian distribution [71], which is characterized by only two statistical parameters: mean and standard deviation. This makes such parametric models fairly simple and computational efficient. Apart from Gaussian, Beta distribution is another widely-used statistical inference for wind power forecasting errors, mostly owing to its fat-tail property that accord with the large kurtosis of wind power forecasting error distribution [72]. In addition, the authors in [73] argued that wind power output should be viewed as the double-bounded variable, instead of unlimited variable following Gaussian distribution, which makes the Beta distribution (ranging from 0 to 1) more appropriate to describe the stochastic process of wind power [35].

Once the predictive distribution shape is determined, one needs to identify a regression model to establish a functional link with the underlying statistical parameters, e.g., location parameter and scale parameter. In most cases, the location parameter is considered as the point forecast, whereas the scale parameter implies the uncertainty.

2.3.1.1 Statistical time series models with parametric noise assumption

The most well-known models used to estimate the scale parameters must be statistical time series models (e.g. ARMA, ARIMA, GARCH), as they are

the basic and standard models that take into account the random nature and underlying time correlations of phenomenon [74, 75]. Such approaches focus on the error terms, that are frequently assumed to be independent and identically distributed (iid) noises with zero means and finite variances (i.e. Gaussian white noise) $\mathcal{N}(0, \sigma_\varepsilon)$.

In ARMA(p, q) model, the current value of the process is expressed linearly in terms of its p past values (autoregressive part) and in terms of q previous values of the residuals (moving average part) [74]. For $q=0$, it yields the most classic autoregressive AR(p) model. The ARMA modeling approach assumes that the time series under study is stationary. If it is not, then a transformation of the series to the stationary form has to be done first. In particular, this transformation can be performed by differencing. The resulting model known as the autoregressive integrated moving average (ARIMA) model. The ARMA-based model can be further extended to so called autoregressive moving average with exogenous variables (ARMAX) model, in which, the exogenous variables (e.g. NWP information) are incorporated to make the modeling process more rational.

However, all the linear ARMA-based models are characterized by homoscedastic nature, i.e. a constant variance and covariance function. From a practical point of view, wind/PV power is non-linear and non-stationary process, thus the heteroskedasticity of residuals must be taken into account. This is successfully addressed by the generalized autoregressive conditional heteroskedastic (GARCH(p, q)) model [75], where the variance is conditional

on the past values of the time series and a moving average of past conditional variances. By using this model, the predictive Gaussian distribution of residuals is no longer a constant one, but varies with time, making it more appealing to forecasters.

Over the past years, the majority of works using statistical time series models to generate the parametric predictive distribution or PIs concerns the wind power [76-79]. Literature [80] introduces a method to obtain 95% PIs for global horizontal irradiance (GHI) based on standard error as a function of solar zenith angle and clear sky index. Two assumptions are made in this work, standard errors follow normal distributions and prediction intervals are central on point forecasts.

Such parametric models, either with homoscedastic or heteroscedastic noise assumption, are popular in probabilistic forecasting of electricity price as well. Related works include [81-86].

2.3.1.2 Artificial intelligence models with bootstrap

Another prevailing technique used to yield parametric probabilistic forecasts is bootstrap [87], which is commonly combined with artificial intelligence (AI) models (e.g. NN, SVM). Bootstrap is a general approach of statistical inference introduced in [87], which aims to provide estimation of the sampling distribution of almost any statistic by uniform sampling with replacements from the original dataset [88]. Once the estimated distribution is obtained, PIs can be extracted by various means. In the process of regression, bootstrap has the merits on taking into account the model uncertainty arising

by the misspecification of the parameters, whereas it is limited by the high computational burden [46].

Khosravi et al. [89] build B multilayer feedforward NN models with different number of layers and neurons per layer to estimate the model uncertainty, and one additional NN to estimate the data noise, the resultant PI is assumed to be located at the mean value of B bootstrapping forecasts, with the total variance associated to the model outcome. Considering the intensive computational efforts are required by using traditional NN, Wan et al. [90] combine the ELM [91] with bootstrap to derive the PIs. In their work, B_M and B_N ELM models are used to estimate the variances of model uncertainty and data noise, respectively. Three bootstrap methods are examined, and they find pair bootstrap is more robust than the other two rivals. In addition, the overall skill of yielded PIs is assessed via interval score, which is a proper scoring rule in terms of PI evaluation. A hybrid model is developed in [92] to provide intra-hour PIs for one-minute averaged direct normal irradiance (DNI). The hybrid model comprises 4 NNs and 1 SVM. The proposed model first uses the SVM to classify the time series of DNI into two categories: low DNI variability period (lv) and high DNI variability period (hv). Then PIs are generated by ANN_{lv} or ANN_{hv} which are trained with data collected in lv and hv periods, respectively. Through comparing with the bootstrap-ANN model that consists of 201 NNs, this hybrid model is demonstrated to be more computationally efficient and can achieve high coverage probability during ramp events [92].

In parallel, bootstrap is also extensively applied to quantify the uncertainty involved in MCP forecasting. In [93], the ELM and a wild bootstrap are combined to produce the point forecasts and residual PIs of the Australian half-hourly MCP. The resultant PIs are the specific quantiles taken from the bootstrap sampling distribution. In a follow-up paper, Wan et al. [88] deem that the prediction uncertainty mainly consists of model uncertainty and data noise. By using a bunch of ELMs (with the number of B) to estimate the variance of model uncertainty via bootstrap, and a separate maximum-likelihood-estimation (MLE) NN to approximate the variance of residual noises, the overall prediction uncertainty can be obtained by adding these two independent variances up. The predictive distribution is assumed as normally distributed, with the mean of B bootstrapped replicates and the variance of overall prediction uncertainty. Therefore, rather than constructing the PIs nonparametrically from the empirical distribution in [93], the resultant PIs are centered around the point forecast symmetrically in [88]. In a more recent work, Rafiei et al. [94] propose a hybrid approach based on improved clonal selection algorithm, wavelet transform and ELM. The original series is firstly decomposed into one approximation and three details series by using wavelet technique. Then, each of them is fitted via NN, and thus the model uncertainty is computed via bootstrap. ELM is subsequently used to quickly train a single-layer NN to estimate the noise uncertainty. Finally, the PIs are established by considering both uncertainties in the same way as developed in [88]. Other

studies of MCP probabilistic forecasts where bootstrap method is used include [95-97].

2.3.2 Nonparametric methods

Although the observations at certain time or location follow a known and well-behaved marginal distribution, there is no guarantee that conditional predictive densities still follow the same distribution, incorrect distributional assumptions may directly cause biases in analyses and results. Without any efforts to make the statistical inference for the quantity of interest in advance, nonparametric approaches do not require the prior knowledge of the distribution shape.

2.3.2.1 Empirical simulation

Empirical simulation only concerns the past measurements or forecasting errors, no assumption is made before for the distribution family, which is essentially forecasting model-independent. The resultant distribution or PIs can be either in unconditional or conditional regime. Unconditional empirical simulation approach assumes the uncertainty is constant in certain location or period, the empirical distribution can be easily achieved as long as abundant past observations are available. Pinson et al. [98] proposes an adaptive resampling method to derive the conditional PI, which is able to inform the situation-dependent uncertainty. This method relies on a classification of recent of forecast errors, a fuzzy inference model and a multisample resampling scheme for combination of probability distribution. Empirical

simulation applied in probabilistic forecasting of electricity price include [26, 34-36].

2.3.2.2 Statistical moments estimation

The underlying rationale of such type of approaches is to estimate the first few moments of the random variable of interest. It is distinct from the traditional methods for time series in that it considers the quantity of interest as realizations of a variable whose distribution function parameters vary dynamically with time and other explanatory variables, whereas the traditional techniques model the expected pattern and the residuals [99].

Generalized Additive Models for Location, Scale and Shape (GAMLSS) is a powerful tool to achieve this goal, and is used to model the time-varying distribution of MCP by Serinaldi [100]. Through establishing the functional relationships between the distribution parameters (location, scale and shape) and the explanatory variables (e.g., past prices, loads, and weather information such as temperature) in a time-adaptive way, the price periodicities, trends and abrupt changes characterizing both the position parameter (linked to the expected value of prices), and the scale and shape parameters (related to price volatility, skewness, and kurtosis) can be explicitly informed by this model.

2.3.2.3 Kernel density estimation (KDE)

The rationale of KDE method is generating smooth histogram to estimate PDF of random variable. Given a finite data sample, a kernel density representing the contribution of each data point is placed at each data point. The estimated PDF can be obtained by adding up all kernel densities. KDE

technique could provide the whole predictive information via a smoothed PDF, instead of the discrete quantiles or PIs.

A multivariate kernel density estimator is employed to predict the PDF of wind power production for the next 60 hours in three wind farms in France [101]. By using the mutual information based feature selection, 4 features are identified from 16 NWP candidates as the inputs, and further modeled by a d -dimensional kernel density estimator. The bi-weight kernel is adopted for all variables in this case due to its low computational burden. Bessa et al. [102, 103] consider the feasibility of different kernel functions to different variables, four types of kernel function are proposed for wind energy, depending on whether the corresponding variables are bounded, unlimited or periodic.

Instead of directly establishing the multivariate relationships between in all variables, this nonparametric estimator can be alternatively used to model the error terms in time series models, particularly for electricity price approximating. In literature [82, 104], the authors employ the iterated Hsieh-Manski estimator (IHM) and smoothed nonparametric maximum likelihood estimator (SN) associated with the AR model to derive the day-ahead PIs for the electricity prices in two markets (Nord Pool and California Market). Comprehensive case study is carried out by examining 12 time series models, including the models driven by empirical noises and Gaussian noises. They conclude that the developed IHMAR and SNAR models yield better PIs than the other models, in terms of both unconditional and conditional coverage.

The major issue influencing the accuracy of density estimation is the bandwidth selection, which is a crucial parameter that controls the smoothness of the estimated PDF. Large bandwidth would lead to oversmooth, while small bandwidth would lead to undersmooth. The optimal bandwidth should be identified based on specific performance criteria. Asymptotic Mean Integrated Square Error (AMISE) is hired as the optimality index in [105], which is subject to minimization by the multivariate plug-in selector. Qin et al. [106] propose to minimize Integrated Square Error (ISE) of two kernel estimation. Then, optimal bandwidth estimation is converted to an unconstrained optimization problem.

2.3.2.4 Quantile regression (QR)

QR is a popular nonparametric probabilistic forecasting tool in the community of econometrics and social sciences, which is firstly introduced in [107]. QR aims at estimating the conditional quantiles of the response variable given certain explanatory variables. Based on the pinball loss function, the QR model can be solved by linear programming algorithms and many variations have been proposed in the past years [35, 46]. Over the last two decades, this classic approach has also attracted much attention in the forecasting activities concerning electricity market.

Local QR model is applied to [31, 108], where the dependence of predictive quantile on explanatory variables is modeled by a linear regression in the neighborhood of explanatory variable. In [109], the predictive quantiles are modeled by cubic B-Spline functions. Quantile regression forests (QRF)

models are proposed in [110], which is essentially an extension of regression forecast based on classifications and regression trees. In [111], multiple quantiles are estimated by the gradient boosted machines (GBM), and in [112], the multiple quantile regression is solved by the alternating direction method of multipliers (ADMM) algorithm. In recent studies, Wan et. al [61] develops a direct quantile regression (DQR) model for wind power in Bornholm Island in Denmark. By utilizing the unique merits of ELM, the complicated artificial NN-based nonparametric probabilistic forecasting is formulated as a linear programming (LP) problem with high computational efficiency. Hatalis et. al [113] performs a smooth approximation of the pinball loss function, thus the problem can be solved by the gradient based back-propagation NN. Moreover, they develop a novel weight initialization scheme to ensure multiple quantiles can be estimated simultaneously without overlapping.

The earliest work on resorting quantile regression to generate the probabilistic forecasts of electricity price can be traced back to year 2014 by Nowotarski and Weron [114, 115]. Their methodology involves applying quantile regression with the point forecasts of a number of individual forecasting models. In other words, the individual point forecasts are used as independent regressors and the corresponding observed target variable as the dependent variable in a standard quantile regression setting. Such methodology containing a batch of individual forecasts is commonly interpreted as ‘quantile regression averaging (QRA)’ [114]. The choice of the number of forecasting

members can be made arbitrarily [46] or, in case of dozens of competing models, using dimension reduction techniques [116].

The regressors can be further expanded to include useful explanatory variables, such as hourly, mean daily and ratios of load forecasts, average daily price forecasts and their squares [117]. Such models has helped two teams achieve spectacular success in the GEFCom2014 [117, 118].

Recently, Jonsson et al.[8] construct a time-adaptive quantile regression model to estimate the 5-95% quantiles for the residuals separately, besides, to obtain robust predictive distribution, they treat the distribution tails with an exponential assumption. Therefore, strictly speaking, this model belongs to semiparametric framework.

2.3.2.5 PI-score-based direct estimation

Lately, the PI-score-based prediction models has gained great popularity in the MCP probabilistic forecasting research. Khosravi et al. [43] proposes a new training method for NN to directly generate two outputs (lower and upper bounds) based on a hybrid PI-based cost function, Coverage-width-based Criterion (CWC), which covers two important aspects of PIs: informativeness (width) and correctness (coverage probability). Through verifying with 10 datasets and comparison with three traditional PI construction method, the proposed method proves to be simpler, faster and more reliable. In a follow-up paper [119], they applied the same training method to GARCH model that directly deals with the quality of PIs, rather than forecasting error. The tuning of GARCH model parameters is driven by minimization of CWC, which

contrasts with the traditional MLE-based methods for GARCH training. Once tuned, the GARCH model is used for construction of PIs around forecasts obtained by the moving block bootstrapped NN. Performance of the proposed method is evaluated using datasets in two Australian and New York City deregulated markets.

However, several researchers argue that CWC is not a proper scoring rule [32, 120], resulting in the illness of resultant PIs. Therefore, the score-oriented training scheme should be implemented on the condition that the proper scoring rule is hired beforehand.

To address this issue, Wan et. al [42] proposes the interval-score-based learning strategy for NN. Owing to the unique merit of ELM, i.e., the hidden neurons are randomly assigned and free to tune, only the output weights are subject to minimization with the interval score. As such, the computational efficiency can be highly improved as compared to the traditional methods. In [41], a modified version of CWC, called deviation information-based criterion (DIC), is proposed as the objective function. The original wind power series is decomposed into three components by the ensemble empirical mode decomposition (EEMD) method. Only the noise component is used to construct the optimal PIs based on ELM. The effectiveness of this hybrid model is demonstrated using the real wind power measurement provided by Australian wind farms.

2.4 Summary

The current use of probabilistic forecasts for VRE generation and electricity price in today's electricity market are briefly described with some typical cases in the beginning of this chapter. Four frequently-used representations of probabilistic prognosis are presented afterwards, with the proper verification tools that applicable to each of them. Despite that the goodness of a forecast can be legitimately determined by these metrics, Pinson [47] argues on the fact that a forecast being considered as better than the others greatly depends on the operational context. For instance, some real-time control activities require the corresponding forecast should be given within seconds. In such cases, the forecasting engine with high learning speed is preferred to that with high forecasting accuracy. Hence, the metrics interpreted in this chapter are more likely to be favored by forecasters as they provide a generic and objective way to evaluate the underlying forecast. But from a practical perspective, decision makers are more inclined to use the prognosis that are tailored to their needs. Lastly, an overview of the current research statues of probabilistic forecasts for VRE generation and electricity price is given. Note that none of the methods discussed in Section 2.3 is restricted to certain variable, once the explanatory variables for the variable is identified explicitly, these approaches can be applied interchangeably to different variables (e.g. wind power, PV power, electricity price, load) in the context of probabilistic forecasting.

Chapter 3 Granule-based Interval Forecast for Solar Irradiance

3.1 Introduction

Predicting solar irradiance is essential but fairly challenging to estimate solar power production. To effectively quantify uncertainties in solar irradiance or PV power forecasting, a great deal of NN-based models has been developed to obtain the PIs following the paradigm of wind power forecasting. Traditional NN-based PI estimating methods always require the point forecasts and the associated error information to construct the PIs, with a prior assumption of the error distribution [121-123]. However, through empirical investigations, the distribution of prediction errors can be rather complicated in the realistic cases, exhibiting different shapes at different locations and look-ahead horizons [70, 124].

Recent advance aims to directly construct the PIs nonparametrically to pursue the best quality of derived PIs, without performing point forecasts or prior knowledge of the associated errors. This idea is originally introduced by Khosravi et. al [43], who uses the coverage width-based criterion (CWC) as

the optimum score to tune the parameters of traditional NNs to directly yield the estimated lower bound and upper bound. The authors further apply the same framework to construct the PIs for load [125] and wind power [89, 126]. Yet, CWC is reported to be not able to measure the overall skill of constructed PIs, and would mislead the construction of optimal PIs as the CWC-based cost function gives biased weights to reliability and sharpness [32, 120]. This issue is successfully addressed by bringing in a proper sorting rule for PI assessment shortly thereafter, Wan et. al [42] develops a direct interval forecast (DIF) approach based on interval score, to account for both coverage probability and sharpness of PIs simultaneously.

Although the DIF approach can ensure optimal quality of PIs in one single optimization process, it should be highlighted that it only concerns on the stochastic uncertainty of non-stationary time series, while the knowledge uncertainty is not considered at all. Stochastic uncertainty and knowledge uncertainty are always regarded as the two main contributors to the prediction errors in AI-based models. The former describes the inherent variability of the observed values due to the natural physical phenomenon, measured error, device failure and the like, while the latter reveals the uncertainty in knowledge

transfer, such as imperfect representation of processes in a model, as well as the imperfect knowledge of the parameters associated with these processes. In addition, the PIs are constructed based on crisp or deterministic inputs without explicitly recognizing the variability involved in the observed dataset.

Lately, information granule (IG) has emerged as a new and powerful tool to deal with the situations characterized by manifold uncertainties [40, 127]. By granulating the underlying chaotic times series and model parameters with proper granularity level, a granular input-output mapping can be established, which can effectively reflect the potential attributes in the process of knowledge transfer. Therefore, in this chapter, a new framework based on the conceptualization of IGs is developed to directly generate the optimal PIs without extra efforts for point forecasting and error modeling. The granular model is built upon the existing crisp NN. ELM is an appealing option due to its fast learning speed, free-tuning of hidden neurons, and excellent approximation capabilities. The granular parameters are tuned through particle swarm optimization (PSO) with a cost function combining the reliability and interval score. Case study is carried out by using the 1-min GHI observations measured by Hong Kong King's Park Meteorological Station. Comparative

studies show the superiority of the proposed GELM model in producing the skillful PIs accounting for both reliability and sharpness. The effect of different time scales and granulation methods are also examined.

3.2 Granular Computing

Granular computing (GrC) is an emerging computing discipline of information processing [40, 128], which is more a theoretical perspective than a coherent set of methods or principles. It encompasses all the methods that concerned with the data abstraction and derivation of knowledge from information or data, where IG arises in a generic representation formalism. In the community of machine learning or data mining, IG is often used to deal with the situations characterized by excess or a lack of data [129]. The first situation occurs when there are collections of entities that exhibit some similarity in terms of their properties or functional appearance. In this case, IG provides a vehicle to abstract the complexity of the data set that one can organize into hierarchies and converts the original problem into manageable subtasks. The second situation occurs, for instance, when noisy data exists.

Here, IG allows modeling the precision of indirect measurements, providing a computationally appealing view of knowledge [130].

IG can be formalized and described in a variety of representations, such as, intervals, fuzzy sets, rough sets, shadow sets, probabilities and so forth [131]. The selection of representation type should be treated with caution, depending on the issue we encounter and the available knowledge we obtain.

In light of this, in the AI-based models where uncertainty, inaccuracy and variability should be taken into account, generalizing (abstracting) the existing model by forming its granular counterpart is a straightforward and appealing way to tackle with these manifold uncertainties. In the granular model, parameters are regarded as IGs with certain type rather than numeric entities [132], correspondingly, the output is also in a granular form, comprising all the potential attributes resulting from the uncertainty of data and models.

3.3 Experimental Data and Empirical Investigation

The GHI time series with 1-min resolution recorded in Hong Kong King's Park Meteorological Station, from 1st Jan 2012 to 31st Dec 2012, are collected for case study in this thesis. GHI is the total amount of solar irradiation incident

on a horizontal surface, measured by Wm^{-2} . The other two crucial components closely related to this quantity is direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI). The value of GHI is of particular interest to the forecasting of PV power output, a series of equations have been developed to convert the GHI parameter into PV power empirically [10, 133]. As studied in [133], the linear correlation between the observed GHI and PV production can be as high as 0.97.

Fig. 3.1 is an illustration of the GHI series for a typical sunny day at King's Park Meteorological Station, high variability and intermittence is observed before noon of the day, whereas a smoother profile is witnessed in the afternoon. The main factor for the fluctuations of GHI in the morning is the cloud motion, such effects can be amplified by the high resolution of underlying time series.

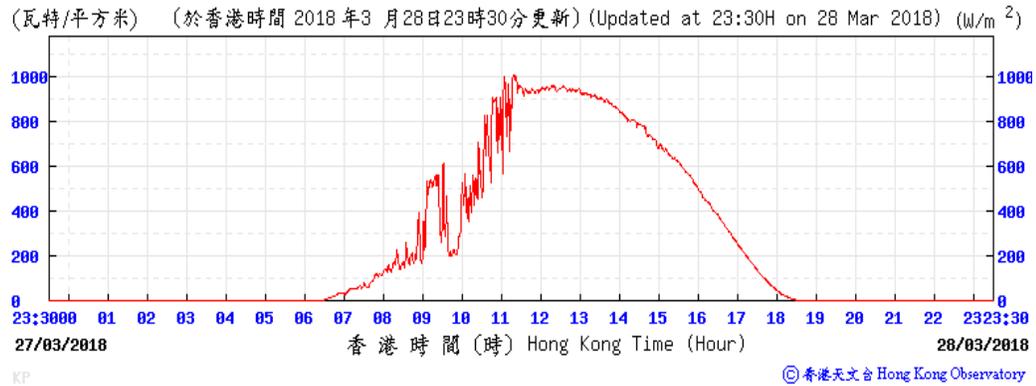


Fig. 3.1 Time series of 1-min GHI, measured at King's Park Meteorological Station on 28th Mar 2018 in Hong Kong. A snapshot downloaded from HKO website [134]

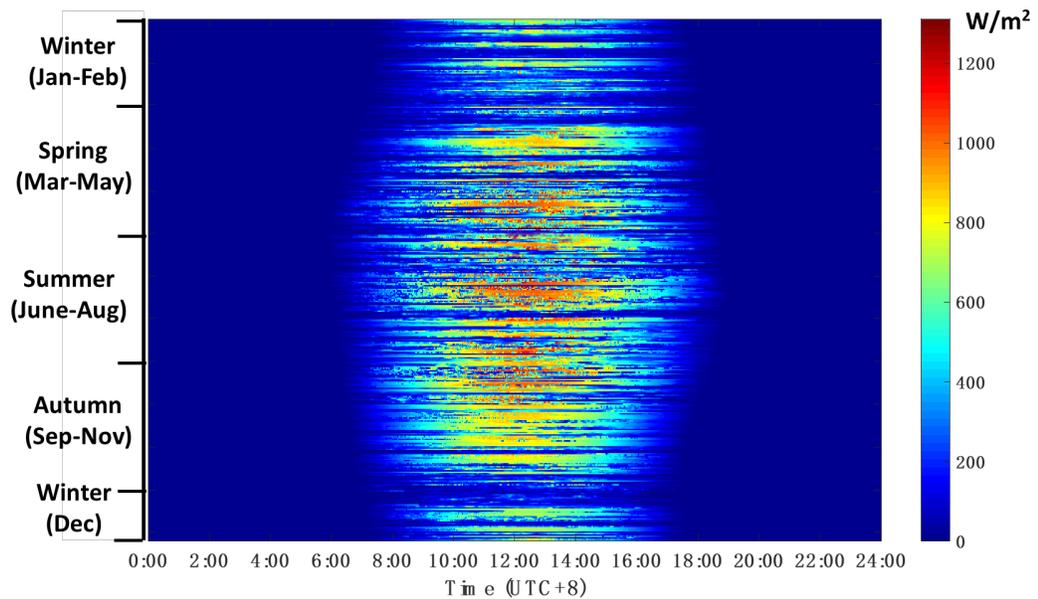


Fig. 3.2 2D plot of GHI throughout the year of 2012

To have a broad view of its seasonality, the GHI throughout the year of 2012 is depicted in Fig. 3.2 in a 2D view, from which we can clearly see the variations of daily sunshine duration and the amount of GHI during the whole year. The amount of solar radiation is more abundant between May and

September than that in the rest months, it peaks approximately in July and hits the bottom in Winter. To have a consistent prediction duration for all days, the nighttime data is removed, only the data from 7:00 to 18:00 is used for the following experiments.

3.4 Extreme Learning Machine (ELM)

ELM has received great popularity lately [84, 88, 93], since iterative parameters' tuning in traditional gradient-based NN are avoided in the learning process of ELM [91]. In essence, this novel learning algorithm is proposed [91] to train single hidden-layer feedforward neural networks (SLFNs), where the input weights and biases of hidden nodes are randomly assigned and free to be tuned further. The output weights of SLFNs are analytically determined by a direct matrix calculation. According the experimental results reported in [91], ELM achieves better generalization performance with extremely fast learning speed. As such, ELM is utilized as the basis forecasting engine to efficiently generate granular outputs in this thesis, the detailed algorithm is introduced as follows.

Given N arbitrary distinct samples $\{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}^m\}_{i=1}^N$, where \mathbf{x}_i denotes the input vector and \mathbf{t}_i denotes the target vector. ELM with a specific activation function $g(\cdot)$ and randomly assigned input weights and biases can efficiently approximate all samples with zero error:

$$f(\mathbf{x}_j; \mathbf{w}, b, \boldsymbol{\beta}) = \sum_{i=1}^L \boldsymbol{\beta}_i g(\mathbf{w}_i^T \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j=1, 2, \dots, N \quad (3.1)$$

$$\sum_{i=1}^L \boldsymbol{\beta}_i g(\mathbf{w}_i^T \mathbf{x}_j + b_i) = \mathbf{o}_j, \quad j=1, 2, \dots, N$$

$$\sum_{j=1}^N \|\mathbf{o}_j - \mathbf{t}_j\| = 0$$

where $\mathbf{w}_i \in \mathbf{R}^n$ is the input weight vector associated with the i -th hidden node and all input nodes; $\boldsymbol{\beta}_i \in \mathbf{R}^m$ is the output weight vector connecting the i -th hidden node and all output nodes; b_i is the threshold of the i -th hidden node; and L is the number of hidden neurons. Eq (3.1) can be rewritten as the following matrix form:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (3.2)$$

where \mathbf{H} is expressed as:

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1^T \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_L^T \mathbf{x}_1 + b_L) \\ \vdots & \vdots & \vdots \\ g(\mathbf{w}_1^T \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_L^T \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (3.3)$$

$\boldsymbol{\beta}$ and \mathbf{T} are respectively expressed as:

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L \times m} \text{ and } \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \quad (3.4)$$

Since the input weights and biases are randomly assigned, training an ELM-based SLFN is equivalent to obtaining the least square solution for output weights $\boldsymbol{\beta}$ of the linear system in (3.1), which can be expressed as:

$$\begin{aligned} & \left\| \mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_L, b_1, \dots, b_L) \hat{\boldsymbol{\beta}} - \mathbf{T} \right\| \\ & = \min_{\boldsymbol{\beta}} \left\| \mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_L, b_1, \dots, b_L) \boldsymbol{\beta} - \mathbf{T} \right\| \end{aligned} \quad (3.5)$$

It can be proven that the smallest norm least-squares solution of the above linear formulation can be achieved by:

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^\dagger \mathbf{T} \quad (3.6)$$

where \mathbf{H}^\dagger is the Moore-Penrose generalized inverse of matrix \mathbf{H} [135] and it is generally derived by singular value decomposition.

ELM effectively overcomes the limitations of traditional gradient-based NNs, such as the local minima, overfitting and the high computational cost. For any infinitely differentiable activation function, the ELM with N hidden layer neurons can learn N distinct samples exactly with zero error. In addition, ELM training can obtain the best results according to the assigned input weights. The training speed is extremely fast due to the simple matrix operation

in (3.6). ELM also distinguishes from traditional NNs in superior generalization capability without the overtraining issue [136, 137].

3.5 Uncertainty Analysis

Prediction plays a crucial role in various decision-making processes nowadays, more than in the realm of power system. For this reason, during the last decade, a large family of works [138-140] have been carried out to find out the fundamental sources that might contribute to the prediction uncertainty. Generally, prediction uncertainty can arise due to the stochastic process of time series, measurement errors, the misspecification of model formulations and alike. These factors can be broadly classified into two distinct types: stochastic uncertainty and knowledge uncertainty [138].

3.5.1 Stochastic uncertainty

Stochastic uncertainty describes the inherent variability of the observed values due to the natural physical phenomenon, measured error, device failure and the like [62]. Hence, it is also named as natural uncertainty, inherent certainty, or aleatory uncertainty in other existing literatures [138, 141, 142].

Natural variability is a property of the natural system, reflecting the stochastic

process of quantity in meteorological forces, thus it is always deemed as the leading source of stochastic uncertainty, that cannot be eliminated. On the other hand, humans can easily improve the measurement accuracy and device reliability in reality, as a result, the uncertainty arising from inaccurate measurement, mistake recording and missing data is reducible. In this sense, to diminish the stochastic uncertainty, great efforts must be spent to make a better knowledge of the stochastic behavior of the underlying quantity.

For GHI series, the uncertainty is heavily influenced by the local cloud states, cloud passages can cause a sudden shade on the measurement site, thus leading to a dramatic drop of GHI within few seconds. To tackle with this issue, a wide range of practices have developed to use satellites or ground-based sky imagers to predict the cloud motions [10, 80, 143]. However, it is still challengeable to model the variability of GHI, since the cloud state is also a complex modeling process, wherein the cloud cover, cloud thickness and duration are hard to estimate.

3.5.2 Knowledge uncertainty

Another possible source of uncertainty in model output results is the imperfect model representation of the system behavior, in terms of the

imprecise representation the model processes (structural uncertainty) and imperfect knowledge of the values of parameters associated with these processes (parameter uncertainty) [138, 142]. Such uncertainties related to the system models are also referred to as model uncertainty or epistemic uncertainty [90, 142].

Knowledge uncertainty exists in the process of knowledge transfer, but it is not always due to a lack of knowledge. Given a model that are calibrated repeatedly by different datasets, different model parameter values would result correspondingly. Those values would further yield different simulated system behavior and, thus, different predictions. If such imprecise specification of parameters were eliminated, then the predictions would be consistent and the parameter uncertainty in the forecasting results would be zero [138]. But, it does not mean this model would be perfectly accurate. Additionally, the misspecification of model structure (e.g., number of layers and number of neurons in each layer in a NN) can also give rise to the variability in model output.

In recognition of the main sources of prediction uncertainties, the key task translates to identify a proper representing approach that can effectively

propagating these uncertainties through computational models. Traditionally, the exercise of uncertainty quantification is conducted mostly based on simulation ways, e.g., bootstrap. Nonetheless, they are always limited by the considerable computational cost, this motivates us to exploit a more straightforward and efficient technique to represent the uncertainty. IG is a powerful tool to bounding solutions under uncertainty, by constructing a granular counterpart to the original crisp model, where the stochastic uncertainty and knowledge uncertainty are well captured by certain type of IGs, the potential outcomes resulting from these uncertainties can be revealed efficiently by the derived granular outputs.

3.6 Formulation of Granular Prediction Model on the basis of

ELM

The proposed granular prediction model aims to employ a proper type of representations to deal with the uncertainty in the system inputs and parameters, thereby propagating it through the model and revealing the resulting uncertainty in the granular output. By dressing with designated criteria, the output granule leads to an equivalent description of PI.

3.6.1 Granular GHI time series

The interpretability of time series is of significant interest in data mining and is still an ongoing challenge, particularly in the era of big data, where the data is characterized by large size, high dimensionality and a stream-like nature. It thus becomes an ever-visible trend to design user-centric models of data, considering that humans are more inclined to perceive and organize knowledge at a higher level of abstraction than the one being supported by numeric models [18].

The abstraction of detailed numeric data can be realized with the aid of IG, through a family of representative forms, including intervals, fuzzy sets, rough sets, probabilities, etc. [131, 144]. These granular representations are commonly built from two points of views: time and space. With regard to temporal granulation, time series is split into finite temporal windows leading in this way to the formation of temporal granules, wherein the key features are perceived. Alternatively, we can construct granules over the space of amplitude by using some clustering techniques, this process is often linked to feature dimension reduction in the modeling systems plagued with sizable features. Given that our experiment only involves univariate – the historical GHI

measurements, while exogenous important variables (e.g., NWP information and cloud motion) are not available, only temporal granules can be constructed over the raw GHI time series. More specifically, the granule is expressed in the form of interval due to its compact form, low computational burden, and coherence with PI. In the remainder of this thesis, the term granule and interval are used interchangeably.

To fully describe the variability in each temporal window, the upper and lower bound of an interval must be selected with caution. Once the interval is determined, all elements included in this temporal segment lose their identity in the sense they become fully indistinguishable. Fig. 3.3 shows the construction of interval-valued time series on a 2-day crisp (single-valued) GHI series.

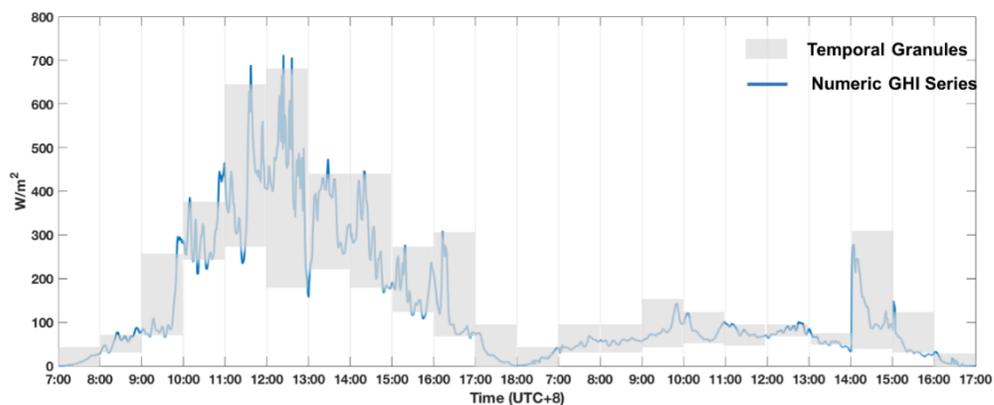


Fig. 3.3 From numeric time series to granular time series

Here, two approaches are offered to derive the temporal intervals with different granularities, which we refer to as min-max-Gr and FCM-Gr.

Min-max-Gr builds the lower and upper bound of interval by simply taking the minimum and maximum value within each temporal segment.

Suppose \mathbf{x}_j is the vector of k observations contained in j -th time window, the

associated interval can be formulated by: $\mathbf{x}_j = \{x_{1,j}, x_{2,j}, \dots, x_{k,j}\} \rightarrow$

$$G_j^{min-max} = [\min(\mathbf{x}_j), \max(\mathbf{x}_j)].$$

FCM-Gr approach takes advantages of fuzzy sets theory that each data point could belong to two or more clusters with different degrees of membership measured in $[0,1]$, thus giving the flexibility to represent the membership relationship of each data. Further, the shape of membership function is not required to be pre-assumed in this method; instead, it depends on the clusters' centers to establish the lower and upper bounds [145]. The algorithm is presented as follows.

Given a vector of numeric measurements $\mathbf{x}_j = \{x_{1,j}, x_{2,j}, \dots, x_{k,j}\}$ within j -th time window, FCM attempts to partition it into d clusters $\mathbf{c}_j = \{c_{1,j}, c_{2,j}, \dots, c_{d,j}\}$. In this process, a partition matrix, $\mathbf{U}_j = u_{mn}^{(j)} \in [0,1], m =$

$1, 2, \dots, k, n = 1, 2, \dots, d$, is obtained, indicating the degree to which element, x_{mj} , belongs to each cluster c_{nj} . Through minimizing an objective function:

$$\arg \min_{c_j} \sum_{m=1}^k \sum_{n=1}^d (u_{mn}^{(j)})^l \|x_{m,j} - c_{n,j}\|^2 \quad (3.7)$$

$$\text{where } u_{mn}^{(j)} = 1 / \sum_{t=1}^d \left(\|x_{m,j} - c_{n,j}\| / \|x_{m,j} - c_{t,j}\| \right)^{\frac{2}{l-1}}.$$

The n -th cluster centroid thus can be expressed as:

$$c_{n,j} = \frac{\sum_{m=1}^k (u_{mn}^{(j)})^l x_{m,j}}{\sum_{m=1}^k (u_{mn}^{(j)})^l}, \text{ where } l \text{ is the fuzziness coefficient. Two}$$

cluster centers are required to describe the variability of each time segment,

i.e., $d=2$. Smaller centroid is taken as the lower bound and larger centroid is

for the upper bound. As a result, the granule derived by FCM-Gr for j -th time

window is given by $G_j^{FCM} = [\min(c_j), \max(c_j)]$.

3.6.2 Granular ELM-based NN

The SLFN based on ELM algorithm is selected as the basis model due to its superiorities described in Section 3.4. To accommodate the granular inputs, the crisp ELM model is generalized by forming its granular counterpart – granular ELM, wherein the parameters are regarded as granules rather than numeric entities. The architecture of GELM is illustrated in Fig. 3.4, where the crisp weights and bias are augmented in an interval-valued form. By doing this,

we form a granular mapping to provide a granular output, which in light of a proper PI skill becomes more in rapport with the granular inputs.

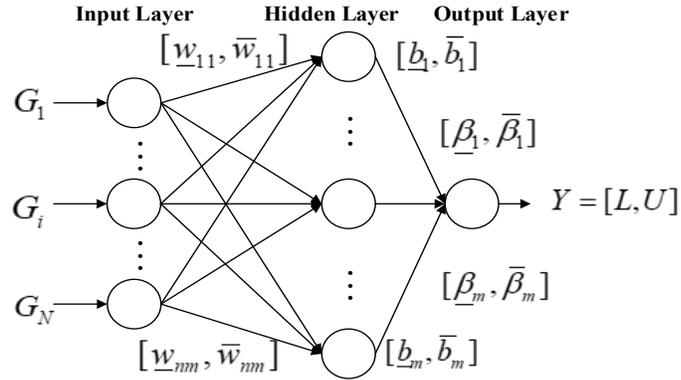


Fig. 3.4 Architecture of GELM

3.6.3 PI-skill-oriented training strategy

3.6.3.1 Fitness function

As the original problem is expected to generate PIs, we resort to intervals score as the optimized perform index to pursue optimal reliability and sharpness of the yielded granular output simultaneously. It has to be stressed that the interval score is able to account for reliability and sharpness, it cannot quantitatively distinguish the contributions of the two aspects [146]. Given that the reliability of PIs is the primary metric in probabilistic forecasting, we highlight the reliability respect by combining it with the interval score to form fitness function. It worth noting that the input weights and bias of GELM are randomly generated, only the granular output weights serve as design variables.

Recall the discussion of proper scoring rules of PIs in Section 2.2.3, the reliability of PI is expressed by ACE, the overall fitness function can thus be written as:

$$\begin{aligned} \mathbf{\beta}_{opt} = \underset{\mathbf{\beta}}{\operatorname{argmin}} & \left(\left| \operatorname{ACE}(\operatorname{PICP}, \alpha) \right| + \sum_{i=1}^{N_t} \left| IS_i(\hat{L}_i, \hat{U}_i, y_i, \alpha) \right| \right) \\ \text{s.t. } & \hat{L}_i \leq \hat{U}_i, 0.1 \leq \hat{L}_i, \hat{U}_i \leq 0.9 \end{aligned} \quad (3.8)$$

where N_t is the size of training samples and $\mathbf{\beta} = [\underline{\beta}_1, \bar{\beta}_1, \dots, \underline{\beta}_m, \bar{\beta}_m]$ is the vector of granular output weights. The derived lower bound and upper bound are restricted within the range of GHI, i.e., the normalized minimum and maximum value of GHI, respectively.

3.6.3.2 Particle swarm optimization (PSO)

The objective function in (3.8) is non-differentiable with respect to the granular output weights, thus requiring a computational optimization tool with high efficiency. PSO is a classic heuristic optimization algorithm, which has been proved to be an efficient, robust and gradient-free optimization tool [147] in numerous science and engineering applications. The major inspiration behind PSO is the flocking behavior of birds. It solves the problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical

formulae over the particle's position and velocity. Besides, the fast convergence speed differs PSO from other heuristic optimization methods [146], hence it is applied to tune the granular output weights to achieve the optimal quality of granular outputs. The algorithm is briefly presented in the following.

Given a group of random particles (solutions) with size N_p , in a S -dimensional search space, the i -th particle can be expressed by $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iS}]$. In the iterative process of PSO, each particle is updated by following two best values. The first is called global optimal solution, i.e., the particle with the best fitness value in the population, which is denoted by \mathbf{p}_g^{best} . The other is the best position of any particles achieved in the previous search, which is stored in a vector and expressed as $\mathbf{p}_i^{best} = [p_{i1}^{best}, p_{i2}^{best}, \dots, p_{iS}^{best}]$, the associated velocity of i -th particle is given as $\mathbf{v}_i = [v_{i1}, v_{i2}, \dots, v_{iS}]$. In every iteration, the particle updates its velocity and position with the following equations:

$$\mathbf{v}_i = w\mathbf{v}_i + c_1R_1(\mathbf{p}_i^{best} - \mathbf{x}_i) + c_2R_2(\mathbf{p}_g^{best} - \mathbf{x}_i) \quad (3.9)$$

$$\mathbf{x}_i = \mathbf{x}_i + \phi\mathbf{v}_i \quad (3.10)$$

where w is the inertia weight; ϕ is a constriction factor controlling and keeping the velocity within the range $[-v_{max}, +v_{max}]$; c_1 and c_2 are learning factors; R_1 and R_2 are random numbers within $[0,1]$. As seen from (3.9) and (3.10), the velocity of the i -th particle is related to three components: the previous velocity of particle, the distance between the previous best position of the particle and its current position, and the distance between the swarm's best success and the particle's current location.

3.6.3.3 Granular parameters tuning via PSO

The underlying optimization procedure is described as follows:

Step 1) The original numeric GHI dataset is normalized into $[0.1,0.9]$, then based on the granulation methods introduced in Section 3.6.1, the corresponding granular time series is constructed with a certain level of granularity (e.g., 10min).

Step 2) To initialize the GELM, a set of granular training samples are formulated as $\{(\mathbf{G}_{t,i}, G_{t+k|t,i})\}_{i=1}^{N_T}$, where $\mathbf{G}_{t,i}$ is the granular input vector consisting of the lagged observed granules available at time t and $G_{t+k|t,i}$ is the target granule for look-ahead time $t+k$. Note that, in this stage, the initial parameters of ELM can be determined by interval regression approaches [148-

150], with the goal of minimizing the interval error by interval arithmetic [40] and genetic algorithm. This is more compliant with the nature of intervals, as the interval-valued NN is originally developed to estimate the unconditional intervals, irrespective of certain crisp entity. However, it seems much computationally expensive to involve an extra optimization process in the initialization stage. To address this issue, we resort ourselves to the wealth of NN's own learning scheme. Specifically, two individual ELM networks are used to train the lower bound and upper bound series of granular samples, separately, then two sets of ELM parameters can be quickly derived to initialize GELM. The core idea behind this manipulation is that searching around the unconditional specified intervals can significantly facilitate the seeking process for optimum solutions.

Step 3) Initialize the population with N_p randomly generated particles around the initial output weights $\boldsymbol{\beta}_{ini}$ obtained in step 2) and velocities \mathbf{v}_{ini} . Owing to the random nature of ELM, i.e., the input weights and bias are randomly determined, the optimization phase is only concerned with the granular output weights $\boldsymbol{\beta} = [\underline{\beta}_1, \bar{\beta}_1, \dots, \underline{\beta}_m, \bar{\beta}_m]$.

. *Step 4)* For each positioned particle, according to (3.9), the corresponding PIs and fitness value can be computed.

Step 5) Do the iterative optimization procedure until the maximum number of iterations is reached or the convergence criterion is met. In each iteration, if the fitness value with the current position $\mathbf{p}_i, i = 1, \dots, N_p$ is better than that of \mathbf{p}_i^{best} , then set \mathbf{p}_i^{best} as the best location. In parallel, the global optimal particle is identified by comparing all current positions, which is stored in the updated vector \mathbf{p}_g^{best} . Then, the velocities are changed to move towards the next positions according to (3.9) and (3.10).

Step 6) Once the optimization phase is ended, the optimal GELM model is determined in regards of the fitness function (3.8) and used to construct PIs for out-of-the-samples.

3.7 Experimental Results

3.7.1 Benchmarking methods

3.7.1.1 Persistence method

As a simple and popular benchmark in deterministic point forecasts, persistence method [98, 151] is commonly used to give the estimation for the very short-term horizon (e.g. one-time step ahead). A probabilistic generalization of it is considered here to benchmark the proposed GELM model. The persistence predictive distribution $\hat{F}_{t+k|t}$ issued at time t for lead time $t+k$ is assumed to be normally distributed $\hat{F}_{t+k|t} \sim N(\hat{\mu}_{t+k|t}, \hat{\sigma}_{t+k|t}^2)$. The mean $\hat{\mu}_{t+k|t}$ is the latest available observation y_t , and the estimated variance $\hat{\sigma}_{t+k|t}^2$ is determined by the past samples over a certain period. Therefore, to obtain a PI with nominal proportion of $(1 - \alpha) \times 100\%$, the quantiles of $\alpha/2$ and $(1 - \alpha/2)$ are extracted from the predictive distribution as the symmetric lower and upper bound, respectively:

$$\begin{aligned}\hat{q}_{t+k|t}^{(\alpha/2)} &= \hat{F}_{t+k|t}^{-1}(\alpha/2 | y_t, \hat{\sigma}_{t+k|t}^2) \\ \hat{q}_{t+k|t}^{(1-\alpha/2)} &= \hat{F}_{t+k|t}^{-1}(1-\alpha/2 | y_t, \hat{\sigma}_{t+k|t}^2) \\ \hat{I}_{t+k|t}^{1-\alpha} &= [\hat{q}_{t+k|t}^{(\alpha/2)}, \hat{q}_{t+k|t}^{(1-\alpha/2)}]\end{aligned}\tag{3.11}$$

where $\hat{F}_{t+k|t}^{-1}$ is the inverse CDF; $\hat{q}_{t+k|t}^{(\tau)}$ is the estimated quantile with nominal level τ ; $\hat{I}_{t+k|t}^{1-\alpha}$ is the resultant PI centered at y_t with nominal proportion $(1 - \alpha) \times 100\%$.

3.7.1.2 Climatology method

Climatology is another widely-used benchmark in probabilistic forecasting, particularly for weather-related processes. It simply relies on the nonparametric fitting of the past observations, regardless of any current available information, hence it is essentially unconditional and unique. The distribution can be fitted empirically or using KDE [70] based on the data for the same time of all days in the same season. Here, the empirical distribution is adopted due to its simple fitting process, and the pair quantiles of $\alpha/2$ and $(1 - \alpha/2)$ can be easily computed via some interpolation techniques. Let us define F_{t+k}^{emp} is the empirical distribution for all past samples at the same look-ahead time $t+k$ for all days in the same season, the central PI can thus be derived as:

$$\begin{aligned} \hat{F}_{t+k|t}^{-1} &= F_{t+k}^{emp} \\ \hat{q}_{t+k|t}^{(\alpha/2)} &= \hat{F}_{t+k|t}^{-1}(\alpha/2), \quad \hat{q}_{t+k|t}^{(1-\alpha/2)} = \hat{F}_{t+k|t}^{-1}(1-\alpha/2) \\ \hat{I}_{t+k|t}^{1-\alpha} &= [\hat{q}_{t+k|t}^{(\alpha/2)}, \hat{q}_{t+k|t}^{(1-\alpha/2)}] \end{aligned} \quad (3.12)$$

3.7.1.3 B-Spline quantile regression (BS-QR)

The above two benchmarks are relatively simple and less challengeable.

To better demonstrate the effectiveness of the proposed model, a classical nonparametric model, BS-QR [109, 110], is introduced as a competitive rival in this thesis. Instead of use a simple linear combination of underlying regressors to estimate the desired quantile, BS-QR uses an additive model combining with some known basis functions (e.g., B-Spline basis). The general form of additive models can be written as:

$$y = a + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon \quad (3.13)$$

where y is the desired variable depending on x_1, \dots, x_p , a is the constant and ε represents the Gaussian white noise. As described in [152], each of the functions $f(\cdot)$ can be approximated by linear combinations of known basis functions of the corresponding explanatory variable, i.e.,

$$f_j(x_j) = \sum_{k=1}^{n_j} b_{jk}(x_j) \theta_{jk} \quad (3.14)$$

where $b_{jk}(\cdot)$ are the basis functions and θ_{jk} are unknown coefficients. As a consequence, substituting (3.14) into (3.13), we can obtain a linear regression

model. This can be further generalized to the quantile regression by modeling the τ -th quantile $q_t^{(\tau)}$ as:

$$q_t^{(\tau)} = a_t^{(\tau)} + \sum_{j=1}^p f_j^{(\tau)}(x_{j,t}) = a_t^{(\tau)} + \sum_{j=1}^p \sum_{k=1}^{n_k} b_{jk}(x_{j,t})\theta_{jk}^{(\tau)} \quad (3.15)$$

with the basis functions constructed under appropriate restrictions on $f_j(\cdot), j = 1, \dots, p$ [109].

Recall the definition of pinball loss function [107],

$$\rho_\tau(e) = \begin{cases} (\tau-1)e & \text{if } e < 0 \\ \tau e & \text{if } e \geq 0 \end{cases} \quad (3.16)$$

quantile regression aims to identify the τ -th quantile by minimizing $\sum_{t=1}^T \rho_\tau(y_t - q_t^{(\tau)})$ [107]. Replacing $q_t^{(\tau)}$ with (3.16) leads to the coefficients estimation:

$$\begin{aligned} & \left(\hat{a}_{t+k|t}^{(\tau)}, \hat{\boldsymbol{\theta}}_{t+k|t}^{(\tau)} \right) \\ & = \arg \min_{a, \boldsymbol{\theta}} \sum_{t=1}^{t-T} \rho_\tau \left[y_t - \left(a_t^{(\tau)} + \sum_{j=1}^p \sum_{k=1}^{n_k} b_{jk}(x_{j,t})\theta_{jk}^{(\tau)} \right) \right] \end{aligned} \quad (3.17)$$

where T is the number of training samples, y_t is the observation at time t , $\hat{a}_{t+k|t}^{(\tau)}, \hat{\boldsymbol{\theta}}_{t+k|t}^{(\tau)}$ are the constant and coefficient vector of basis functions estimated at time t for time $t+k$, respectively.

Once deriving the regression coefficients $\hat{a}_{t+k|t}^{(\tau)}$ and $\hat{\boldsymbol{\theta}}_{t+k|t}^{(\tau)}$, the forecasted quantile $\hat{q}_{t+k|t}^{(\tau)}$ can be readily computed nonparametrically via

(3.15). Similar to climatology PI, the resulting PI by BS-QR is also supposed to be centered on its median, that is, constituted by $\hat{q}_{t+k|t}^{(\alpha/2)}$ and $\hat{q}_{t+k|t}^{(1-\alpha/2)}$.

3.7.1.4 Bootstrap ELM (BELM)

BELM [90] is another powerful approach to derive PIs by taking data uncertainty and model uncertainty into account. Both uncertainties are assumed to be Gaussian distribution, wherein the noise uncertainty is modeled by the variance σ_ε^2 , and the model uncertainty is modeled by σ_y^2 . Assuming they are statistically independent, the total prediction uncertainty for time t associated with input vector \mathbf{x}_t is mathematically given by:

$$\sigma_i^2(\mathbf{x}_i) = \sigma_\varepsilon^2(\mathbf{x}_i) + \sigma_y^2(\mathbf{x}_i) \quad (3.18)$$

To estimate the model uncertainty, bootstrap is used to construct B_M training datasets by resampling with replacement from the original training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ and B_M ELMs are correspondingly applied to these bootstrapped datasets to generate B_M predictions. Suppose $\hat{y}_l(\mathbf{x}_i)$ is the output of l -th bootstrapped ELM, the mean of model uncertainty distribution is calculated by:

$$\hat{y}(\mathbf{x}_i) = \frac{1}{B_M} \sum_{l=1}^{B_M} \hat{y}_l(\mathbf{x}_i) \quad (3.19)$$

and the variance of model uncertainty is calculated by:

$$\sigma_{\hat{y}}^2(\mathbf{x}_i) = \frac{1}{B_M - 1} \sum_{l=1}^{B_M} (\hat{y}_l(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i))^2 \quad (3.20)$$

To estimate the uncertainty of data noise, by replacing the original target y_i with $(\hat{y}(\mathbf{x}_i) - y_i)^2$, a transformed training dataset can be obtained as $\{(\mathbf{x}_i, (\hat{y}(\mathbf{x}_i) - y_i)^2)\}_{i=1}^N$. Likewise, bootstrap is used to generate B_N predictions via B_N ELMs, suppose $\hat{r}_l(\mathbf{x}_i)$ is the output of l -th bootstrapped ELM, the estimated variance of noise $\hat{\sigma}_\varepsilon^2(\mathbf{x}_i)$ and variance related to the model $\sigma_{\hat{r}}^2(\mathbf{x}_i)$ are computed by:

$$\hat{\sigma}_\varepsilon^2(\mathbf{x}_i) = \hat{r}(\mathbf{x}_i) = \frac{1}{B_N} \sum_{l=1}^{B_N} \hat{r}_l(\mathbf{x}_i) \quad (3.21)$$

$$\sigma_{\hat{r}}^2(\mathbf{x}_i) = \frac{1}{B_N - 1} \sum_{l=1}^{B_N} (\hat{r}_l(\mathbf{x}_i) - \hat{r}(\mathbf{x}_i))^2 \quad (3.22)$$

Thus, the total noise uncertainty can be obtained through:

$$\sigma_\varepsilon^2(\mathbf{x}_i) = \hat{\sigma}_\varepsilon^2(\mathbf{x}_i) + \sigma_{\hat{r}}^2(\mathbf{x}_i) \quad (3.23)$$

Eventually, the PI with nominal coverage rate of $(1 - \alpha) \times 100\%$ can be expressed by:

$$\hat{I}_{t+k|t}^{1-\alpha} = \left[\hat{y}(\mathbf{x}_i) - z_{1-\alpha/2} \sqrt{\sigma_\varepsilon^2(\mathbf{x}_i)}, \hat{y}(\mathbf{x}_i) + z_{1-\alpha/2} \sqrt{\sigma_\varepsilon^2(\mathbf{x}_i)} \right] \quad (3.24)$$

where $z_{1-\alpha/2}$ is the critical value of standard Gaussian distribution.

3.7.1.5 Direct interval forecasting (DIF)

As the crisp counterpart of the proposed GELM, DIF approach is more appealing than most traditional PI construction methods on that it can directly produces the optimal PIs in terms of interval score [42, 153] without any statistical inference of the forecasting errors. Compared with our GELM model, the main difference is that DIF simply depends on the crisp inputs and crisp ELM, whereas the fitness function is identical to (3.8) in our model.

3.7.2 Comparative studies and discussion

3.7.2.1 Skill verification on different months

The modeling capabilities of the proposed GELM are verified against the five benchmarks from perspectives of reliability, sharpness and overall skill.

The mathematic expressions for these verification metrics have been described in Section 2.2.3, here we rewrite them as follows. Given the test dataset

$\{(\mathbf{x}_{t,i}, y_{t+k|t,i})\}_{i=1}^{N_D}$, and the corresponding PIs with $(1 - \alpha) \times 100\%$ nominal

coverage, the reliability of PIs is determined by ACE,

ACE=PICP-PINC

$$\text{PICP} = \frac{1}{N_D} \sum_{i=1}^{N_D} I_i \quad (3.25)$$

$$I_i = \begin{cases} 1 & \text{if } y_i \in [\hat{L}_i, \hat{U}_i] \\ 0 & \text{if } y_i \notin [\hat{L}_i, \hat{U}_i] \end{cases}$$

The sharpness of PI is expressed by

$$\text{PINAW} = \frac{1}{N_D R} \sum_{i=1}^{N_D} (\hat{U}_i - \hat{L}_i) \quad (3.26)$$

and the IS is written as

$$\text{IS}(\hat{L}_i, \hat{U}_i, y_i, \alpha) = \begin{cases} -2\alpha(\hat{U}_i - \hat{L}_i) - 4(\hat{L}_i - y_i) & \text{if } y_i < \hat{L}_i \\ -2\alpha(\hat{U}_i - \hat{L}_i) & \text{if } y_i \in [\hat{L}_i, \hat{U}_i] \\ -2\alpha(\hat{U}_i - \hat{L}_i) - 4(y_i - \hat{U}_i) & \text{if } y_i > \hat{U}_i \end{cases} \quad (3.27)$$

$$\overline{\text{IS}} = \frac{1}{N_D} \sum_{i=1}^{N_D} \text{IS}(\hat{L}_i, \hat{U}_i, y_i, \alpha)$$

Model training is implemented using a sliding window of two previous months to forecast the third month. For example, the data in January and February is used for training, and the data in March is used for model validating. Subsequently, training is performed for February and March, and prognosis is done for April. This rolling process continues until the last month in our dataset is evaluated. Besides, considering that the PIs with high confidence levels are more practically appealing to system operators and a shorter time scale is

coherent with the requirements of micro-grid dispatch, in this case study, PIs are evaluated with PINCs of 90%, 95% and 99% (i.e., $\alpha=10\%$, 5% and 1%), respectively. The granularity level is chosen as 10-min, meaning that each PI is constructed for the average GHI of next 10 mins. In this sense, the benchmarking models can be built with two input patterns, i.e., the previous 1-min raw data or the lagged 10-min averages. Through our numerical comparisons, see Appendix A, the models using 1-min raw data as inputs can yield better performance than that with the latter setting, therefore all results of benchmarks are obtained using the 1-min crisp inputs. Furthermore, to give the unbiased outcomes, each test is carried out for 10 times for DIF and the proposed model, the mean performance value is taken as the validating result. Finally, the results are summarized in Table 3.1, Table 3.2 and Table 3.3.

Table 3.1 Skill of proposed GELM model against benchmarks in terms of 90% PI over different months in 2012.
Best performance values are marked in bold.

Month	Persistence			Climatology			BS-QR			BELM			DIF			Proposed GELM		
	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS
Mar	98.24%	44.64%	-5.66%	62.22%	39.98%	-12.12%	86.70%	13.58%	-2.65%	90.71%	15.30%	-2.66%	86.46%	24.68%	-5.57%	87.22%	14.77%	-2.50%
Apr	95.71%	45.67%	-5.97%	74.14%	51.07%	-9.54%	75.05%	12.98%	-4.23%	82.63%	19.82%	-4.10%	84.34%	31.66%	-7.53%	88.79%	23.00%	-4.34%
May	96.04%	52.13%	-7.98%	75.12%	50.73%	-11.10%	78.00%	16.71%	-5.96%	82.36%	19.20%	-5.45%	83.24%	19.33%	-5.86%	86.02%	24.05%	-4.89%
Jun	94.90%	50.38%	-7.55%	92.86%	59.77%	-8.88%	86.51%	23.35%	-5.08%	87.63%	23.29%	-5.10%	86.77%	21.95%	-5.30%	86.57%	32.06%	-6.59%
Jul	95.50%	52.74%	-8.39%	76.25%	55.30%	-9.96%	85.38%	27.03%	-6.30%	88.51%	27.27%	-6.35%	86.36%	30.43%	-7.05%	90.62%	31.31%	-6.06%
Aug	96.53%	54.76%	-8.06%	97.84%	59.81%	-8.26%	94.43%	31.41%	-4.97%	93.26%	27.25%	-4.85%	94.89%	32.54%	-5.44%	89.05%	24.60%	-4.47%
Sep	96.76%	60.61%	-8.72%	91.66%	59.03%	-8.31%	90.55%	28.88%	-5.32%	87.78%	27.33%	-5.54%	91.11%	31.72%	-5.94%	91.72%	32.40%	-5.28%
Oct	98.28%	61.67%	-7.11%	89.59%	61.67%	-7.63%	82.56%	29.87%	-5.88%	88.22%	26.48%	-3.64%	94.33%	28.58%	-3.95%	93.30%	29.03%	-3.88%
Nov	97.88%	44.71%	-4.95%	69.65%	64.42%	-9.20%	78.11%	24.50%	-5.16%	84.85%	21.49%	-2.86%	92.63%	21.26%	-3.01%	91.36%	18.58%	-2.47%
Dec	97.61%	50.20%	-4.83%	94.92%	65.68%	-6.10%	95.11%	18.98%	-2.34%	90.78%	18.57%	-2.32%	92.86%	29.84%	-3.55%	89.30%	16.54%	-2.22%
Avg	96.74%	51.75%	-6.92%	82.42%	56.75%	-9.11%	85.24%	22.73%	-4.79%	87.67%	22.60%	-4.29%	89.30%	27.20%	-5.32%	89.39%	24.64%	-4.27%

Table 3.2 Skill of proposed GELM model against benchmarks in terms of 95% PI over different months in 2012.
Best performance values are marked in bold.

Month	Persistence			Climatology			BS-QR			BELM			DIF			Proposed GELM		
	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS
Mar	98.92%	53.19%	-3.38%	65.74%	42.63%	-8.61%	90.08%	16.69%	-1.72%	93.70%	17.94%	-1.63%	92.62%	23.80%	-2.83%	94.04%	24.02%	-2.00%
Apr	97.42%	54.42%	-3.54%	78.43%	54.23%	-5.65%	84.64%	17.07%	-2.66%	87.27%	23.32%	-2.64%	85.35%	17.27%	-3.42%	91.31%	25.77%	-2.49%
May	97.51%	62.12%	-4.75%	81.04%	54.94%	-6.22%	85.09%	21.65%	-4.11%	86.22%	22.90%	-3.56%	88.91%	22.25%	-3.94%	89.35%	24.36%	-3.65%
Jun	96.92%	60.04%	-4.50%	94.60%	63.49%	-4.76%	93.43%	30.04%	-3.07%	89.95%	27.86%	-3.33%	92.32%	29.04%	-3.34%	91.72%	29.76%	-3.69%
Jul	97.07%	62.84%	-4.98%	84.16%	59.29%	-5.26%	90.91%	33.55%	-3.72%	91.15%	32.64%	-3.80%	92.28%	36.31%	-4.24%	93.16%	34.54%	-3.69%
Aug	97.95%	65.25%	-4.77%	98.97%	63.73%	-4.37%	97.21%	39.38%	-2.99%	95.41%	32.48%	-2.90%	97.75%	39.68%	-3.04%	96.09%	37.35%	-3.29%
Sep	98.03%	72.22%	-5.20%	96.97%	64.43%	-4.37%	94.89%	38.42%	-3.31%	89.90%	32.59%	-3.53%	95.20%	37.23%	-3.60%	94.90%	36.82%	-3.49%
Oct	99.27%	73.49%	-4.22%	93.16%	66.78%	-4.04%	83.67%	40.65%	-3.20%	89.64%	31.98%	-2.22%	97.61%	37.62%	-2.38%	97.95%	39.07%	-2.50%
Nov	98.74%	53.28%	-2.96%	83.64%	70.65%	-4.50%	85.60%	34.46%	-2.63%	86.41%	25.41%	-1.75%	95.51%	27.66%	-1.78%	96.21%	32.71%	-1.96%
Dec	98.53%	59.82%	-2.88%	97.61%	68.58%	-3.17%	97.40%	26.48%	-1.84%	91.89%	22.09%	-1.46%	95.26%	30.14%	-1.93%	95.11%	27.21%	-1.80%
Avg	98.04%	61.67%	-4.12%	87.43%	60.87%	-5.10%	90.29%	29.84%	-2.92%	90.15%	26.92%	-2.68%	93.28%	30.10%	-3.05%	93.98%	31.16%	-2.86%

Table 3.3 Skill of proposed GELM model against benchmarks in terms of 99% PI over different months in 2012.
Best performance values are marked in bold.

Month	Persistence			Climatology			BS-QR			BELM			DIF			Proposed GELM		
	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS
Mar	99.76%	69.90%	-0.90%	99.66%	67.40%	-0.93%	97.99%	27.03%	-0.57%	96.97%	23.64%	-0.57%	98.00%	34.31%	-0.68%	98.44%	32.33%	-0.53%
Apr	99.04%	71.51%	-0.94%	82.58%	57.40%	-2.32%	93.28%	26.40%	-1.04%	91.52%	30.68%	-1.06%	95.56%	41.34%	-1.34%	96.05%	31.69%	-0.93%
May	99.17%	81.64%	-1.30%	85.92%	57.30%	-2.34%	91.10%	30.19%	-1.93%	93.84%	36.64%	-1.33%	96.43%	35.88%	-1.14%	96.58%	35.93%	-1.10%
Jun	98.64%	78.90%	-1.25%	97.02%	66.85%	-1.18%	97.22%	42.43%	-0.95%	94.04%	36.47%	-1.32%	96.52%	39.32%	-1.16%	97.32%	44.13%	-0.97%
Jul	98.78%	82.59%	-1.31%	92.52%	63.07%	-1.34%	97.31%	51.13%	-1.06%	94.92%	43.03%	-1.28%	97.02%	48.89%	-1.17%	97.61%	51.72%	-1.15%
Aug	99.17%	85.76%	-1.27%	98.97%	63.73%	-0.93%	99.75%	56.22%	-0.87%	97.21%	42.63%	-0.92%	99.27%	58.16%	-0.92%	99.12%	41.81%	-0.76%
Sep	99.19%	94.92%	-1.41%	98.48%	68.29%	-0.95%	98.48%	55.26%	-0.94%	91.16%	43.23%	-1.43%	98.48%	57.32%	-1.08%	98.03%	52.30%	-1.09%
Oct	99.90%	96.58%	-1.12%	94.38%	71.42%	-0.98%	95.99%	63.50%	-0.79%	91.10%	42.05%	-0.70%	99.46%	58.33%	-0.70%	99.56%	58.03%	-0.68%
Nov	99.29%	70.02%	-0.79%	88.23%	75.25%	-1.32%	95.10%	53.70%	-0.69%	87.68%	33.58%	-0.55%	97.37%	49.88%	-0.58%	97.17%	42.52%	-0.51%
Dec	99.51%	78.62%	-0.78%	98.19%	71.22%	-0.67%	99.41%	41.24%	-0.42%	93.94%	29.01%	-0.51%	97.80%	38.38%	-0.44%	97.31%	39.09%	-0.54%
Avg	99.25%	81.04%	-1.11%	93.60%	66.19%	-1.30%	96.56%	44.71%	-0.93%	93.24%	36.10%	-0.97%	97.59%	46.18%	-0.92%	97.72%	42.96%	-0.83%

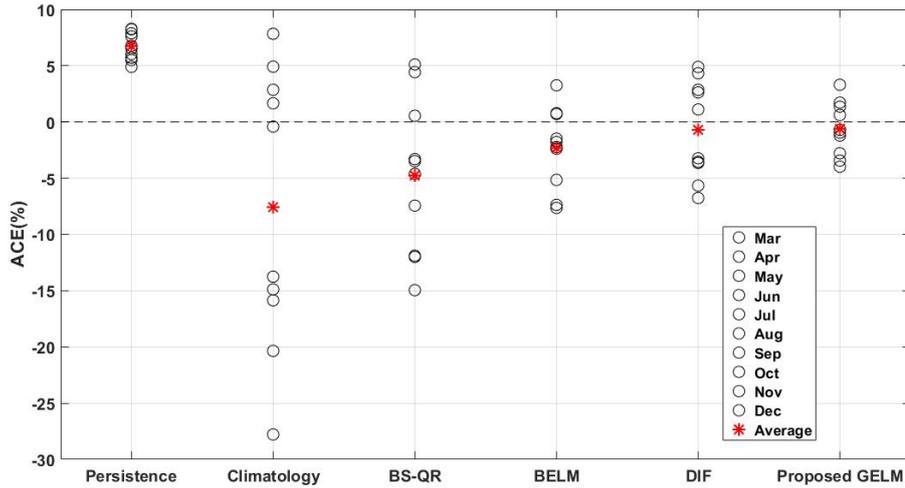
Inspecting the reliability of PIs with PICP in above three tables, we can see the proposed GELM model generally assures a superior performance on reliability, particularly in terms of 90% PIs, where the PIs in most months are identified to give the closest coverage rates to PINCs. As the PINC increases, this dominance fades away, but can still outperform the benchmarks with regard to average ACEs of 95% PI and take the second rank under 99% PI. This can be further illustrated by the means and standard deviations of ACEs listed in Table 3.4 and the scatters delineated in Fig. 3.5.

Table 3.4 Means and standard deviations of ACEs in terms of different PINCs over 10 months in 2012.

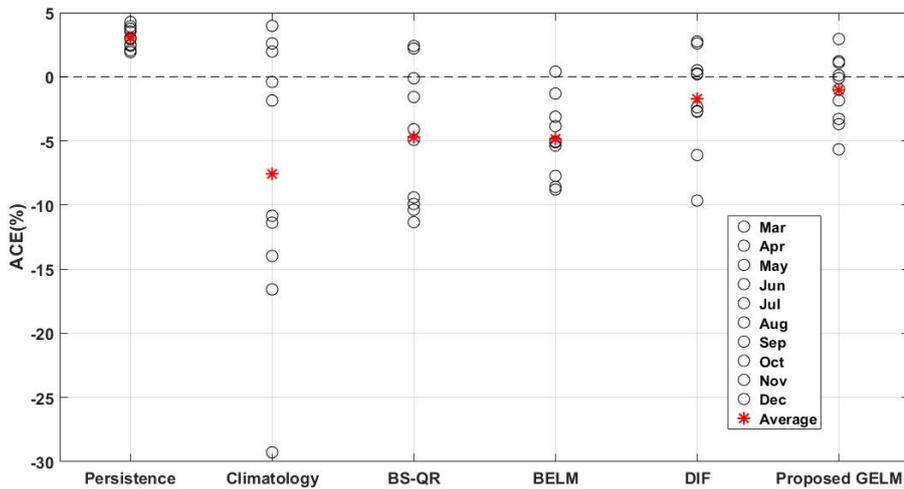
PINC	Statistics of ACEs	Persistence	Climatology	BS-QR	BELM	DIF	Proposed GELM
90%	Mean	6.74%	-7.58%	-4.76%	-2.33%	-0.70%	-0.61%
	Std	1.21%	12.34%	6.90%	3.54%	4.32%	2.37%
95%	Mean	3.04%	-7.57%	-4.71%	-4.85%	-1.72%	-1.02%
	Std	0.81%	10.69%	5.32%	3.02%	3.88%	2.62%
99%	Mean	0.25%	-5.40%	-2.44%	-5.76%	-1.41%	-1.28%
	Std	0.40%	6.09%	2.75%	2.92%	1.26%	1.09%

It is clearly observed from Table 3.4 that the persistence benchmark is more likely to give the unbiased ACEs, with the smallest standard deviations among all models. This is also visualized in Fig. 3.5, where the black circles closely spread around their means, this can be explained by its persistent nature. Despite the low variability of ACEs obtained in different months, it remarkably deviates from the PINCs in terms of 90% and 95% PIs, as it tends to over-estimate the PICP. It is also for this reason, the persistence benchmark

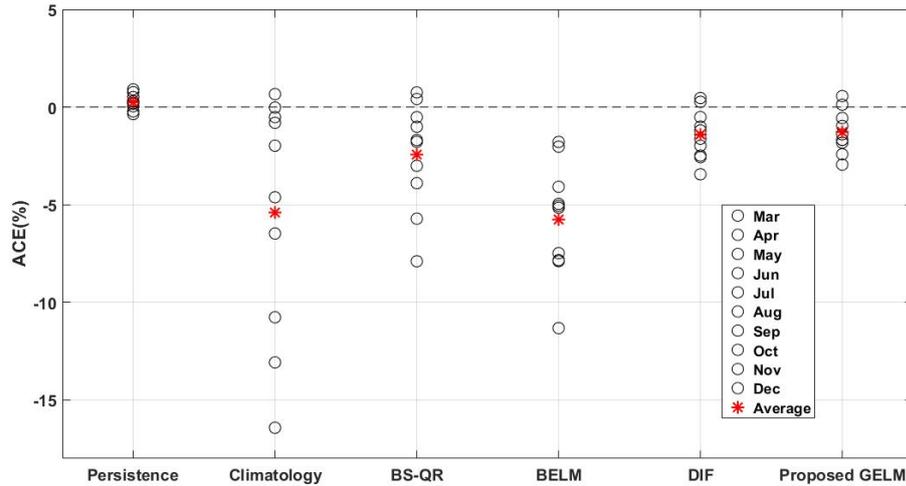
performs the best on both average and standard deviation of ACE for 99% PIs. Yet, by further looking into the sharpness indices with PINAW values in Table 3.1-3.3, the persistence model can hardly be used in practice due to the large PI width for all verified PINCs.



(a) Scatters of ACEs in terms of 90% PI



(b) Scatters of ACEs in terms of 95% PI



(c) Scatters of ACEs in terms of 99% PI

Fig. 3.5 ACE Scatters of different approaches over 10 months in 2012 in terms of different PINCs.

The proposed GELM model is able to guarantee the smallest average ACEs for both 90% and 95% PIs, while for 99% PI, it surpasses the other benchmarks (excluding persistence model), as can be seen in Table 3.4 and Fig. 3.5. In light of the standard deviations, the GELM model achieves the lowest values as compared to the benchmarks (excluding persistence model) for all PINCs, indicating its robustness in yielding reliable PIs in different situations.

Looking at the PINAW values recorded in Table 3.1-3.3 we see that, the BELM model is capable of generating the sharpest PIs for all evaluated PINCs. However, as shown in Fig. 3.5, BELM model is apt to under-estimate the PINCs, particularly for 95% and 99% PIs. In other words, the sharp PIs derived by BELM benchmark are achieved at the expense of lower PI reliability, which

is undesirable in probabilistic forecasts. The similar cases are discovered for BS-QR approach in March, April and May.

In respect of the overall skill, BELM comes as the most competitive rival to our GELM model, showing minor gaps on average IS for 90% PI, and even surpassing the GELM for the 95% PI. The fundamental reason underneath can be also attributed to the under-coverage property of BELM model, as collated in Table 3.4, larger mean ACEs and spread are identified compared with the GELM model. Given reliability is the foremost metric in probabilistic forecast, ACE should be assessed in priority. Therefore, the proposed GELM model is obviously more compelling than benchmarks as it can ensure both reliability and overall skill simultaneously for PIs with different nominal coverages.

Among the benchmarks, the DIF model appears to be more capable of producing PIs with high reliability, notwithstanding the inferior sharpness and IS. BS-QR and BELM can secure the best performance on PI sharpness, hence not surprisingly, the absolute IS values in several months are lower than that of GELM model. However, from Table 3.4 and Fig. 3.5, one can clearly notice that both BS-QR and BELM fail to provide reliable PIs, they tend to underestimate the corresponding PINCs and show poor robustness with regard to different cases. In this view, to give a comprehensive and justified assessment of the PI quality, the indices of ACE and IS should be integrated into consideration, which confirms the intention of our objective function (3.8). Overall, the proposed GELM model is able to guarantee the high reliability of

yielded PIs while keeping the PI width within an acceptable level, revealing its superior capability of constructing PIs with different levels.

Fig. 3.6 depicts the PIs with 90% confidence level estimated for 10 successive days in November, where most observations are noticed to be well included by the derived PIs with descent sharpness, even for the days with high volatility, e.g., the fourth and seventh day.

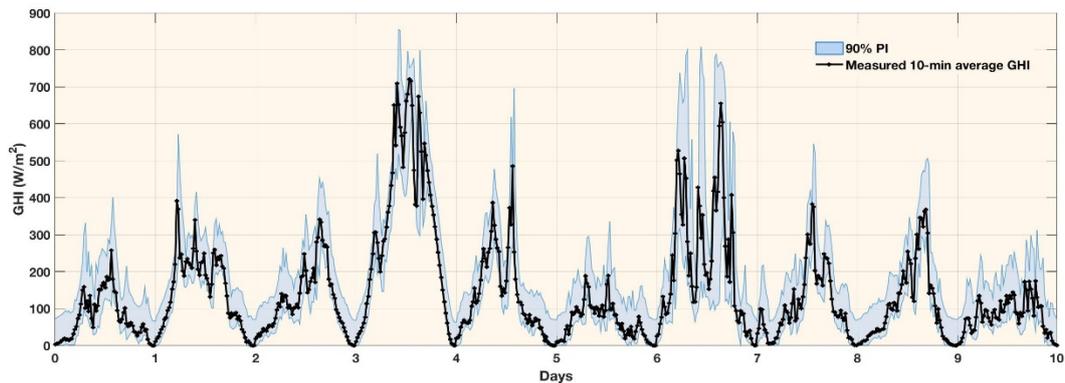


Fig. 3.6 Snapshot of 90% PIs associated with the 10-min average GHI observed in 10 successive days in November 2012.

3.7.2.2 Skill verification on different granularities and granulation methods

This section investigates the performance of different granulation methods with three different levels of granularity (temporal scales): 10 minutes, 30 minutes and 1 hour. The data in November is used for validating. As seen from Table 3.5, the range of PIs gets wider as the granularity level increases, so does the absolute interval score. This can be explained by the fact that more variability and uncertainty is involved in larger granules. In general, both granulation techniques are effective in generating PIs with high reliability,

while min-max-Gr is likely to achieve narrower PIs with different nominal levels, thus leading to a better overall performance than that of FCM-Gr. This might be due to that the FCM-Gr offers more conservative temporal granules than min-max-Gr.

Table 3.5 Skills of different granularities and different granulation methods evaluated in November 2012.

PINC	Granularity	10min			30min			60min		
	Granulation Methods	PICP	PINAW	IS	PICP	PINAW	IS	PICP	PINAW	IS
99%	min-max-Gr	97.27%	34.83%	-0.46%	98.94%	39.34%	-0.71%	98.18%	55.67%	-0.69%
	FCM-Gr	97.02%	36.04%	-0.50%	99.09%	57.91%	-0.73%	99.09%	69.72%	-0.87%
95%	min-max-Gr	95.86%	24.82%	-1.57%	97.58%	39.39%	-2.46%	95.76%	43.39%	-2.61%
	FCM-Gr	95.00%	31.22%	-1.96%	97.12%	43.06%	-2.59%	98.79%	54.47%	-3.09%
90%	min-max-Gr	91.36%	18.58%	-2.47%	91.52%	26.03%	-3.42%	89.70%	36.86%	-4.76%
	FCM-Gr	92.98%	24.75%	-3.17%	90.30%	36.15%	-4.80%	91.76%	43.44%	-5.05%

3.8 Summary

Generally, different decision-makers in power systems have different lookahead time preferences ranging from minutes to days for renewable generation forecasts according to their own operational requirements. Very short-term solar power or solar irradiance prognosis is crucial for micro-grid economic dispatch, the operation of storage systems associated with temporal market regulations such as Australian National Electricity Market with 5 minutes resolution, and the TSO which aims to optimally dispatch reserves for the continuous balance of the power system.

The goal of this research is to quantitatively represent the uncertainty of the AI-based forecasting process, originating both from the uncertainty in inputs and in the forecasting model itself. Inspired by the conceptualization of IGs, the variability involved in the raw detailed numeric data are abstracted by specified temporal granules thus providing granular model inputs. In parallel, the original crisp prediction model is granulated as well to accommodate the granular inputs. As such, the uncertainty can be well reflected in the resulting granular output. In this thesis, we rely on the ELM-based NN, whose powerful mapping ability could immensely benefit the learning process of its generalized granular counterpart. Specifically, the 10-min granules are focused that given to the model inputs, which can faithfully capture the variability within 10 minutes. Then, PSO is used to train the underlying granular model in a PI-score-oriented framework. Through verifying against five benchmarks, including its crisp counterpart - DIF, the proposed GELM model proves to be effective in yielding PIs with higher reliability and overall quality. Additionally, in regards of the two granulation methods we offer, the min-max-Gr achieves slightly better PI skill than FCM-Gr in this experiment.

All in all, compared to traditional crisp models, the advantages of proposed GELM are manifold. Firstly, in the situation of massive data with high resolution and variability, by constructing IGs over a lower temporal scale, the mapping capability of the underlying model could be improved as the granulated time series virtually mitigate the variability of raw complicated numeric data with refined time scales. In the meanwhile, the key features are

retained thus being able to give a faithful quantification of the natural variability. Secondly, this framework is not solely limited to temporal variability, while it can be flexibly extended to data measuring noises if the information within certain temporal scale is unavailable in a more detailed level. In such cases, the information granules, depending on the gaussian noise or the empirical measurement errors of sensors, are advocated to build for inputs. Thirdly, unlike most of existing PI construction methods, GELM does not require any quantile analysis of point forecasting errors involving statistical inferences. Instead, it only involves one optimization stage to directly derive the optimal PIs.

Note that this research only provides a generic framework to quantify the uncertainties of NN-based models, the future improvement can be emphasized on the following aspects:

(1) Since our model is completely dependent on the historical observations, satisfactory predictive skill can be achieved over horizons only within few time-steps. To accurately predict the solar irradiance or PV power for longer look-ahead times, more influential factors are advocated to be included, such as cloud motion, pressure, sky clearness index and geographical coordinates [154]. Furthermore, if the ramp events are focused in the very short-term forecast for the small-scaled PV, sky cameras are commonly deployed to track the positions and velocities of sun and cloud, which proves to be a reliable tool for ramp rates management. Therefore, future work is anticipated to construct the granular counterparts for more potential predictors.

(2) This study builds up the temporal granules in an intuitive and straightforward manner, i.e., min-max-Gr and FCM-Gr. Some sophisticated granulation approaches are highly needed to be exploited to construct a more justifiable granule. Currently, a comprehensive discussion on the realization of IGs is evidently lacking, further work can be focused on the optimal construction of IGs based on a well-defined justifiable principle, e.g., legitimate and specificity [144]. Additionally, instead of assigning a fixed temporal scale for all granules in this thesis, a scalable granularity is more appealing to adhere to the underlying principle of IGs.

(3) Other formalisms of IGs that might be appropriate to quantify the uncertainties are highlighted to investigate. Interval Type-2 Fuzzy Sets (IT2 FS) is demonstrated to be a vigorous tool to handle the uncertainties [155]. Further work would attempt to incorporate the IT2 FS into AI models to yield the probabilistic forecasts.

Chapter 4 Conditional Density Forecast of Electricity

Price based on Ensemble ELM and Logistic EMOS

4.1 Introduction

With the deregulation of electricity market worldwide, electricity price has become an indispensable signal for all participants involved in this competitive environment, as it constitutes considerably valuable information for their decision-makings. Despite a variety of periodicities (e.g., daily and weekly) can be tracked in spot price series, it normally exhibits volatile and irregular patterns, even unanticipated spikes [36, 156] that impose great challenges to trading activities of various market agents. A good knowledge of future prices in advance can help market participants submit effective bids with low risks and make sensible bilateral transaction decisions [36].

Extensive research work has been carried out on the point prediction of electricity prices over last two decades [9]. Nevertheless, such forecasts have limitations due to their inability to inform the inevitable error information involved, which is fairly crucial for sagacious decision makings considering diverse uncertainties. This boosts the shift towards a more informative forecast tool based on a probabilistic framework [46]. Generally, the probabilistic forecasts can be represented by prediction intervals (PIs) or probability density functions (PDFs). The majority of existing PIs construction approaches rely on the forecasting residuals to derive the PIs via either Gaussian noise assumption [81, 82, 84] or bootstrap method [88, 93]. Alternatively,

the PIs can be obtained by combining several specific quantiles of interest in nonparametric or semiparametric ways, such as quantile regression (QR) [61, 114, 157], kernel density estimation [82, 104] and statistical moments estimation [100]. Recent advance [42, 158, 159] aims to construct optimal PIs by using a PI-score-based cost function for the parameters' estimation via a direct optimization approach.

Despite pervasive studies done in interval forecast, density forecast is believed to provide more utility to decision makers since it provides full distributional information on the future uncertainties than merely a single or a set of PIs. Most importantly, due to miscellaneous uncertainties in modern market, decision makers are inclined to consider their objectives in a stochastic framework [14, 16, 17] instead of conventional deterministic ones. Predictive density is therefore a valuable input for, e.g. stochastic programming based decision-making models, as it allows the generation of temporal scenarios. Similar to interval forecasts, the predictive errors are commonly assumed as normally distributed and a variety of generalized autoregressive conditional heteroscedasticity (GARCH) based models were developed [85, 86]. A time-varying model for the first few moments estimation of price densities was presented in [100] to overcome the limitations in traditional simulation-based PI construction approaches, but the authors only discussed the coverage accuracy of several extracted PIs without evaluating the performance of full predictive distribution. [83] proposed a vector autoregressive model with a skew t-distribution for price forecasting noises, despite the lack of back-transforming the logarithmic price series, the skill assessment for the issued densities by CRPS is fully described and found to be more appealing than Probability Integral Transform Score

(PITS) and Logarithmic Score [85] for density forecasts. Further, predictive density can be approximately built on the basis of finite quantiles. [160] applied QR to generate 5%-95% (in steps of 5%) quantiles for forecasting residuals and treated predictive distribution tails with an exponential assumption. Although this semiparametric hybrid model is demonstrated to be superior in several benchmarks, it can only provide a discrete description of the predictive distribution.

This work aims to develop a conditional density forecast for the Marginal Clearing Price (MCP) on the purpose of offering the users a comprehensive understanding of the statistical description of stochastic uncertainty and knowledge uncertainty –the two of which are considered to constitute the total uncertainty involved in a modeling system [138]. The former is subject to the natural variability of the underlying dataset and the measurement errors, whereas the latter chiefly accounts for the imperfect representation of processes in a model and imperfect knowledge of parameters associated with these processes. A NN based on ELM [91] is adopted as the point estimator in this thesis. ELM-based NNs have received great popularity lately since iterative parameters' tuning in traditional gradient-based NN are avoided in the learning process of ELM [62, 88, 90, 93]. However, what is controversial for ELM is the instability owing to its random nature [161], i.e., the generalization performance varies with different initial parameter settings, which gives rise to large uncertainty in the outputs. In this sense, an appropriate representation of the knowledge uncertainty in ELM based NNs is imperative as it consists of more uncertainty concerning models' setting than traditional NNs.

An aggregate structure comprising collective ELM networks is employed to generate multiple predictions in this study. Later, these ensemble forecasts are fed into a post-processing unit - EMOS to derive the predictive distribution. Current applications of EMOS [64] are largely confined to the numerical weather prediction, in which the bias and dispersion errors requires to be corrected by the post-processing unit. This inspires us to apply it into our model to quantify the uncertainty of ELM. Through fitting the EMOS model in a performance-oriented framework with CRPS, both the reliability and sharpness of the predictive density can be guaranteed. Additionally, given Normal distribution is too conservative to model the MCPs with occasional spikes, the Logistic distribution is proposed to model the ensemble forecasting errors due to its property of heavy tails that are more robust to the outliers.

The main contributions of this work are summarized as follows: (1). To the best of the authors' knowledge, this research is the first to apply the Logistic-EMOS model to derive the predictive density of MCP conditional on the ensemble forecasts via ELMs. By assessing the predictive skill for the full distribution with CRPS, the entire ELM-Logistic-CRPS-based EMOS (ELC-EMOS) model proves to outperform the benchmarks. (2). A comprehensive analysis of significant price drivers is conducted with respect to the case market, where substantial explanatory variables can be collected. (3). By taking advantages of the well-calibrated predictive density and Gaussian copula, the next-day price scenarios are produced. When dealing with the temporal interdependence, a highly efficient approach relying on the empirical correlation coefficients of the observed probabilistic forecasts in the similar weekdays is proposed to determine the covariance structure, thus avoiding

complicated computations and assumptions in traditional scenario generation methods. At last, the estimated price scenarios based on the developed covariance structure are verified against benchmarks with two proper scoring rules: energy score and variogram score. Results demonstrate that the proposed covariance structure can yield most skillful scenarios.

4.2 The Data

Nord Pool Spot operates two different physical operation markets, the day-ahead market (Elspot) and the intraday market (Elbas), in 15 bidding areas over 7 Nordic countries [7]. In Elspot, participants are required to submit their sell and buy orders, including prices and volumes, to Nord Pool no later than the deadline of clearing for the following operation day. Based on these bids, Nord Pool releases the system price in hourly format at 12:00 noon before the following day.



Fig. 4.1 Four bidding areas (SE1, SE2, SE3, SE4) in Sweden with corresponding MCP at certain hour on 23rd Apr 2018

The Swedish Elspot is selected as the case market in this thesis and the average MCPs of 4 bidding areas are used as the targets. Fig. 4.1 is a snapshot of 4 bidding areas in Sweden. Considering that different areas in Sweden show minor differences on MCPs, the selected market be regarded as an entirety. The Swedish MCPs and the other explanatory dataset can be publicly accessed on Nordic Pool Website [162], Swedish TSO Website [163] and the Global Weather Information Website [164]. Totally 31 explanatory variables are collected from 1st Jan 2013 to 31st Jul 2016, as listed in Appendix B. The clock-change issue is neglected in this study, for all missing

variables at 02:00 am in March, the mean value of its previous and next hour is interpolated; for the duplicates occurring at 2:00 am in October, we averaged them to ensure the equal size of samples for each day.

4.3 Quantification of Uncertainty in Swedish MCP Forecasting

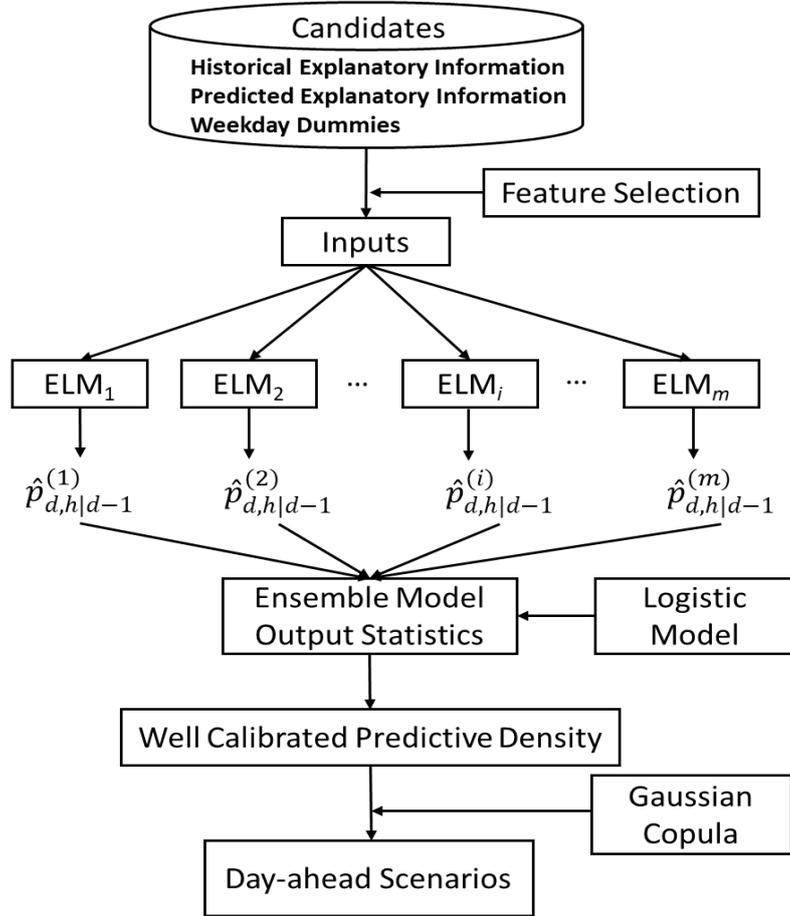


Fig. 4.2 Schematic of proposed forecasting strategy

The entire forecasting strategy is illustrated in Fig. 4.2. At first, all candidates listed in Appendix B are collected to perform correlation analysis to extract the most relevant features as inputs. Then, these inputs are fed into several individual ELM networks to generate a set of forecasts $\hat{p}_{d,h|d-1}^{(i)}$. These ensemble forecasts are further

incorporated into EMOS model to derive the conditional density with proper distribution. Finally, we can generate the next-day scenarios based on the predictive densities and Gaussian copula. $\hat{p}_{d,h|d-1}^{(i)}$ is the point estimate of MCP for hour h in day d issued in the previous day $d-1$ via the i -th ELM network, $i = 1, 2, \dots, m$. m is the total number of ELMs. Details of each block are described in the following subsections.

4.3.1 Feature Selection

Feature selection is an important task in the process of machine learning, aiming to seek the most relevant candidates as inputs to improve generalizing performance and reduce overfitting [156]. This can be achieved by a set of metrics, such as mutual information, distance correlation, learning-based search and linear correlation analysis. The former three are commonly restricted by the abundant time complexities of dealing with a number of explanatory variables and the associated lags. Linear correlation has the merits of its simple interpretation and low computational burden for large datasets. Moreover, it can measure both the degree and direction of the correlation between two series. Pearson correlation coefficient [165] is used to identify the appropriate input features of ELMs in this phase.

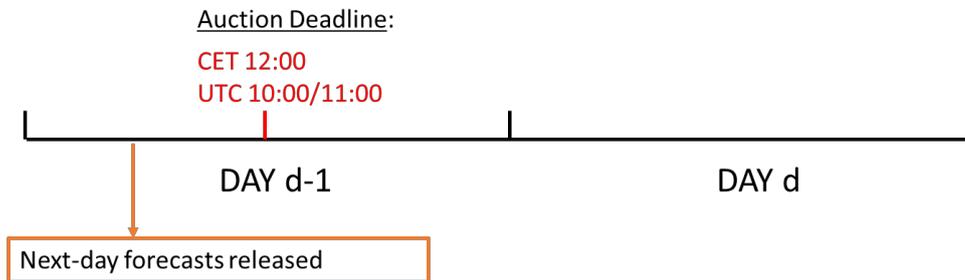


Fig. 4.3 Timeline of Nord Pool Elspot

The potential influential variables are listed in Appendix B. All the forecasted variables are available before the auction deadline, as Fig. 4.3 illustrates. Correlation analysis is executed between each candidate feature (with different lags) and the target Swedish MCP under the mechanisms of Elspot. Through comprehensive analysis, some interesting findings are presented as follows.

The total Nordic system production was identified to have a higher correlation with Swedish MCP than domestic generation, the same observation holds for the demand. This could be explained by that the area prices are not purely dependent on the area demand and supply but also reckon on the transmission capacity between the connected bidding areas. Hence, from a broad view, the Swedish MCP is more correlated with the aggregated load and production in the Nordic system.

Some literatures report that the reservoir level and the nuclear power are the prominent price drivers, since the hydro and nuclear power dominate the production in Nordic. However, through our experiments, both energy forms have weak correlations with the Swedish MCP. This might be owing to that the hydrologic level and nuclear production is quite stable, thus only having strong impacts on variations of MCP in the long-run, rather than the hourly behavior, i.e. the higher the reservoir level the lower the prices and vice versa. This can also be used to explain the minor influence of temperature on the hourly spot prices. As the leading source in renewables, wind power generation shows a negative but small effect on Swedish MCPs due to its zero fuel cost.

The predominant price contributors were exposed to be the net demand, generation from conventional units and the import/export power. The term ‘net

demand' refers to the gross demand minus the generation from non-dispatchable renewables (wind power and PV power). In parallel, 'generation from conventional units' represents the gross electricity production minus the generation from non-dispatchable renewables. It is worth noting that the exchange power shows a hysteretic effect on the MCP, through analysis, the highest correlation with MCP was identified in historical exchange power before 54 hours and the forecasting exchange power of 7 hours ago.

In light of the correlation coefficients, the most relevant variables for MCP are its lagged observations, particularly for those at the same hour lagged by 1, 2, 6 and 7 days, as Fig. 4.4 shows. Recent study [166] advises us to take the most recent available MCP (i.e. the last observation of the previous day) into account, as a result, a strong correlation is also observed in our case, as clearly seen in Fig. 4.5, especially for the nighttime during 21:00-06:00. In addition, the weekday dummies have a notable impact on the daily patterns of MCP. Through examining the average intraday price series with respect to different weekdays (Mon, Tue, ..., Sun) in Swedish Elspot, as visualized in Fig. 4.6, the mean daily price series within Monday to Thursday are noticed to be extremely close, whereas the profiles in Saturday and Sunday show clear deviations. The price series in Friday during 12:00-24:00 exhibits a distinct difference from that in both weekend and other working days, hence, four weekday dummy variables - $\phi_1, \phi_2, \phi_3, \phi_4$ - representing Monday, Friday, Saturday and Sunday, are taken into account for the weekly seasonality.

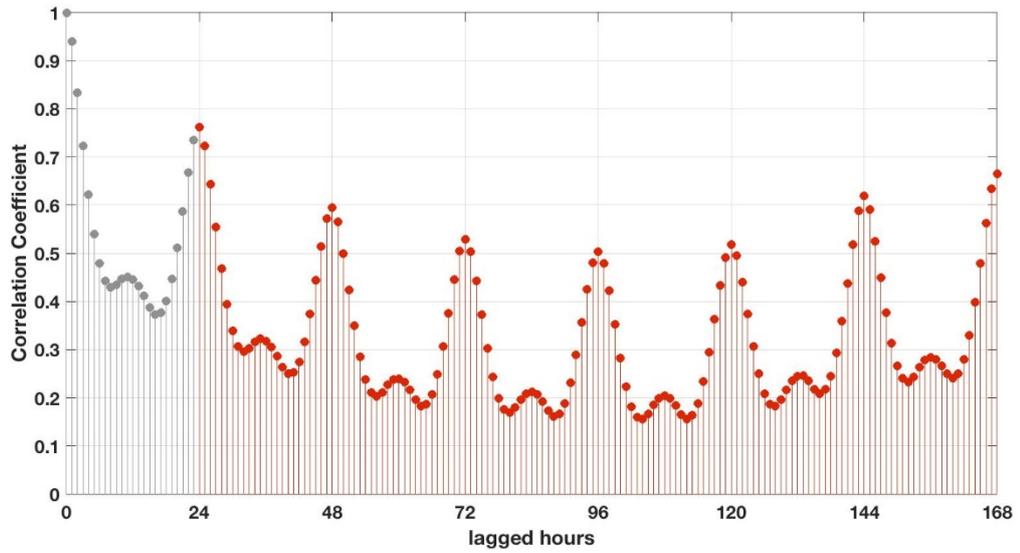


Fig. 4.4 Autocorrelation coefficients of Swedish MCP in 2013. Grey lines mean data that are not available at forecast.

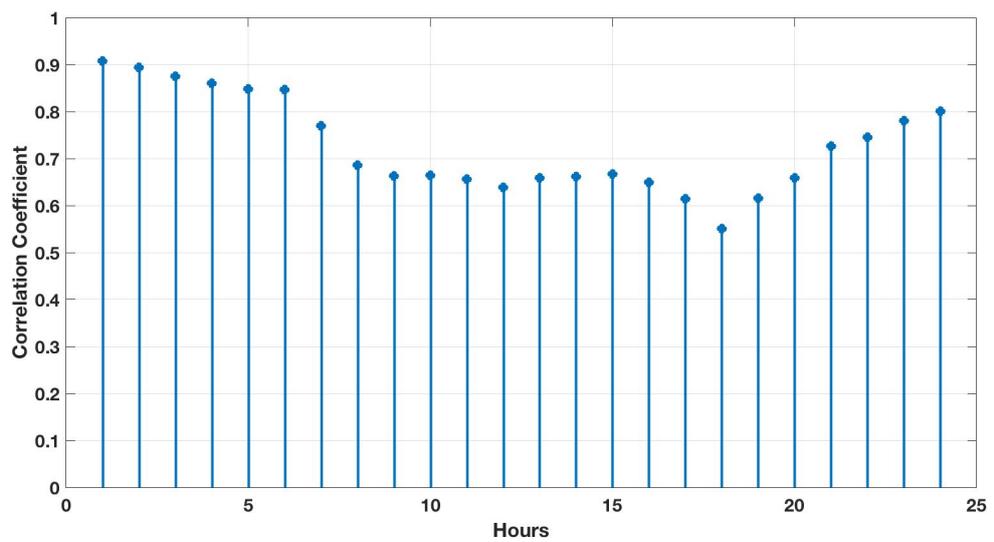


Fig. 4.5 Correlation coefficients between MCP and its most recent price over 24 hours in 2013.

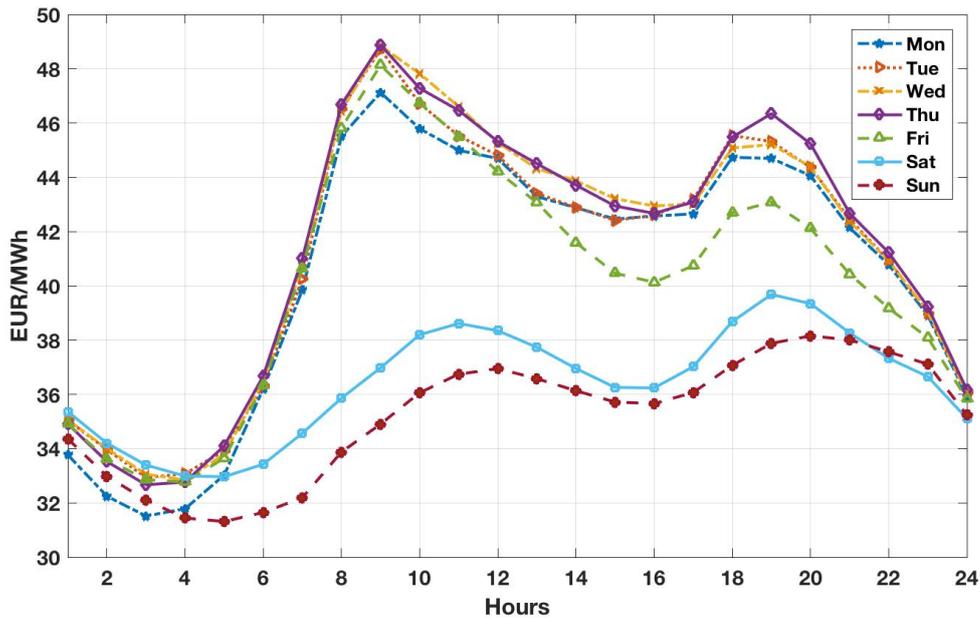


Fig. 4.6 Intraday profiles of average MCP in terms of different weekdays in 2013.

Table 4.1 Selected features for Swedish MCP

Predictor	Symbol	Units
24 hour ahead domestic MCP	$P_{SE,t-24}$	€/MWh
48 hour ahead domestic MCP	$P_{SE,t-48}$	€/MWh
144 hour ahead domestic MCP	$P_{SE,t-144}$	€/MWh
168 hour ahead domestic MCP	$P_{SE,t-168}$	€/MWh
The most recent available domestic MCP	$p_{SE,t}^{last}$	€/MWh
54 hour ahead actual import/export power	E_{t-54}	MWh
Prognosis domestic net demand	$\hat{D}_{SEnet,t}$	MWh
Prognosis system net demand	$\hat{D}_{SYSnet,t}$	MWh
Prognosis system generation from conventional units	$\hat{G}_{SYSconv,t}$	MWh
Prognosis import/export power at 7 hours before the time of interest	\hat{E}_{t-7}	MWh
Weekday dummies (Mon, Fri, Sat, Sun)	$\phi_i(i=1,2,3,4)$	-

Performing the correlation analysis with regard to the endogenous and exogenous variables, respectively, the irrelevant attributes are removed and the rest are further examined by the redundancy analysis [156] to filter out the redundant features that would also lead to the degradation of forecasting performance. Eventually, 11 predictors are selected in Table 4.1. It should be noted that the variable selection in such an ad hoc manner is somehow simple, since our study does not focus on the sophisticated feature engineering, but to showcase the feasibility of our method as a forecasting model given a certain set of features and compare it with benchmarks. To further improve the forecasting skill and efficiency, regression models with automated variable selection or shrinkage ability can be applied, such as LASSO [166], elastic nets [167], random forests [164], regularized ELM [168], etc.

4.3.2 Ensemble Forecasts on the basis of ELM Networks

ELM is an emerging training algorithm for single hidden-layer feedforward neural networks (SLFNs). The essence of ELM is the input weights and biases of hidden layer are randomly assigned and free to be tuned further [91].

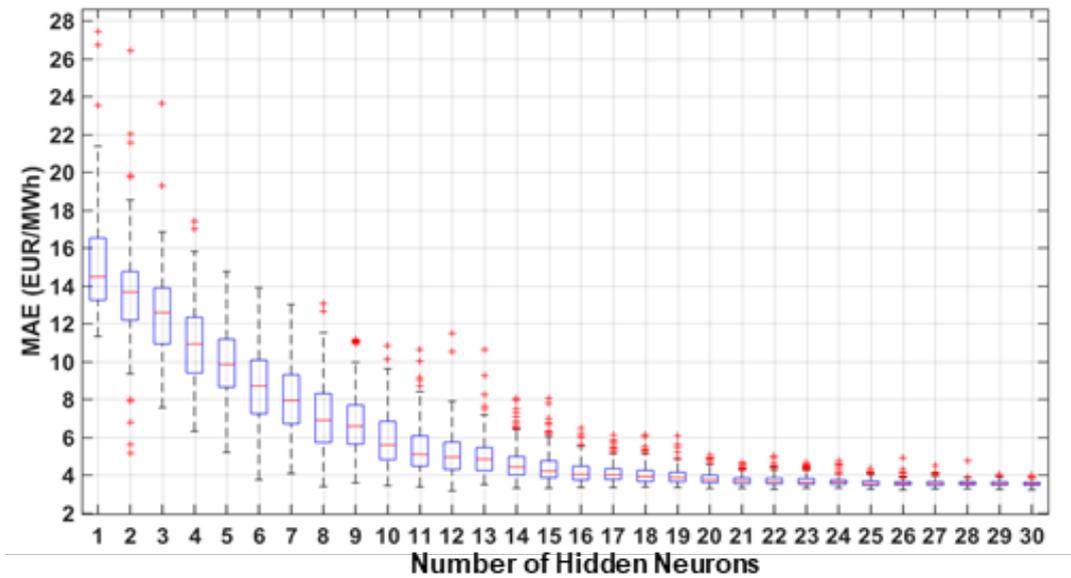


Fig. 4.7 Typical generalization performance with respect to different ELM structures.

Despite a variety of superiorities over traditional networks, such as better generalization performance with extremely fast learning speed, evasion of local minima, learning rate, momentum rate and over-fitting [91]. The biggest issue of ELM is the instability due to its random nature [161], that is, it is not able guarantee the optimal performance in the training phase, not even for an unseen test dataset. This can be illustrated in Fig. 4.7, where the generalization performance with respect to different ELM structures (with hidden neurons from 1 to 30) but the same training dataset are examined for forecasting of the hourly Swedish MCP in Jun 2015. For each structure, a group of ELMs with different initial parameter settings are implemented to yield the statistical results of generalization performance indicated by mean absolute errors (MAEs). As depicted in Fig. 4.7, large variability and outliers are discovered with a small number of hidden neurons, indicating significant model

uncertainty. With the increase of hidden neurons, MAE drops off distinctly and finally converges at around 4 EUR/MWh. However, a variability range of roughly 1 EUR/MWh is still noticeable at the end, which cannot be ignored and needs proper quantification in the output.

In view of the above, combining a number of ELMs can be an effective way to quantify such model uncertainty. Suppose m ELMs with predefined same structure are independent of each other, by randomly assigning the initial parameters to each ELM, m different outputs can be derived and are further fed into EMOS module to generate predictive densities, thus the model uncertainty can be well informed. In this study, the individual ELM is fitted by the samples over the lagged 84 days (12 weeks) in a time-adaptive way for the next day.

4.3.3 Logistic Distribution based EMOS

EMOS [64] is essentially a post-processing technique to calibrate the ensemble forecasts, where the unknown parameters are estimated with a pre-defined distribution in a time-adaptive learning manner [169]. The EMOS is found to be more compelling than Bayesian Moving Average (BMA) due to its conceptual simplicity [64]. To identify the most appropriate shape of ensemble forecasts' distribution, numerous ELMs with different initial settings are run to generate a series of potential predictions of MCPs. Through examining various ensemble forecast errors at different time slot, a notably underdispersive character among the forecasts was explored, which means the ensemble ELMs either underestimate or overestimate the validating observations. As illustrated in Fig. 4.8 (a), most forecasts are below the real MCPs before 7:00 am, and turn to overestimate in the following period. Such

bias may result from the structural model deficiencies shared among all ensemble members or the insufficient explanatory information [170]. Broadly, the ensemble errors exhibit a unimodal and symmetric property, as shown in Fig. 4.8 (b), which can be well fitted by Normal or Logistic distribution [171].

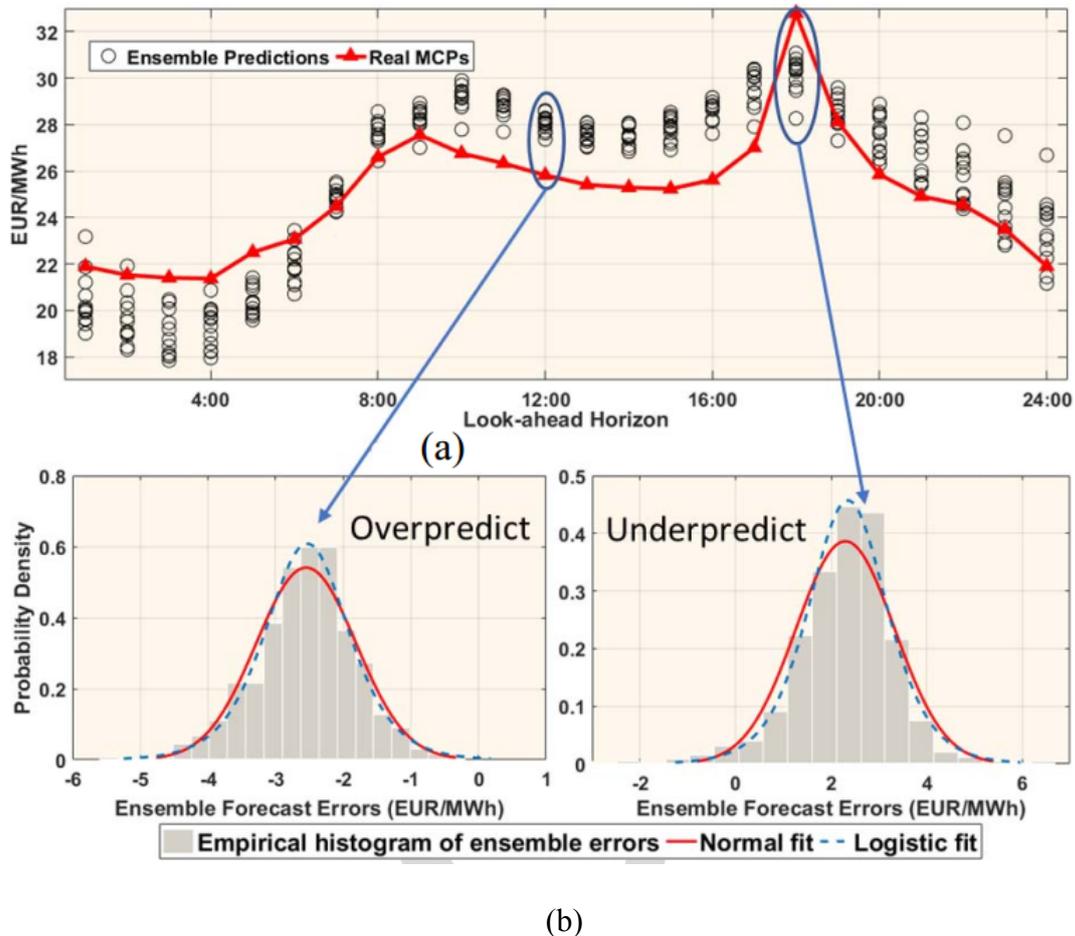


Fig. 4.8 (a) The potential ensemble predictions and real MCPs on 17th Nov 2015. (b) Histogram of potential ELM's ensemble errors and the corresponding Normal and Logistic fits

Gaussian-based EMOS model has been fully developed and extensively used to calibrate the ensemble forecasts for weather quantities [172]. Logistic distribution [171] is close to Normal distribution in shape but has slightly fatter tails (higher

kurtosis), thus being more robust to the ensemble outliers and small spike prices in this case. The comparative results between Normal and Logistic models will be elaborated in Section 4.5.

The probability density function (PDF) of Logistic distribution is defined as

$$f(x) = \frac{e^{-(x-\mu)/\zeta}}{\zeta \left(1 + e^{-(x-\mu)/\zeta}\right)^2} \quad (4.1)$$

where μ is the location parameter, ζ is the scale parameter and is proportional to its standard deviation via $\sigma^2 = \pi^2 \zeta^2 / 3$. According to the formulated process of Gaussian based EMOS [64], the predictive mean is the bias-corrected weighted average of the ensemble forecasts $\hat{p}_{d,h|d-1}^{(1)}, \hat{p}_{d,h|d-1}^{(2)}, \dots, \hat{p}_{d,h|d-1}^{(m)}$ and the variance is a linear function of the ensemble variance $S_{d,h}^2$. Hence, the Logistic predictive distribution for hour h in day d issued in the previous day $d-1$ can be viewed as conditional on the ensemble estimates, which is parameterized by

$$\begin{aligned} \hat{\mu}_{d,h|d-1} &= \alpha_{d,h}^{(0)} + \alpha_{d,h}^{(1)} \hat{p}_{d,h|d-1}^{(1)} + \dots + \alpha_{d,h}^{(m)} \hat{p}_{d,h|d-1}^{(m)} \\ \hat{\sigma}_{d,h|d-1}^2 &= \beta_{d,h}^{(0)} + \beta_{d,h}^{(1)} S_{d,h}^2 \end{aligned} \quad (4.2)$$

As a result, the entire process translates to the estimation of parameters, $\alpha_{d,h} = (\alpha_{d,h}^{(0)}, \dots, \alpha_{d,h}^{(m)})$ and $\beta_{d,h} = (\beta_{d,h}^{(0)}, \beta_{d,h}^{(1)})$, for a linear model.

4.3.4 Parameter Estimation

In this stage, a proper scoring rule must be determined as the optimum objective since $\alpha_{d,h}, \beta_{d,h}$ are desired to be estimated in a performance-oriented framework. In order to assess the reliability and sharpness of predictive density simultaneously, the CRPS is considered as a robust choice. Unlike the other metrics, such as interval score and quantile score [58], CRPS does not focus on any specific point of the probability

distribution, but considers the distribution as whole. The general form of CRPS has been given in (2.11).

In order to facilitate the further manipulation by optimization tool, the integral form of CRPS should be replaced by a generalized closed-form expression, which has been done concerning Normal distribution [64]. In this research, the Logistic distribution based closed-form analytic expression of CRPS is derived using the pinball loss function [58], as expressed in (4.3). The detailed mathematical derivations are given in Appendix C.

$$CRPS(F_{\mathcal{L}(\mu, \zeta)}, p) = (p - \mu)(2F(p) - 1) - 2\zeta F(p) \logit(F(p)) - 2\zeta \log(1 - F(p)) - \zeta \quad (4.3)$$

where p is the evaluated point and $F(p)$ is the associated probability, \logit is the inverse Logistic function, given by $\logit(p) = \log(p) - \log(1 - p)$.

Therefore, the regression coefficients $\boldsymbol{\alpha}_{d,h}$, $\boldsymbol{\beta}_{d,h}$ can be estimated by

$$(\boldsymbol{\alpha}_{d,h}, \boldsymbol{\beta}_{d,h}) = \underset{\boldsymbol{\alpha}_{d,h}, \boldsymbol{\beta}_{d,h}}{\operatorname{argmin}} \left(\frac{1}{T} \sum_{d=1}^{d-T} \sum_{h=1}^{24} CRPS(F_{\mathcal{L}(\mu_{d,h}, \zeta_{d,h})}, P_{d,h}) \right) \quad (4.4)$$

where T is the length of training days, the identification of optimal T will be discussed in the following section. The optimization is implemented by Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [173]. Finally, the predictive density can be obtained by (4.2).

4.4 Temporal Dependency Modeling

Based on the well calibrated marginal predictive distribution obtained in the last section, a set of temporal forecasting scenarios can be generated by using the popular multivariate statistical tool, Gaussian copula [68].

By introducing a multivariate random vector $\mathbf{Z}_d = [Z_{d,1}, \dots, Z_{d,h}, \dots, Z_{d,24}]$ for the 24 hours in day d , where $Z_{d,h} \in \mathbb{R}^{1 \times 1}$ is a random variable following standard normal distribution, J realizations can be issued. \mathbf{Z}_d follows multivariate normal distribution with zero expectation vector $\mu_d \in \mathbb{R}^{24 \times 1}$ and a covariance matrix $\mathbf{R}_d \in \mathbb{R}^{24 \times 24}$. Subsequently, using the CDF of standard normal distribution Φ and the inverse CDF $\hat{F}_{d,h|d-1}^{-1}$ of the marginal predictive distribution, the realization $P_{d,h}$ at hour h in day d can be expressed as [68]

$$P_{d,h} = \hat{F}_{d,h|d-1}^{-1}(\Phi(Z_{d,h})), \quad h=1,2,\dots,24 \quad (4.5)$$

This transformation process can be illustrated by Fig. 4.9, where the arrows indicate the inverse transformation from standard normal random variable $Z_{d,h}$ to the prediction realization $P_{d,h}$.

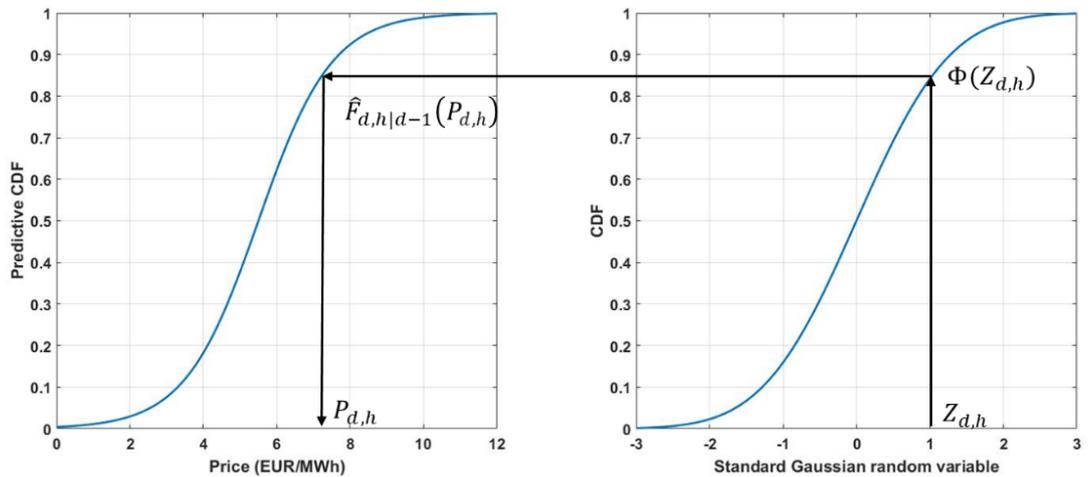


Fig. 4.9 Schematic diagram of inverse transformation for a realization in predictive margins.

The critical issue of this approach is the determination of the covariance structure \mathbf{R}_d for the multivariate random variable \mathbf{Z}_d , which informs the interdependence of look-ahead horizons. A recursive estimation approach was applied to adaptively estimate the covariance structure in literature [34]. But it suffers from the lack of theoretical backgrounds to define analytically or numerically an optimal forgetting factor for tracking the type of covariance matrices. Further, [174] used an exponential function to express the covariance structure, in which the optimal range parameter was determined by the consistency between the variability of measured wind power and the generated scenarios. However, the range parameter needs to be examined in an enumerative way before each forecast, which is computationally expensive.

In this study, we assume that the intercorrelation of multivariate Gaussian random vector \mathbf{Z}_d is consistent with that of the transformed multivariate vector $\tilde{\mathbf{Z}}_K^\phi = [\tilde{Z}_{K,1}^\phi, \dots, \tilde{Z}_{K,h}^\phi, \dots, \tilde{Z}_{K,24}^\phi]$ of the observed probabilistic forecasts series in the past K similar weekdays, the matrix form of vector $\tilde{\mathbf{Z}}_K^\phi$ can be expressed as

$$\mathbf{Z}_K^\phi = \begin{bmatrix} z_{1,1}^\phi & \cdots & z_{1,24}^\phi \\ \vdots & z_{k,h}^\phi & \vdots \\ z_{K,1}^\phi & \cdots & z_{K,24}^\phi \end{bmatrix} \quad (4.6)$$

where ϕ is the weekday dummy, $z_{k,h}^\phi = \Phi^{-1}(\hat{F}_{k,h|k-1}(p_{k,h}))$, $k = 1, 2, \dots, K$, is the transformed value of the probabilistic forecast $\hat{F}_{k,h|k-1}(p_{k,h})$ for hour h in similar weekday k through inversed standard normal CDF Φ^{-1} . This process can be also illustrated in Fig. 4.9 by following opposite direction of the arrow marked. $p_{k,h}$ is the real MCP at hour h in day k , and $\hat{F}_{k,h|k-1}$ is the corresponding forecasted CDF issued at its previous day.

After deriving the Gaussian transformed matrix of the observed probabilistic forecasts in (4.6), its temporal interdependence can be informed by the Pearson correlation coefficient between hour n and hour m

$$\tilde{\rho}_{n,m}^{\phi,K} = \frac{\text{cov}(\tilde{Z}_{K,n}^{\phi}, \tilde{Z}_{K,m}^{\phi})}{\sigma_{\tilde{Z}_{K,n}^{\phi}} \sigma_{\tilde{Z}_{K,m}^{\phi}}}, \quad n, m = 1, 2, \dots, 24 \quad (4.7)$$

where $\text{cov}(\cdot)$ and σ denote the covariance and standard deviation, respectively.

Likewise, we can write the correlation coefficient of multivariate Gaussian random vector \mathbf{Z}_d

$$\rho_{n,m}^d = \frac{\text{cov}(Z_{d,n}, Z_{d,m})}{\sigma_{Z_{d,n}} \sigma_{Z_{d,m}}}, \quad n, m = 1, 2, \dots, 24 \quad (4.8)$$

which is considered equal to the empirical correlation coefficient $\tilde{\rho}_{n,m}^{\phi,K}$. Note that $\sigma_{Z_{d,n}} = \sigma_{Z_{d,m}} = 1$, we can quickly calculate the covariance between each pair of

look-ahead times by $\text{cov}(Z_{d,n}, Z_{d,m}) = \tilde{\rho}_{n,m}^{\phi,K}$, thus yielding its covariance matrix \mathbf{R}_d .

In this sense, without introducing any extra hypothetic parameters or involving any recursive or enumerative process, the proposed empirical correlation based dependence structure determination method could significantly decrease the computational cost as compared to traditional ones.

4.5 Results

The whole dataset described in Section 4.2 is divided into four parts. The explanatory analysis is conducted using the dataset in year 2013. The dataset within the period from 1st Jan 2014 to 30th Jun 2014 is used to determine the structure of individual ELM, while EMOS's training period and ensembles' dimensionality are

determined by the samples from 1st Jul 2014 to 31st Jan 2015. The out-of-sample test period is from 1st Feb 2015 to 31st Jul 2016.

4.5.1 Determination of ensembles' dimensionality and EMOS's training periods

For each ensemble member, an optimal structure should be determined beforehand to guarantee the minimum uncertainty arising from the individual ELM structure. The k -fold cross-validation is used to identify such structure using the samples from 1st Jan 2014 to 30th Jun 2014. Then, multiple ELMs dressed with this structure but different initial parameters are combined to produce ensemble forecasts of MCP.

The dimensionality of ensembles should also be examined on the account of the efficiency of the subsequent optimization. The EMOS model is fitted in a time-adaptive way with short training periods, which allows for a rapid adaption to changes in environmental conditions while longer training periods reduce the statistical variability in parameter estimation [64].

Fig. 4.10 depicts the forecasting skill indicated by average CRPS values tested from 1st Jul 2014 to 31st Jan 2015 by varying the training sample size ($T = 1, 2, \dots, 20, 30, \dots, 60$ days) and number of ensemble members ($m = 2, 3, \dots, 20, 30, 40, 50$). As can be seen clearly from Fig. 4.10, extremely short training periods yield inferior performance, so do the larger length of training samples. The optimal training period is located at around 14 days. No obvious trend is discovered in the performance with respect to the number of ensembles. Given the computational burden of optimization,

a small number of ensembles is preferred. In this case, the minimum average CRPS value comes at $T = 14$ days and $m = 13$.

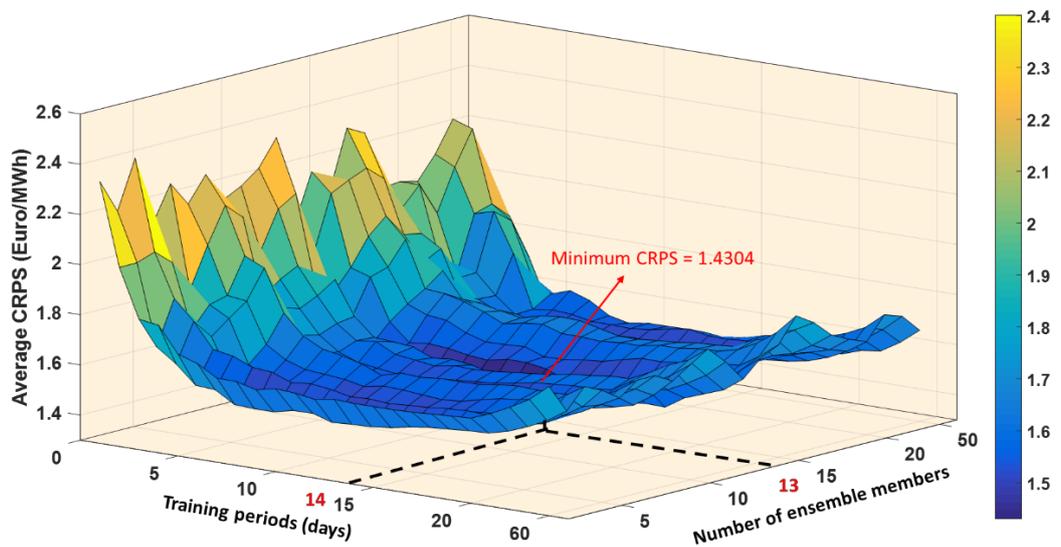


Fig. 4.10 Forecasting skill over period from 1st Jul 2014 to 31st Jan 2015 in terms of different training days and dimensionalities of ensembles.

4.5.2 Discussion of the predictive density via ELC-EMOS

With the well-trained EMOS model, the density of each hour for next-day can be estimated independently. Fig. 4.11 shows the 24 hourly predicted density curves of MCPs on 28th Jul 2016 in a 3D view.

Generally, the price variation perceives a nature pattern like the demand, which is much predictable. Somehow, in Nordic system, with increasing wind energy and frequent congestions occurring at the interconnectors [175], the variation doesn't always show cyclic property but involves fluctuations and spikes with various magnitudes, posing great challenges to forecasters. In order to evaluate the performance of the proposed model from a global perspective, three weekly scenarios are considered in this thesis, i.e., (1) weeks with normal trend, (2) weeks with small

spikes and (3) weeks with large spikes. The number of weeks belonging to each scenario during the test period is listed in TABLE 4.2. It should be noted here, as ELM and EMOS are fitted by the samples of the lagged 84 days and 14 days in a rolling way, respectively, the density forecast is available from 10th May 2016. The time-adaptive fitting strategy is illustrated in Fig. 4.12.

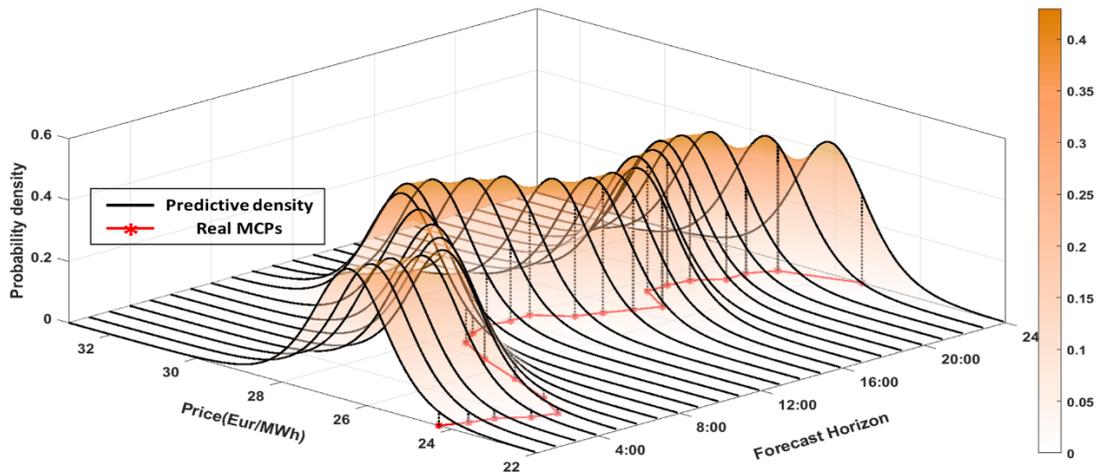


Fig. 4.11 3D visualization of predictive densities of Swedish MCPs on 28th Jul 2016

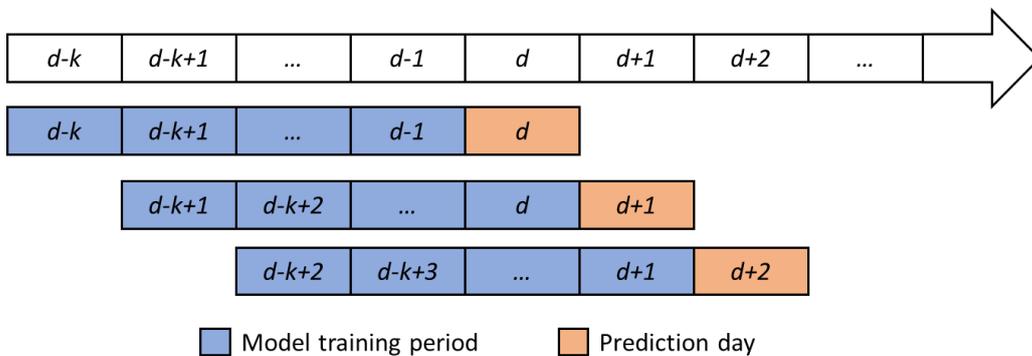


Fig. 4.12 Time-adaptive model fitting strategy

Typical price series of each weekly pattern associated with the prediction intervals (PIs) extracted from the Logistic predictive distribution is illustrated in Fig. 4.13. In the first block, a normal and stable trend is observed during the entire week,

the resultant PIs are extremely sharp accordingly. In the second block, the series becomes more volatile with some unexpected spike occurred on the second day of the week, but far less severe, so the forecasting skill over the entire week is acceptable as most observations fall into the PIs with satisfactory sharpness except for the spike. When the spikes emerge with large magnitude (as shown in the last block), the PIs are unable to cover most of the prices, particularly for the huge spikes. The sharpness turns to be rather poor, revealing the large uncertainty involved in this situation. This can be further reflected by the average CRPS value with regard to each weekly scenario in Table 4.2. Stationary price series appears in nearly half of the test period with the lowest average CRPS value (1.0054 EUR/MWh). When rare small spikes occur, performance of the predictive density degrades, giving a larger CRPS around 2 EUR/MWh. The worst case comes in the weeks with significant spikes, yielding a 330% decline of the overall forecasting skill as compared with that in weeks with normal trend.

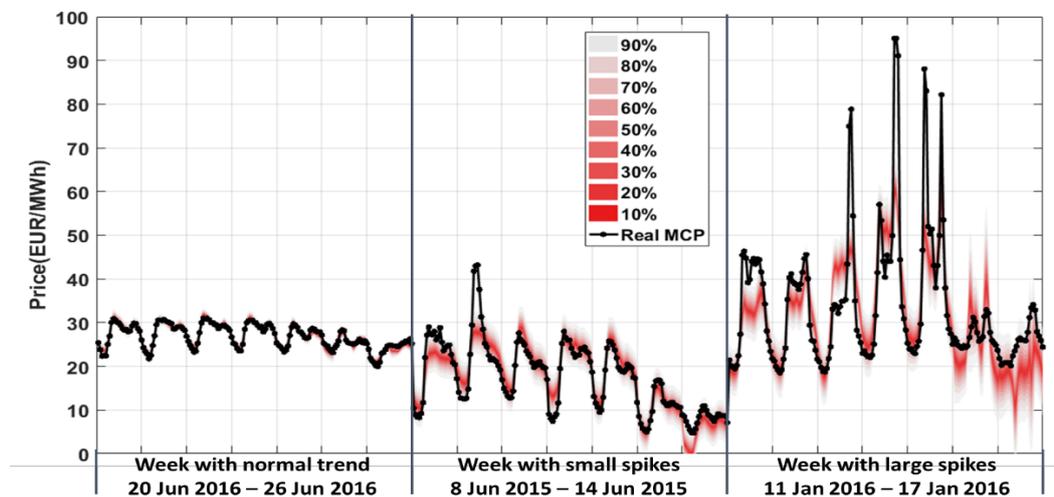


Fig. 4.13 MCP series with prediction intervals in weeks with different price patterns.

The fundamental reasons behind the poor forecasting performance at huge price spikes were studied comprehensively in [176]. From the economic perspective, high price volatility occurs in the regions where the demand and supply are very inelastic. In this sense, models that simply depend on the past price observations and forecasted supply/demand are not so favorable for predicting the large spikes. The authors in [176] have showed that the spikes can be well detected if the underlying mechanic of the price process could be obtained in advance. They utilize the structural approaches to learn from the bidding behavior itself (i.e. forecasting the supply and demand curves) to accommodate forecast with large spikes [176]. In this thesis, extreme price events are not focused due to the lack of real auction information and other related processes.

Table 4.2 Average CRPS (EUR/MWh) of Normal- and Logistic- EMOS Models in the Context of Different Weekly Scenarios

Weekly Pattern	Number of weeks	Gaussian based model	Logistic based model
Weeks with normal trend	33	1.0070	1.0054
Weeks with small spikes	18	1.9992	1.9890
Weeks with large spikes	14	4.3090	4.2944

4.5.3 Comparison with Gaussian EMOS model

The well-known Gaussian EMOS model is tested in comparison with the Logistic model. Full analytic expressions of CRPS for Gaussian model is well established in [64]. From the results in Table 4.2, the average CRPS values of two

distributions show minor difference, whereas the Logistic distribution yields a slight improvement over the Gaussian one in all three weekly scenarios. Given the improvement of average CRPS is imperceptible, the significance test needs further implementation.

Diebold-Mariano (DM) test [177] allows to assess the statistical significance of the CRPS differences between two models. Denote the loss differential series as

$$d_i = CRPS_1(\hat{F}_{1,p_i}, p_i) - CRPS_2(\hat{F}_{2,p_i}, p_i) \quad (4.9)$$

where $CRPS_n(\cdot)$ is the CRPS value for model n . Then the testing statistic is computed as

$$t_N = \sqrt{N} \mu_{d_i} / \sigma_{d_i} \quad (4.10)$$

where N is the length of the testing period, μ_{d_i} and σ_{d_i} are the sample mean and standard deviation of series d_i , respectively. Under the null hypothesis of equal predictive performance, i.e. $\mathbf{E}(d_i)=0$, t_N follows a standard normal distribution. Negative values of t_N indicate a better forecasting skill of model 1, whereas model 2 is preferred in case of positive values of t_N . The statistical significance of the test statistic can be assessed by calculating the corresponding p -values (probability) under the null hypothesis, values that are significant at the 0.05 level are marked with ‘*’. Table 4.3 summarizes the DM testing results of the forecasts via Logistic-EMOS and Gaussian-EMOS.

As clearly seen from Table 4.3, all DM testing statistics t_N are negative, implying that the better predictive skill can be obtained in the context of Logistic density for all three scenarios. However, the p value in weeks with normal trend is greater than the significance level of 0.05, showing the equal performance for both models in this

situation. In contrast, in weeks with price spikes, the proposed Logistic-EMOS significantly outperforms the Gaussian-EMOS in terms of CRPS.

Table 4.3 Test Statistics and Probability of DM Test of Equal Forecast Performance for Comparison of Logistic-EMOS and Gaussian-EMOS
 ‘*’ Symbols Mark Statistically Significant Difference in CRPS

	Weeks with normal trend	Weeks with small spikes	Weeks with large spikes
t_N	-1.50	-2.81	-3.04
p	0.133	0.0049*	0.0024*

Another assessment is performed with regard to the predictive reliability. For density forecast, the reliability is usually indicated by the absolute probabilistic deviation (APD) through examining finite quantiles, which can be written as

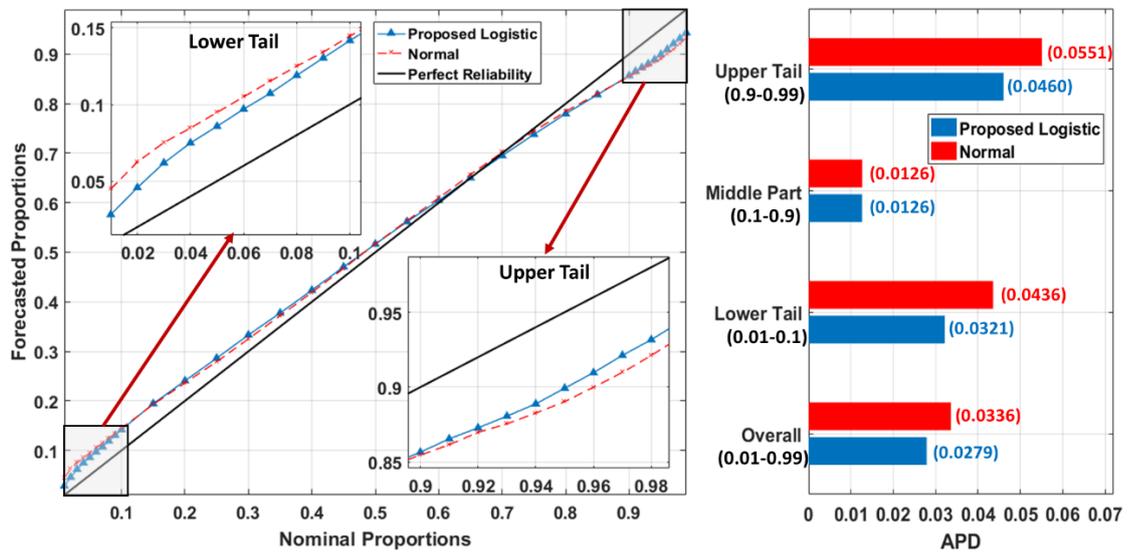
$$APD = |\tau - \hat{\tau}| \quad (4.11)$$

$$\hat{\tau} = \frac{1}{T_p} \sum_{d=1}^{T_p} \sum_{h=1}^{24} \mathbf{1}(p_{d,h} \leq \hat{F}_{d,h|d-1}^{-1}(\tau))$$

where T_p is the length of verifying horizon, $\tau \in (0,1)$ is the nominal quantile level of interest and $\hat{F}_{d,h|d-1}^{-1}(\tau)$ denotes the predictive quantile of level τ for hour h in day d issued at the last day $d-1$. It can be seen from (4.11) that high reliability can be obtained when APD turns to zero.

Since the performance of distribution tail is paramount in this case, the whole predictive distribution is evaluated by three parts, the lower tail ($\tau=0.01 \sim 0.1$, in steps of 0.01), the middle part ($\tau=0.1 \sim 0.9$, in steps of 0.05) and the upper tail ($\tau=0.9 \sim 0.99$, in steps of 0.01). Fig. 4.14 interprets the predictive proportions and APD values at all evaluated quantiles for Normal and Logistic predictive distribution. As noticed Fig. 4.14 (a), both models perform poorly towards the distributional ends. It is particularly

true for Normal distribution (the red dashed line in Fig. 4.14 (a)). Fig. 4.14 (b) displays the APDs at different quantiles. The Logistic distribution achieves the improvement of 19.78% and 35.83% APDs at upper and lower tail over Normal model, respectively. Both models show higher reliability at middle quantiles (0.1-0.9), sharing the same average APDs of 0.0126. Lastly, by averaging all APDs over the evaluated quantiles, a 20.43% improvement over the Normal model can be obtained, indicating the Logistic model is capable of producing more reliable quantiles.



(a) Reliability diagram

(b) APDs of different quantiles

Fig. 4.14 Quantile reliability diagram for the Normal and Logistic models.

4.5.4 Comparison with benchmarks of density forecasts

The last comparative study is carried out against five benchmarks in the framework of density forecast, whose technical details are given as follows,

4.5.4.1 Empirical unconditional density forecast (EU)

In this model, the predictions within each lead time are simply assumed to obey a constant empirical distribution by fitting the past observations of the same look-ahead time,

$$\hat{F}_{d,h|d-1} = F_h, \quad h=1,2,\dots,24 \quad (4.12)$$

where $\hat{F}_{d,h|d-1}$ denotes the predictive CDF of hour h for day d issued in its previous day $d-1$, F_h represents the empirical CDF of the past MCP observations at hour h . We used the past one year's samples to fit F_h . Recall the empirical quantile decomposition [67] of CRPS in (2.15), which allows us to examine the finite quantiles.

4.5.4.2 Multivariate kernel density estimator (KDE)

The general formulation of this model is expressed as [101],

$$f_{P_{d,h}|\mathbf{X}_d} = \frac{f_{P_{d,h},\mathbf{X}_d}}{f_{\mathbf{X}_d}} \quad (4.13)$$

which intends to compute the conditional PDF of the variable $P_{d,h}$ for hour h in day d , given the available information \mathbf{X}_d up to day d . The key task of this approach is to estimate the underlying marginal and the joint PDFs using the kernel density estimator [178], the conditional PDF of MCP for hour h in day d evaluated at price p thus can be expressed by

$$\begin{aligned} \hat{f}_{P_{d,h}|\mathbf{X}_d}(p, \mathbf{x}) &= \frac{1}{\lambda} \sum_{i=1}^N \omega(\mathbf{x}, \mathbf{x}_i) \psi\left(\frac{p-p_i}{h}\right) \\ \omega(\mathbf{x}, \mathbf{x}_i) &= \frac{\psi(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{x}_i))}{\sum_{j=1}^N \psi(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{x}_j))} \end{aligned} \quad (4.14)$$

where N is the number of samples considered; ψ is the multivariate Gaussian kernel function; \mathbf{H} and λ are the smoothing matrix and smoothing parameter used to control the smoothness of conditional variables \mathbf{x} and the resultant predictive distribution,

respectively. p_i is the MCP corresponding to explanatory vector \mathbf{x}_i and p is the evaluated point.

4.5.4.3 Gradient boost machines based quantile regression with exponentially distributed tails (GBM-QRE)

Quantile regression is another classic technique that doesn't require any prior knowledge of the distribution shape. Here, the gradient boosting machines (GBM) [179] is employed to estimate the conditional quantiles ($\tau=0.05\sim 0.95$, in steps of 0.01) separately. The core idea of this powerful algorithm is that it builds the model in a stage-wise fashion and then generalizes them by allowing optimization of an arbitrary differential loss function. By following the detailed algorithm interpreted in [179], the estimation of the τ -th quantile function u_k^τ at k -th iteration can be formulated as

$$\begin{aligned} \hat{u}_k^\tau &\longleftarrow \hat{u}_{k-1}^\tau + \rho_k h(\mathbf{X}, \theta_k) \\ (\rho_k, \theta_k) &= \arg \min_{\rho, \theta} \sum_{i=1}^N L(p_i, \hat{u}_{k-1}^\tau) + \rho h(\mathbf{x}_i, \theta) \end{aligned} \quad (4.15)$$

where ρ_k is the step-size at each iteration, $h(\mathbf{x}_i, \theta)$ is the base-learner function, which heavily influences the properties of the GBM model and needs to be specified beforehand. In order to capture interaction between variables in a computationally-feasible way, the decision-trees is used as the base-learner function here. $L(\cdot)$ is the pinball loss function given by

$$L(p_i, u^\tau(\mathbf{x}_i)) = (\mathbf{1}(p_i < u^\tau(\mathbf{x}_i)) - \tau)(u^\tau(\mathbf{x}_i) - p_i) \quad (4.16)$$

where \mathbf{x}_i is the vector of explanatory variables and p_i is the MCP.

GBM aims to estimate the parameters in a gradient descent way for the loss function as

$$g_k(\mathbf{x}_i) = \left[\frac{\partial L(p_i, u^\tau(\mathbf{x}_i))}{\partial u^\tau(\mathbf{x}_i)} \right]_{u^\tau(\mathbf{x}_i) = \hat{u}_{k-1}^\tau(\mathbf{x}_i)} \quad (4.17)$$

Finally, (4.15) can be rewritten as

$$(\rho_k, \theta_k) = \arg \min_{\rho, \theta} \sum_{i=1}^N [-g_k(\mathbf{x}_i) + \rho h(\mathbf{x}_i, \theta)]^2 \quad (4.18)$$

Considering that a robust quantile regression model is always coupled with a justified modeling process for the extremes, the exponential distribution is employed here to model the tails beyond the last quantiles ($\tau=0.05, 0.95$) estimated via GBM. The detailed modeling process of exponential distributed tails for extreme electricity prices can be found in [160]. The overall forecasting skill of this GBM-QRE model is then verified over 99 quantiles, including the center quantiles (0.05-0.95 in steps of 0.01) estimated by GBM and additional tail quantiles (0.01-0.04 in steps of 0.01 and 0.96-0.99 in steps of 0.01).

4.5.4.4 GARCH

Gaussian distribution attempts to quantify the uncertainties arising from forecasting errors in GARCH model [160], which is formulated by

$$p_{d,h} = \hat{p}_{d,h|d-1} + \mathbf{z}_{d,h} \sigma_{d,h} \quad (4.19)$$

where $\sigma_{d,h}$ is the standard deviation of error $\varepsilon_{d,h} = p_{d,h} - \hat{p}_{d,h|d-1}$. $\mathbf{z}_{d,h}$ is a sequence of independent and identically distributed (i.i.d) standard Gaussian variable, which translates the error to be a Gaussian variable $\varepsilon_{d,h} \sim \mathcal{N}(0, \sigma_{d,h}^2)$.

The classical GARCH process of order (p, q) for estimating the conditional variance is established as

$$\sigma_{d,h}^2 = \alpha_{d,h}^{(0)} + \sum_{i=1}^q \alpha_{d,h}^{(i)} \varepsilon_{d-i}^2 + \sum_{j=1}^p \beta_{d,h}^{(j)} \sigma_{\varepsilon, d-j}^2 \quad (4.20)$$

where $\alpha_{d,h}^{(0)}$ is constant, $\alpha_{d,h}^{(i)}$ is the ARCH parameter and $\beta_{d,h}^{(j)}$ is the GARCH parameter. The GARCH(1,1) specification is adopted. Besides, to achieve good regressive performance, the model is fitted separately for each forecast horizon h . The point forecast series is also obtained by ELM.

4.5.4.5 Maximum likelihood estimation for EMOS (MLE-EMOS)

The alternative way to estimate the statistical parameters in (4.2) is maximum likelihood [64]. The ignorance score (IGN) is commonly used in such problems, which is defined as the negative of the logarithm of the predictive density \hat{f} (likelihood function) at the verifying value p ,

$$\text{ign}(\hat{f}, p) = -\log \hat{f}(p) \quad (4.21)$$

In case of the logistic predictive PDF with location $\mu_{d,h}$ and scale $\varsigma_{d,h}$, the following expression is derived,

$$\begin{aligned} \text{IGN}(\mathcal{L}(\mu_{d,h}, \varsigma_{d,h}), p_{d,h}) = \\ (p_{d,h} - \mu_{d,h}) / \varsigma_{d,h} + \ln \varsigma_{d,h} + 2 \ln(1 + e^{-(p_{d,h} - \mu_{d,h}) / \varsigma_{d,h}}) \end{aligned} \quad (4.22)$$

The IGN is also a negatively oriented score like CRPS. As a result, the EMOS coefficients $\alpha_{d,h}, \beta_{d,h}$ are estimated by

$$(\alpha_{d,h}, \beta_{d,h}) = \underset{\alpha_{d,h}, \beta_{d,h}}{\text{argmin}} \left(\frac{1}{T} \sum_{d=1}^{d-T} \sum_{h=1}^{24} \text{IGN}(\mathcal{L}(\mu_{d,h}, \varsigma_{d,h}), p_{d,h}) \right) \quad (4.23)$$

4.5.4.6 Comparative results and discussion

The comparative results in terms of monthly average CRPS during the verifying period are summarized in Table 4.4. Among the benchmarks, not surprisingly, EU shows the worst skill due to its persistent nature, which totally depends on the past observations of MCPs. KDE and GBM-QRE significantly improve the predictive

performance over EU by 63.9% and 68.8%, respectively, since they both incorporate the abundant explanatory information into the statistical models. The GARCH model gives the better average CRPS than the former three benchmarks regardless of its parametric properties. Both EMOS-based models outperform other benchmarking models, with around 14% improvement over the GARCH model, this can be explained by the rationale of the underlying models: the density estimated via GARCH is conditional on a single point prediction without calibration, and the conditional noise variance is estimated separately via a heteroscedastic model. In contrast, EMOS-based models are conditional on multiple estimates, both predictive mean and variance can be well calibrated by the optimum score. In the monthly perspective, the ELC-EMOS model can lead best forecasts in 12 months out of 15 months, even in the highly volatile month (Jan 2016), showing its robustness in dealing with various situations. However, by comparing ELC-EMOS model and MLE-EMOS model (the last two columns in Table 4.4), it can be found that the two models struggle to distinguish which is the better for MCP density forecast due to the minor difference between them. In light of this, we performed the DM test of equal forecast performance to benchmark the ELC-EMOS model in terms of CRPS. The results are given in Table 4.5.

It can be observed in Table 4.5 that the lowest t_N occurs in the EU model, which is followed by KDE, GBM-QRE and GARCH model successively, which confirms the outcomes reported in Table 4.4. The MLE-EMOS model comes with a highest t_N of -4.21 over the underlying test period. But its corresponding probability is still far

less than 0.05, indicating that the improvement over MLE-EMOS can be considered statistically significant.

Table 4.4 Average CRPS (EUR/MWh) against benchmarking models on a monthly Basis. Smallest CRPS values are marked in bold

Month	EU	KDE	GBM-QRE	GARCH	MLE-EMOS	ELC-EMOS
May 2015	6.3638	2.0407	2.4361	2.1766	2.0492	2.0395
Jun 2015	12.1296	2.8042	3.4128	2.4638	2.4204	2.4013
Jul 2015	18.8402	1.5446	1.8136	1.4233	0.9238	0.9152
Aug 2015	13.5270	4.3783	3.9751	3.2385	3.0397	3.0878
Sep 2015	8.7320	3.6494	3.4344	2.9301	2.8016	2.7927
Oct 2015	7.1295	3.1223	2.0465	2.8363	2.4705	2.4818
Nov 2015	4.8133	3.1117	3.3250	2.9163	1.9156	1.9030
Dec 2015	10.1976	2.8726	4.2576	2.1541	1.9768	1.9192
Jan 2016	10.3806	8.1768	5.5380	6.1248	6.1836	5.7863
Feb 2016	8.6886	2.7563	1.2686	2.0761	1.5050	1.5123
Mar 2016	6.6276	3.5239	2.1651	1.8352	1.3573	1.3317
Apr 2016	6.2289	2.678	1.3873	1.6557	0.9315	0.9199
May 2016	5.3685	1.9385	1.5451	1.197	1.1602	1.1505
Jun 2016	2.8358	1.3257	1.6454	0.8931	0.7244	0.7238
Jul 2016	3.5280	1.4000	0.8247	0.8517	0.7534	0.7443
Average	8.3594	3.0215	2.6050	2.3182	2.0142	1.9806

Table 4.5 Test Statistics and Probability of DM Test of Equal Forecast Performance for Comparison of ELC-EMOS and Benchmarks.

	vs EU	vs KDE	vs GBM-QRE	vs GARCH	vs MLE-EMOS
t_N	-101.36	-27.67	-22.16	-18.09	-4.21
p	0*	0*	0*	0*	2.55E-05 *

By further looking into the average CRPS as a function of the look-ahead horizons, as visualized in Fig. 4.15, the skill generally follows the typical intraday MCP pattern. Satisfactory performance is observed in night and early morning, while it deteriorates sharply from 7:00, and reaches the highest CRPS at 9:00. A ‘valley’ shape is noticed during the working hours, followed by another peak CPRS occurring at 18:00.

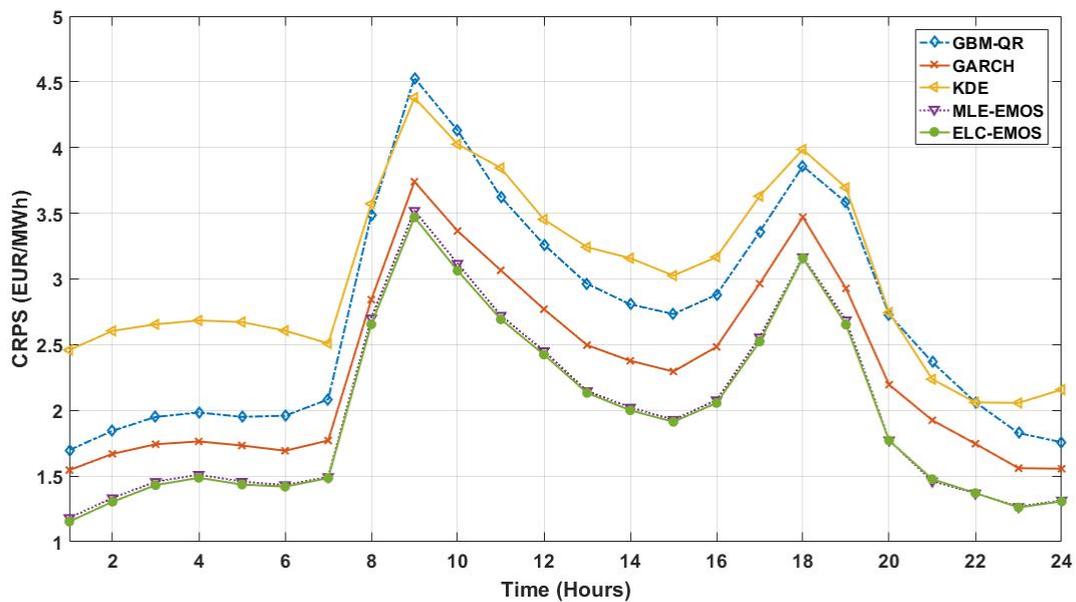


Fig. 4.15 Average CRPS in terms of 24 hours

Table 4.6 Testing for Quantile Reliability by Averaging APDs over Quantiles From 0.01 to 0.99 with 0.1 Increment

	EU	KDE	GBM-QRE	GARCH	MLE-EMOS	ELC-EMOS
Avg. APD	0.3893	0.1071	0.0817	0.0284	0.0188	0.0273

Finally, the quantile reliability test is conducted for all models, 99 quantiles ranging from 0.01 to 0.99 with increment of 0.01 are verified using (4.11). Table 4.6 summarizes the average APDs of each model. Apparently, poor reliability is noticed in EU, KDE and GBM-QRE models. In contrast, GARCH, MLE-EMOS and ELC-EMOS models give more acceptable average APDs over the entire test period. The MLE-EMOS model exhibits the highest overall reliability therein, whereas ELC-EMOS and GARCH model is slightly inferior to the MLE-EMOS, which is inconsistent with the results of CRPS. This is mainly due to the nature of scoring rules, that is, IGN score (maximum likelihood) gives harsher penalties to the poor probabilistic forecast to pursue a higher reliability. Nonetheless, given the worse overall performance (higher CRPS) achieved by GARCH and MLE-EMOS as compare to ELC-EMOS, the sharpness of both ML-oriented methods must be poorer, as we know, predictive density with unsatisfied sharpness can scarcely provide the end-users with valuable information.

One may argue that the proposed ELC-EMOS model seems a little bit complex as it requires extra point estimates while nonparametric approaches, such as KDE and QR, can directly produce the density without deterministic estimates and pre-assumption of distribution shape. However, from the results reported above, the parametric models (GARCH, MLE-EMOS, ELC-EMOS) conditional on the point forecasts outperform the nonparametric ones regarding to the predictive performance of full distribution, particularly for the models that the multiple potential forecasts are dressed with proper distribution and the desired scoring rule is used.

4.5.5 Day-ahead scenarios generation and evaluation

In this sub-section, by using the dependency modeling strategy introduced in Section 4.4, we present the derived day-ahead price scenarios based on the well calibrated predictive densities. The empirical correlation coefficients are determined by the observed probabilistic forecasts in the similar weekdays over the past 6 months. The quality of forecasted scenarios are verified against four rivals, including the uncalibrated ensemble forecasts, independent sampling, Gaussian copula approaches based on empirical covariance [34] and exponential covariance structure [174], respectively. Energy score (ES) and variogram score (VS) [69] are used to evaluate the multivariate skill. The verification period is from 1st Nov 2015 to 31st Jul 2016, and $J = 1000$ scenarios are generated for each verifying day.

4.5.5.1 Uncalibrated ensemble forecasts

In a natural way, the raw forecasts generated by ensemble ELMs can be regarded as the predicted scenarios. To be specific, each time trajectory is represented by the deterministic forecasts via one individual ELM, aggregating a large number of ELMs would be a straightforward way to mimic the potential scenarios.

4.5.5.2 Independent sampling

Independent sampling benchmark, as its name indicates, is introduced to uniformly sample from the same marginal predictive distributions obtained by the proposed ELC-EMOS model but neglects the temporal dependency.

4.5.5.3 Gaussian copula with recursive empirical covariance (GC-REC)

This classic method introduced by Pinson et al. [34] generally relies on the adaptive estimation of the interdependence structure by an exponential smoothing scheme. At time t , the covariance structure is recursively updated with

$$\mathbf{R}_t = \lambda \mathbf{R}_{t-1} + (1-\lambda) \mathbf{X}_{t-H} \mathbf{X}_{t-H}^T \quad (4.24)$$

where

$$\mathbf{X}_{t-H} = \left[\Phi^{-1}\left(\widehat{F}_{t-H+1|t-H}(p_{t-H+1})\right), \dots, \Phi^{-1}\left(\widehat{F}_{t|t-H}(p_t)\right) \right]^T \quad (4.25)$$

is the vector of past observations transformed through the probabilistic forecasts series issued at time $t-H$, and then through the probit function Φ^{-1} . The initial covariance structure is by setting all its off-diagonal elements to 0 and its diagonal elements to 1. λ is the forgetting factor, $\lambda \in [0,1)$. However, the main deficiency of this approach is the lack of theoretical backgrounds to define analytically or numerically an optimal forgetting factor [174].

4.5.5.4 Gaussian copula with exponential covariance (GC-EC)

Another prevailing way to derive the covariance is based on the exponential function [174], which can be written as

$$\text{cov}(Z_{d,n}, Z_{d,m}) = \exp\left(-\frac{|n-m|}{\varepsilon}\right), \quad 0 \leq n, m \leq 24 \quad (4.26)$$

where ε is the range parameter controlling the strength of the correlation of normal random variables among set of look-ahead horizons in day d . By further introducing an indicator I_ε to measure the difference in the statistical characteristics of the variability between the observations and generated scenarios, we had to examine the different values of I_ε by varying ε and eventually determine the optimal range

parameter at the minimum I_ϵ . Obviously, this enumerative way will give rise to large computational cost.

4.5.5.5 Comparative results and discussion

ES has a good discriminating ability to evaluate forecasts relying on marginals with correct variances but biased means, but it fails to detect the misspecified dependence structures between elements of a multivariate quantity. VS is introduced as a proper score to offset the deficiency of ES by considering the pairwise differences of the elements of a multivariate quantity. Therefore, two scores are combined to give a justified assessment of the generated time trajectories. Following (2.16) and (2.17) presented before, the ES and VS for day d in this case can be written as

$$ES_d = \frac{1}{J} \sum_{j=1}^J \|\mathbf{p}_d - \mathbf{s}_d^{(j)}\|_2 - \frac{1}{2J^2} \sum_{i=1}^J \sum_{j=1}^J \|\mathbf{s}_d^{(i)} - \mathbf{s}_d^{(j)}\|_2 \quad (4.27)$$

$$VS_d^l = \sum_{m=1}^{24} \sum_{n=1}^{24} w_{m,n} \left(|p_{d,m} - p_{d,n}|^l - \frac{1}{J} \sum_{j=1}^J |s_{d,m}^{(j)} - s_{d,n}^{(j)}|^l \right) \quad (4.28)$$

In this experiment, VS is normalized by setting $w_{m,n}$ as $1/(24*24)$, implying that the significance of all pairwise differences of observations and forecasts are treated as equivalent. l is fixed as 0.5 due to its best discrimination ability [69].

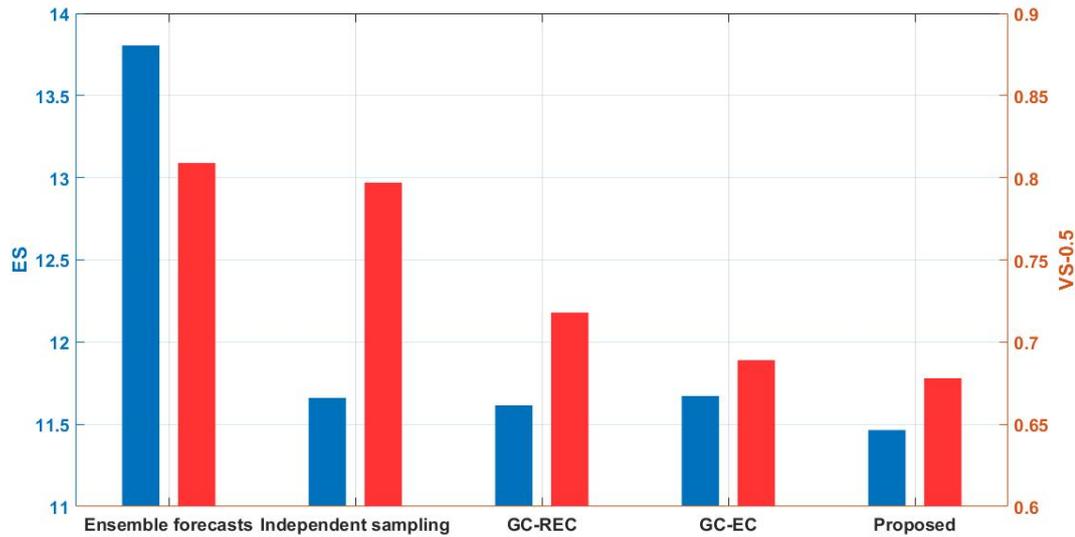


Fig. 4.16 Average ES (EUR/MWh) and VS-0.5 ($[\text{EUR/MWh}]^{0.5}$) of scenarios obtained by different approaches. Blue bar and red bar represent ES and VS-0.5, respectively

Fig. 4.16 visualizes the average ES and VS over all verifying days. Not surprisingly, the scenarios represented by the uncalibrated ensemble forecasts give the worst performance in terms of both ES and VS-0.5, which highlights the need of calibration of the independent forecasts. ES is considerably improved when these forecasts are deployed with probabilistic distribution for each look-ahead time. It deserves noting that the independent sampling shows comparative ES with respect to that of copula-based methods by taking advantage of the well-calibrated marginal probabilistic forecasts. However, when the metric turns to VS, a distinguishable difference is noticed as compared to copula-based methods. This supports the argument in [69] that ES is weak at detecting the miscalibrated dependency structure between the multivariate components. Fig. 4.17 further emphasizes the importance of taking into account the temporal correlations for generating time trajectories. The

similar weekdays for 1st Nov 2015 in the past 6 months are collected to derive the correlation matrix of the transformed gaussian random variables of the observed probabilistic forecasts. As depicted in this figure, there are obvious correlations between in most temporal pairs. Strong temporal dependency appears to emerge in the adjacent hours and decreases as the temporal difference increases. Hence, through establishing a proper dependency structure, the Copula-based approaches notably outperform the independent sampling method in terms of VS. The proposed empirical correlation based approach considering the similar weekdays' patterns is more able to mimic the inherent dependency structure for the prices in different look-ahead horizons, thus leading to a slight improvement over GC-REC and GC-EC.

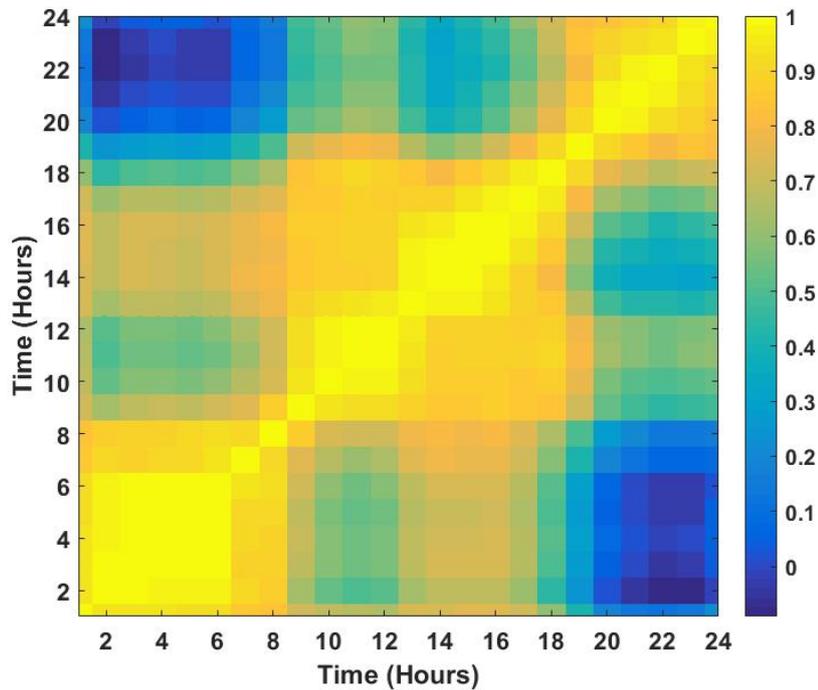


Fig. 4.17 Empirical correlation matrix of the transformed gaussian random variables of the observed probabilistic forecasts for 1st Nov 2015. Highest to lowest correlations are represented by yellow to blue colors.

Given that a small number of representatives are more favored by decision-makers, 20 time trajectories are extracted from the initial set by backward reduction approach [180], as shown in Fig. 4.18, where the three daily patterns (normal trend, small spike and large spike) can be well captured visually by the reduced scenarios with discrepancy, respectively.

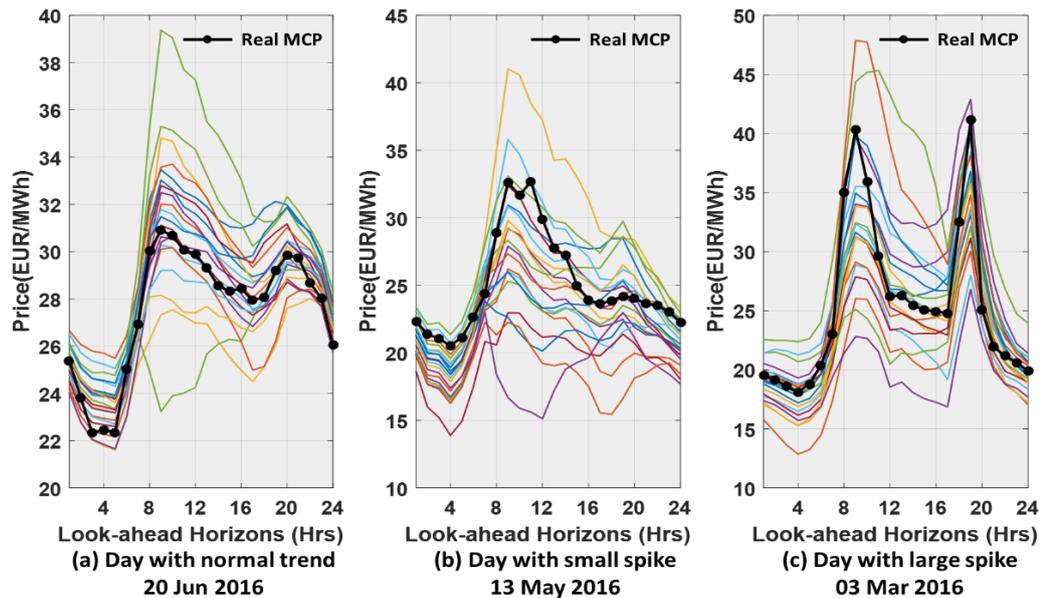


Fig. 4.18 20 representative day-ahead scenarios for different daily pattern.

It should be remarked that a justified multivariate evaluation should be performed by considering several different scores based on different desired properties, since there doesn't exist a single scoring rule that can fulfill all purposes. From a more intuitive and sound perspective, the quality of the estimated scenarios may be assessed by the benefits from their use as input to various decision-making cases [68].

4.6 Summary

An effective density forecast strategy for MCP is proposed in this thesis. In recognizing that the knowledge uncertainty arising from the forecasting models necessitates proper quantifications, the Logistic distribution is proposed to model the potential ensemble forecasts, which is demonstrated to be more robust than normal distribution. Moreover, by comparing with the existing popular density forecast methods, the ELC-EMOS outperforms them in respect of the full predictive distribution assessing by CRPS. Lastly, based on the well calibrated marginal densities, an efficient covariance structure based on the empirical correlation coefficients is developed to quickly generate the highly skillful day-ahead price scenarios without any computational complication.

To further enhance the performance of the proposed hybrid model, future work will be emphasized on three aspects: (1) It is advisable to fit underlying model separately for each forecast horizon [166], and different feature sets need to be selected to accommodate the specific property of each hour. (2) To achieve the better performance at spikes, real auction information will be acquired and incorporated into our model by studying on other markets where the data are available. (3) Instead of ELM, some other powerful forecast engines, such as regularized ELM, random forests, generalized additive models, etc., can be adopted with Logistic-EMOS for further increasing the predictive robustness.

Chapter 5 Conclusions and Future Scope

5.1 Conclusions

The rapid growing penetration of renewable generations into the main grid and the widespread deregulating of electricity markets worldwide forces the system operators and planners to reconsider the mechanisms of their decision-making processes. Diversified uncertainties are brought into each part of the grid unprecedentedly, which requires advanced tools developed to cater for these uncertainties so that the security and economy of power system can be ensured. Over the last two decades, prediction has played a major role in power system, allowing the system operators to implement corresponding actions in advance, e.g., reserve setting, unit commitment, storage sizing, etc. However, the deterministic decision dependent on one plausible forecast appears fairly inadequate in today's market, the market participants are more inclined to evaluate their revenues and risks under all possible scenarios. This highlights the necessity of probabilistic forecasts, which gives a probabilistic description of the potential predictions.

This thesis endeavors to develop some effective approaches to model the potential uncertainties in a modeling system and prognosis them in the output, the contributions are categorized into three distinctive while related types in probabilistic forecasts, which are interval forecast, density forecast and scenario forecast.

In the context of global warming, as the largest renewable energy resource available on our planet, solar energy is increasingly exploited globally, aiming to substitute the traditional fossil fuels integrated into the main grid. Nonetheless, the

extensive installation of PV plants raises a grid integration concern in particular due to its difficulty to dispatch the energy. While storing the PV energy has long been seen as a solution to tackle this problem, a precise forecast of the available energy is necessary for the proper control. Additionally, to manage the viable power output, an accurate PV forecast is also needed to assist in the setting of alternative compensatory controllable resources. However, PV power output forecast is still relatively recent as compared to wind power, particularly in the domain of probabilistic forecasts. This work proposes a novel framework to construct the very short-term PI in a nonparametric way. Inspired by the fundamentals of IGs, the traditional crisp input and model parameters are granulated by intervals, accounting for both input uncertainty and model uncertainty. PI score is used to directly tune the interval parameters so that the optimal PIs can be ensured. The effectiveness of the developed GELM model is verified against benchmarks on a real irradiance measurement station in Hong Kong. It should be noted that this study only develops the generic framework of granule-based AI model to derive the probabilistic forecasts, to further improve the PI skill, this model can be flexibly enhanced in the aspects of input parameters, fitness functions, basis NN, IG forms and granularity levels.

Another initiative to fulfil the vision of Smart Grid [162] is to establish a fully liberalized electricity market. The market deregulation is heating up globally, as the world's largest electricity consumer, China is currently undergoing reforms towards a spot electricity market. In spot market, the generation companies compete to produce and sell electricity through an "auction" instead of a fixed tariff, making their bids to produce a specified quantity of electricity at a particular price in advance.

Therefore, forecasting the electricity price is essential to the revenues of market participants. As reported in [181], a 1% improvement in the mean absolute percentage error (MAPE) in forecasting accuracy would lead to 0.1%-0.35% cost reductions from short-term forecast of electricity price. For a typical medium-scale utility with 5GW peak load, this would be equal to a remarkable savings of \$1.5 million per year [167, 182]. By far, most studies still focus on the point forecast of electricity price, yet the forecasting errors cannot be avoided, and can be rather significant sometimes that bring about considerable economic losses as previously mentioned. Better understanding and proper quantification of the uncertainties in the forecasting process can guide the participants to make more reasonable behaviors, thus maximizing their revenues while considering the associated risks. This study develops an effective density forecasting model to provide a full picture of the potential uncertainties the model may be subject to. CRPS-based EMOS is used to recalibrate the ensemble forecasts from a collection of ELMs, where Logistic distribution is proposed to characterize the potential forecasting errors due to its robustness to the price outliers. The quality is verified against benchmarks over Swedish Elspot in terms of CRPS and Diebold-Mariano (DM) test. In addition, to model the aggregated uncertainties within contiguous lead times, the joint or multivariate distribution is constructed via Gaussian copula through the well-calibrated marginal predictive distribution. In this process, an efficient dependency modeling approach is developed to quickly generate the time trajectories without hypothetical parametrization or enumeration computations. The skill of scenarios is also evaluated against benchmarks to show both superiorities of the derived predictive marginal density and the developed covariance structure.

5.2 Future Scope

This thesis dedicates to provide several effective models to yield the highly skillful probabilistic forecasts, the possible specific enhancement with respect to each model has been given in the end of related chapters. In view of a more generalized prediction framework, the research perspectives are outlined as follows,

1. ***Robust probabilistic forecasts***: In regression problems, the concept of ‘robust’ was originally proposed to ensure the predictions are robust to data missing and perturbations [183, 184]. For the evaluation, in addition to the classic metrics, such as MAPE and RMSE, the worst-case error (WCE) and the standard deviation (STD) are always required to examine the robustness of the results. While in probabilistic forecasts, the resultant performance (e.g., reliability) is also anticipated to spread around the expectation within a small range in the presence of contaminated data. Even though our proposed granule-based regression model has improved the STD of reliability to some extent as compared to other crisp models over different datasets, the data perturbations in different datasets is not taken into account. The relevant work is underway to incorporate the robust criteria into the probabilistic forecasting formulations, with the purpose of constructing a robust model to handle various types of data uncertainty.
2. ***Automated feature selection via Regularized ELM***: In the statistical modeling field, selecting appropriate features is always a crucial and challenging task, which attracts a number of researchers to optimize and standardize this process. Typically, this can be achieved in an ad hoc fashion, which identifies a subset of

predictors that are believed to be influential according to certain metrics (e.g., Chi-square, ANOVA, Pearson's correlation) [163]. Our practices reckoning on correlations to select the predictors exactly belongs to this scope. Another prevalent practice is based on 'wrapper', that is, different subsets of predictors are used to fit different models, the best subset is selected through validating on a test dataset. However, these two methods are limited by the large computational burden in case of high-dimensional explanatory variables. Recent works [134, 164, 166, 167] highly promote the usage of regularization regression models (e.g., LASSO, Elastic net) to fit the full model with all potential variables. Such techniques involve a shrinking (regularization) process which penalizes the coefficients of the regressors shrinking some of them towards zero. In this way, the variables with zero coefficients are excluded from the model, while the rest are retained as the final predictors. Following this idea, the original ELM can be restricted by regularization penalties to avoid the feature selection process. Feasible regularizations on ELM have already been studied, such as L1 penalty (LARS), L2 penalty (Tikhonov regularization) and both of them [168]. Further work is focused on integrating such regularized models into our proposed probabilistic forecasting framework.

3. **Case-oriented probabilistic forecasts:** Most of existing probabilistic forecasts are qualitatively assessed through certain mathematic evaluation tools, which is widely accepted by forecasters or statisticians. However, what the decision makers really care about is the benefits (e.g., profits, security, reliability) by using these forecasts. Hence, the true value of probabilistic forecasts is supposed

to be evaluated by deploying them as inputs to related practical cases, e.g., energy bidding, reserve setting, wind turbine control. To this end, great efforts will be devoted to the study of case-oriented optimal PI or predictive density. On the other hand, apart from reaching the certain foreseen objectives as described previously, possible scenarios or applications that can benefit from these probabilistic forecasts are needed to be further exploited.

Appendix

A. Comparisons between benchmarking models with different input patterns.

Table A. 1 Skills of 90% PI obtained by BS-QR model with past 10-min GHI averages (BS-QR-V1) versus with past 1-min raw data (BS-QR-V2).

Month	BS-QR-V1			BS-QR-V2		
	PICP	PINAW	IS	PICP	PINAW	IS
March	71.80%	13.72%	-4.92%	86.70%	13.58%	-2.65%
April	78.69%	15.58%	-4.24%	75.05%	12.98%	-4.23%
May	73.80%	18.75%	-6.90%	78.00%	16.71%	-5.96%
June	85.90%	25.84%	-5.83%	86.51%	23.35%	-5.08%
July	87.63%	30.33%	-6.52%	85.38%	27.03%	-6.30%
August	93.74%	33.47%	-5.62%	94.43%	31.41%	-4.97%
September	89.64%	31.84%	-5.96%	90.55%	28.88%	-5.32%
October	85.19%	35.68%	-4.72%	82.56%	29.87%	-5.88%
November	81.01%	24.48%	-3.63%	78.11%	24.50%	-5.16%
December	80.50%	21.54%	-3.39%	95.11%	18.98%	-2.34%
Average	82.79%	25.12%	-5.17%	85.24%	22.73%	-4.79%

Table A. 2 Skills of 90% PI obtained by BELM model with past 10-min GHI averages (BELM-V1) versus with past 1-min raw data (BELM-V2).

Month	BELM-V1			BELM-V2		
	PICP	PINAW	IS	PICP	PINAW	IS
March	79.28%	14.72%	-2.83%	90.71%	15.30%	-2.66%
April	78.28%	18.33%	-4.18%	82.63%	19.82%	-4.10%
May	80.74%	22.53%	-5.34%	82.36%	19.20%	-5.45%
June	84.85%	27.38%	-5.69%	87.63%	23.29%	-5.10%
July	83.58%	30.29%	-6.67%	88.51%	27.27%	-6.35%
August	92.08%	32.22%	-5.36%	93.26%	27.25%	-4.85%
September	85.15%	31.62%	-5.73%	87.78%	27.33%	-5.54%
October	85.14%	30.07%	-3.89%	88.22%	26.48%	-3.64%
November	76.97%	22.92%	-3.15%	84.85%	21.49%	-2.86%
December	85.63%	20.34%	-2.56%	90.78%	18.57%	-2.32%

Average	83.17%	25.04%	-4.54%	87.67%	22.60%	-4.29%
---------	--------	--------	--------	---------------	---------------	---------------

Table A. 3 Skills of 90% PI obtained by DIF model with past 10-min GHI averages (DIF-V1) versus with past 1-min raw data (DIF-V2).

Month	DIF-V1			DIF-V2		
	PICP	PINAW	IS	PICP	PINAW	IS
March	87.05%	17.21%	-3.47%	86.46%	24.68%	-5.57%
April	80.10%	25.09%	-7.41%	84.34%	31.66%	-7.53%
May	82.80%	24.42%	-5.95%	83.24%	19.33%	-5.86%
June	87.31%	29.81%	-6.04%	86.77%	21.95%	-5.30%
July	87.39%	30.26%	-7.11%	86.36%	30.43%	-7.05%
August	93.21%	34.77%	-6.16%	94.89%	32.54%	-5.44%
September	89.75%	31.84%	-6.13%	91.11%	31.72%	-5.94%
October	94.13%	33.24%	-4.45%	94.33%	28.58%	-3.95%
November	94.29%	30.41%	-3.76%	92.63%	21.26%	-3.01%
December	92.28%	25.65%	-3.20%	92.86%	29.84%	-3.55%
Average	88.83%	28.27%	-5.37%	89.30%	27.20%	-5.32%

B. Input variable candidates for Swedish MCP

Table B. 1 Input variable candidates for Swedish MCP

Categories	Sub-categories	Candidates
Historical Explanatory Information	Price	Domestic market clearing price (EUR/MWh) System market clearing price (EUR/MWh)
	Demand	Domestic demand (MWh) System demand (MWh) Domestic net demand (MWh) System net demand (MWh)
	Generation	Domestic generation (MWh) System generation (MWh) Domestic wind power generation (MWh) System wind power generation (MWh) Domestic hydro power generation (MWh) Domestic nuclear power generation (MWh) Domestic thermal power generation (MWh) Domestic generation from conventional units (MWh)

		System generation from conventional units (MWh)
	Exchange Power	Import/export power (MWh)
	Weather Info	Temperature ($^{\circ}\text{C}$)
	Reservoir Levels	Domestic reservoir levels (%) System reservoir levels (%)
Prognosis Explanatory Information	Demand	Domestic demand (MWh) System demand (MWh) Domestic net demand (MWh) System net demand (MWh)
	Generation	Domestic generation (MWh) System generation (MWh) Domestic wind power generation (MWh) System wind power generation (MWh) Domestic generation from conventional units (MWh) System generation from conventional units (MWh)
	Exchange Power	Import/export power (MWh)
Weekday Dummies	-	Mon, Tue, ..., Sun

C. Derivation of closed-form analytic expression of CRPS for

Logistic distribution

The quantile score expression for CRPS can be written as:

$$\begin{aligned} \text{CRPS}(F, p) = & 2 \int_0^{F(p)} \tau (p - F^{-1}(\tau)) d\tau \\ & - 2 \int_{F(p)}^1 (1 - \tau) (p - F^{-1}(\tau)) d\tau \end{aligned} \quad \text{C-1}$$

where F is the CDF, p is the observed value and $\tau \in (0, 1)$ is the quantile level.

For Logistic distribution, the τ -th quantile is given by

$$F_{\mathcal{L}(\mu, \varsigma)}^{-1}(\tau) = \mu + \varsigma \text{logit}(\tau) \quad \text{C-2}$$

Suppose $a = F(p)$, $b = p - \mu$. Do the indefinite integral and

substitution for the first term in C-1 we obtain

$$\begin{aligned}
& 2\int \tau(b - \zeta \text{logit}(\tau))d\tau \\
&= b\tau^2 - 2\zeta \int \tau \log(\tau)d\tau + 2\zeta \int \tau \log(1-\tau)d\tau \\
&= b\tau^2 - \zeta\tau^2 \text{logit}(\tau) - \zeta \log(1-\tau) - \zeta\tau
\end{aligned} \tag{C-3}$$

Likewise, the second term of C-1 can be derived as

$$\begin{aligned}
& 2\int (1-\tau)(b - \zeta \text{logit}(\tau))d\tau \\
&= 2b\tau - b\tau^2 - 2\zeta \int (1-\tau) \log(\tau)d\tau + 2\zeta \int (1-\tau) \log(1-\tau)d\tau \\
&= 2b\tau - b\tau^2 + \zeta(1-\tau)^2 \text{logit}(\tau) - \zeta \log(\tau) + \zeta\tau + 2\zeta
\end{aligned} \tag{C-4}$$

Finally, by calculating the results of Eq. C-3 and C-4 on $[0, a]$ and $[a, 1]$,

respectively, and substituting $F(p)$ and $p-\mu$ back, we can obtain the analytical

expression of C-1 as

$$\begin{aligned}
CRPS(F_{L(\mu, \zeta)}, p) &= (p - \mu)(2F(p) - 1) \\
&- 2\zeta F(p) \text{logit}(F(p)) - 2\zeta \log(1 - F(p)) - \zeta
\end{aligned} \tag{C-5}$$

Reference

- [1] J. L. Sawin *et al.*, "Renewables 2017 Global Status Report," 2013.
- [2] J. Twidell and T. Weir, *Renewable energy resources*. Routledge, 2015.
- [3] G. Crabtree *et al.*, "Integrating renewable electricity on the grid," in *AIP Conference Proceedings*, 2011, vol. 1401, no. 1, pp. 387-405: AIP.
- [4] A. Bloom *et al.*, "It's Indisputable: Five Facts About Planning and Operating Modern Power Systems," *IEEE Power and Energy Magazine*, vol. 15, no. 6, pp. 22-30, 2017.
- [5] C. Harris, *Electricity markets: pricing, structures and economics*. John Wiley & Sons, 2011.
- [6] J. Gosens, T. Kåberger, and Y. Wang, "China's next renewable energy revolution: goals and mechanisms in the 13th Five Year Plan for energy," *Energy Science & Engineering*, vol. 5, no. 3, pp. 141-155, 2017.
- [7] S. Voronin, "Price spike forecasting in a competitive day-ahead energy market," *Acta Universitatis Lappeenrantaensis*, 2013.
- [8] T. Jónsson, P. Pinson, H. Madsen, and H. Nielsen, "Predictive Densities for Day-Ahead Electricity Prices Using Time-Adaptive Quantile Regression," *Energies*, vol. 7, no. 9, pp. 5523-5547, 2014.
- [9] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International journal of forecasting*, vol. 30, no. 4, pp. 1030-1081, 2014.
- [10] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38-46, 2015.
- [11] X. Wang, P. Guo, and X. Huang, "A review of wind power forecasting models," *Energy procedia*, vol. 12, pp. 770-778, 2011.
- [12] J. Dobschinski *et al.*, "Uncertainty Forecasting in a Nutshell: Prediction Models Designed to Prevent Significant Errors," *IEEE Power and Energy Magazine*, vol. 15, no. 6, pp. 40-49, 2017.
- [13] S.-E. Fleten and T. K. Kristoffersen, "Stochastic programming for optimizing bidding strategies of a Nordic hydropower producer," *European Journal of Operational Research*, vol. 181, no. 2, pp. 916-928, 9/1/ 2007.
- [14] S.-E. Fleten and T. K. Kristoffersen, "Short-term hydropower production planning by stochastic programming," *Computers & Operations Research*, vol. 35, no. 8, pp. 2656-2671, 2008.
- [15] W. Lei, M. Shahidehpour, and L. Tao, "Stochastic Security-Constrained Unit Commitment," *Power Systems, IEEE Transactions on*, vol. 22, no. 2, pp. 800-811, 2007.
- [16] J. M. Morales, A. J. Conejo, and J. Pérez-Ruiz, "Short-term trading for a wind power producer," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 554-564, 2010.

- [17] D. T. Nguyen and L. B. Le, "Optimal Bidding Strategy for Microgrids Considering Renewable Energy and Building Thermal Dynamics," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1608-1620, 2014.
- [18] P. Kall, S. W. Wallace, and P. Kall, *Stochastic programming*. Springer, 1994.
- [19] J. Ruiwei, W. Jianhui, and G. Yongpei, "Robust Unit Commitment With Wind Power and Pumped Storage Hydro," *Power Systems, IEEE Transactions on*, vol. 27, no. 2, pp. 800-810, 2012.
- [20] X. Wang, "Chapter 8 - Advances in Market Management Solutions for Variable Energy Resources Integration," in *Renewable Energy Integration*, L. E. Jones, Ed. Boston: Academic Press, 2014, pp. 101-116.
- [21] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of operations research*, vol. 23, no. 4, pp. 769-805, 1998.
- [22] J. Li, Z. Xu, J. Zhao, and C. Wan, "A Coordinated Dispatch Model for Distribution Network Considering PV Ramp," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1107-1109, 2018.
- [23] J. Zhao, Z. Xu, J. Wang, C. Wang, and J. Li, "Robust Distributed Generation Investment Accommodating Electric Vehicle Charging in a Distribution Network," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1-1, 2018.
- [24] J. Li, C. Wan, and Z. Xu, "Robust offering strategy for a wind power producer under uncertainties," in *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2016, pp. 752-757.
- [25] M. A. Matos and R. J. Bessa, "Setting the Operating Reserve Using Probabilistic Wind Power Forecasts," *Power Systems, IEEE Transactions on*, vol. 26, no. 2, pp. 594-603, 2011.
- [26] N. Menemenlis, M. Huneault, and A. Robitaille, "Computation of Dynamic Operating Balancing Reserve for Wind Power Integration for the Time-Horizon 1–48 Hours," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 4, pp. 692-702, 2012.
- [27] J. Wang *et al.*, "Wind power forecasting uncertainty and unit commitment," *Applied Energy*, vol. 88, no. 11, pp. 4014-4023, 11// 2011.
- [28] A. Botterud *et al.*, "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois," in *Power and Energy Society General Meeting (PES), 2013 IEEE*, 2013, pp. 1-1.
- [29] H. Bludszweit and J. A. Domínguez-Navarro, "A probabilistic method for energy storage sizing based on wind power forecast uncertainty," *Power Systems, IEEE Transactions on*, vol. 26, no. 3, pp. 1651-1658, 2011.
- [30] E. D. Castronuovo and J. P. Lopes, "On the optimization of the daily operation of a wind-hydro power plant," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1599-1606, 2004.
- [31] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47-54, 2004.
- [32] P. Pinson and J. Tastu, "Discussion of "Prediction intervals for short-term wind farm generation forecasts" and "Combined nonparametric prediction intervals for wind power generation"," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 1019-1020, 2014.

- [33] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *Energy Conversion, IEEE Transactions on*, vol. 24, no. 3, pp. 775-782, 2009.
- [34] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind energy*, vol. 12, no. 1, pp. 51-62, 2009.
- [35] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, no. 0, pp. 255-270, 4// 2014.
- [36] N. Amjady and M. Hemmati, "Energy price forecasting-problems and proposals for such predictions," *IEEE Power and Energy Magazine*, vol. 4, no. 2, pp. 20-29, 2006.
- [37] M. Olsson and L. Soder, "Modeling real-time balancing power market prices using combined SARIMA and Markov processes," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 443-450, 2008.
- [38] Q. Zhou, L. Tesfatsion, and C.-C. Liu, "Scenario generation for price forecasting in restructured wholesale power markets," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, 2009, pp. 1-8: IEEE.
- [39] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2013.
- [40] W. Pedrycz, A. Skowron, and V. Kreinovich, *Handbook of Granular Computing*. Wiley-Interscience, 2008, p. 1148.
- [41] G. Zhang, Y. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H.-C. Iu, "An Advanced Approach for Construction of Optimal Wind Power Prediction Intervals," *Power Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1-10.
- [42] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Optimal prediction intervals of wind power generation," *Power Systems, IEEE Transactions on*, vol. 29, no. 3, pp. 1166-1174, 2014.
- [43] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *Neural Networks, IEEE Transactions on*, vol. 22, no. 3, pp. 337-346, 2011.
- [44] A. Botterud, "Chapter 11 - Forecasting Renewable Energy for Grid Operations," in *Renewable Energy Integration*, L. E. Jones, Ed. Boston: Academic Press, 2014, pp. 137-147.
- [45] EnergiNet, "Environmental report for Danish electricity and CHP for 2016 status year," 2017.
- [46] J. Nowotarski, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renewable and Sustainable Energy Reviews*, 2017.
- [47] P. Pinson, "Estimation of the uncertainty in wind power forecasting," École Nationale Supérieure des Mines de Paris, 2006.
- [48] P. Pinson, C. Chevallier, and G. N. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *Power Systems, IEEE Transactions on*, vol. 22, no. 3, pp. 1148-1156, 2007.

- [49] R. J. Bessa *et al.*, "Reserve Setting and Steady-State Security Assessment Using Wind Power Uncertainty Forecast: A Case Study," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 4, pp. 827-836, 2012.
- [50] J. Saez-Gallego, J. M. Morales, H. Madsen, and T. Jónsson, "Determining reserve requirements in DK1 area of Nord Pool using a probabilistic approach," *Energy*, vol. 74, pp. 682-693, 2014.
- [51] A. Botterud *et al.*, "Unit commitment and operating reserves with probabilistic wind power forecasts," in *PowerTech, 2011 IEEE Trondheim*, 2011, pp. 1-7.
- [52] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang, and K. Subbarao, "Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures," *Sustainable Energy, IEEE Transactions on*, vol. 2, no. 4, pp. 433-442, 2011.
- [53] J. Dowell, G. Hawker, K. Bell, and S. Gill, "A Review of probabilistic methods for defining reserve requirements," in *Power and Energy Society General Meeting (PESGM), 2016*, 2016, pp. 1-5: IEEE.
- [54] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control*. John Wiley & Sons, 2012.
- [55] X. Yan, D. Abbes, and B. Francois, "Uncertainty analysis for day ahead power reserve quantification in an urban microgrid including PV generators," *Renewable Energy*, vol. 106, pp. 288-297, 2017.
- [56] R. Koenker and K. F. Hallock, "Quantile regression," *Journal of economic perspectives*, vol. 15, no. 4, pp. 143-156, 2001.
- [57] T. Heskes, "Practical confidence and prediction intervals," in *Advances in neural information processing systems*, 1997, pp. 176-182.
- [58] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359-378, 2007.
- [59] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, pp. 125-151, 2014.
- [60] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243-268, 2007.
- [61] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767-2778, 2017.
- [62] S. Chai, Z. Xu, and W. K. Wong, "Optimal Granule-Based PIs Construction for Solar Irradiance Forecast," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3332-3333, 2016.
- [63] G. Biau and B. Patra, "Sequential quantile prediction of time series," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1664-1674, 2011.
- [64] T. Gneiting, A. E. Raftery, A. H. Westveld III, and T. Goldman, "Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation," *Monthly Weather Review*, vol. 133, no. 5, pp. 1098-1118, 2005.

- [65] P. Friederichs and T. L. Thorarinsdottir, "Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction," *Environmetrics*, vol. 23, no. 7, pp. 579-594, 2012.
- [66] H. Hersbach, "Decomposition of the continuous ranked probability score for ensemble prediction systems," *Weather and Forecasting*, vol. 15, no. 5, pp. 559-570, 2000.
- [67] F. Laio and S. Tamea, "Verification tools for probabilistic forecasts of continuous hydrological variables," *Hydrology and Earth System Sciences Discussions*, vol. 11, no. 4, pp. 1267-1277, 2007.
- [68] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12-20, 2012.
- [69] M. Scheuerer and T. M. Hamill, "Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities," *Monthly Weather Review*, vol. 143, no. 4, pp. 1321-1334, 2015.
- [70] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation—With application to solar energy," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3850-3863, 2016.
- [71] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [72] H. Bludszweit, J. A. Domínguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983-991, 2008.
- [73] A. Luig, S. Bofinger, and H. G. Beyer, "Analysis of confidence intervals for the prediction of regional wind power output," in *Proceedings of the European wind energy conference, Copenhagen, Denmark, 2001*, pp. 725-728.
- [74] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*. Springer Science & Business Media, 2013.
- [75] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307-327, 1986.
- [76] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *IEEE Transactions on Energy Conversion*, vol. 24, no. 3, p. 775, 2009.
- [77] A. Lau and P. McSharry, "Approaches for multi-step density forecasts with application to aggregated wind power," *The Annals of Applied Statistics*, pp. 1311-1341, 2010.
- [78] P.-J. Trombe, P. Pinson, and H. Madsen, "A general probabilistic forecasting framework for offshore wind power fluctuations," *Energies*, vol. 5, no. 3, pp. 621-657, 2012.
- [79] J. Jeon and J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66-79, 2012.
- [80] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 2, no. 1, pp. 2-10, 2009.

- [81] A. Misiorek, S. Trueck, and R. Weron, "Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models," *Studies in Nonlinear Dynamics & Econometrics*, vol. 10, no. 3, 2006.
- [82] R. Weron and A. Misiorek, "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models," *International Journal of Forecasting*, vol. 24, no. 4, pp. 744-763, 10// 2008.
- [83] A. Panagiotelis and M. Smith, "Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions," *International Journal of Forecasting*, vol. 24, no. 4, pp. 710-727, 10// 2008.
- [84] J. H. Zhao, Z. Y. Dong, Z. Xu, and K. P. Wong, "A statistical approach for interval forecasting of the electricity price," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 267-276, 2008.
- [85] C. Huurman, F. Ravazzolo, and C. Zhou, "The power of weather," *Computational Statistics & Data Analysis*, vol. 56, no. 11, pp. 3793-3807, 11// 2012.
- [86] A. K. Diongue, D. Guégan, and B. Vignal, "Forecasting electricity spot market prices with a k-factor GIGARCH process," *Applied Energy*, vol. 86, no. 4, pp. 505-510, 4// 2009.
- [87] B. Efron, "Bootstrap methods: another look at the jackknife," in *Breakthroughs in statistics*: Springer, 1992, pp. 569-593.
- [88] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong, "A hybrid approach for probabilistic forecasting of electricity price," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 463-470, 2014.
- [89] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction intervals for short-term wind farm power generation forecasts," *IEEE Transactions on sustainable energy*, vol. 4, no. 3, pp. 602-610, 2013.
- [90] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *Power Systems, IEEE Transactions on*, vol. 29, no. 3, pp. 1033-1044, 2014.
- [91] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 12// 2006.
- [92] Y. Chu, M. Li, H. T. Pedro, and C. F. Coimbra, "Real-time prediction intervals for intra-hour DNI forecasts," *Renewable Energy*, vol. 83, pp. 234-244, 2015.
- [93] X. Chen, Z. Y. Dong, K. Meng, Y. Xu, K. P. Wong, and H. W. Ngan, "Electricity Price Forecasting With Extreme Learning Machine and Bootstrapping," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2055-2062, 2012.
- [94] M. Rafiei, T. Niknam, and M. H. Khooban, "Probabilistic electricity price forecasting by improved clonal selection algorithm and wavelet preprocessing," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3889-3901, 2017.
- [95] A. M. Alonso, C. García-Martos, J. Rodríguez, and M. Jesús Sánchez, "Seasonal dynamic factor analysis and bootstrap inference: application to electricity market forecasting," *Technometrics*, vol. 53, no. 2, pp. 137-151, 2011.

- [96] A. Khosravi, S. Nahavandi, and D. Creighton, "Quantifying uncertainties of neural network-based electricity price forecasts," *Applied energy*, vol. 112, pp. 120-129, 2013.
- [97] A. Bello, J. Reneses, and A. Muñoz, "Medium-term probabilistic forecasting of extremely low prices in electricity markets: Application to the Spanish case," *Energies*, vol. 9, no. 3, p. 193, 2016.
- [98] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *Power Systems, IEEE Transactions on*, vol. 25, no. 4, pp. 1845-1856, 2010.
- [99] R. A. Rigby and D. M. Stasinopoulos, "Generalized additive models for location, scale and shape," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 54, no. 3, pp. 507-554, 2005.
- [100] F. Serinaldi, "Distributional modeling and short-term forecasting of electricity prices by Generalized Additive Models for Location, Scale and Shape," *Energy Economics*, vol. 33, no. 6, pp. 1216-1226, 11// 2011.
- [101] J. Juban, L. Fugon, and G. Kariniotakis, "Probabilistic short-term wind power forecasting based on kernel density estimators," in *European Wind Energy Conference and exhibition, EWEC 2007*, 2007, p. <http://ewec2007proceedings.info/>.
- [102] R. J. Bessa, J. Mendes, V. Miranda, A. Botterud, J. Wang, and Z. Zhou, "Quantile-copula density forecast for wind power uncertainty modeling," in *PowerTech, 2011 IEEE Trondheim*, 2011, pp. 1-8: IEEE.
- [103] R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang, "Time-adaptive quantile-copula for wind power probabilistic forecasting," *Renewable Energy*, vol. 40, no. 1, pp. 29-39, 2012.
- [104] R. Cao, J. Hart, and á. Saavedra, "Nonparametric maximum likelihood estimators for AR and MA time series," *Journal of Statistical Computation and Simulation*, vol. 73, no. 5, pp. 347-360, 2003.
- [105] J. Zhang, S. Chowdhury, A. Messac, and L. Castillo, "Multivariate and multimodal wind distribution model based on kernel density estimation," in *ASME 2011 5th International Conference on Energy Sustainability*, 2011, pp. 2125-2135: American Society of Mechanical Engineers.
- [106] Z. Qin, W. Li, and X. Xiong, "Estimating wind speed probability distribution using kernel density method," *Electric Power Systems Research*, vol. 81, no. 12, pp. 2139-2146, 2011.
- [107] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33-50, 1978.
- [108] J. B. Bremnes, "A comparison of a few statistical models for making quantile wind power forecasts," *Wind Energy*, vol. 9, no. 1- 2, pp. 3-11, 2006.
- [109] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1- 2, pp. 95-108, 2006.
- [110] J. Juban, L. Fugon, and G. Kariniotakis, "Uncertainty estimation of wind power forecasts: Comparison of probabilistic modelling approaches," in *European Wind Energy Conference & Exhibition EWEC 2008*, 2008, pp. 10 pages-<http://www.ewec2008proceedings.info/>: EWEC.

- [111] M. Landry, T. P. Erlinger, D. Patschke, and C. Varrichio, "Probabilistic gradient boosting machines for GEFCom2014 wind forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1061-1066, 2016.
- [112] R. Juban, H. Ohlsson, M. Maasoumy, L. Poirier, and J. Z. Kolter, "A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1094-1102, 2016.
- [113] K. Hatalis, A. J. Lamadrid, K. Scheinberg, and S. Kishore, "Smooth Pinball Neural Network for Probabilistic Forecasting of Wind Power," *arXiv preprint arXiv:1710.01720*, 2017.
- [114] J. Nowotarski and R. Weron, "Computing electricity spot price prediction intervals using quantile regression and forecast averaging," *Computational Statistics*, vol. 30, no. 3, pp. 791-803, 2015.
- [115] J. Nowotarski and R. Weron, "Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices," in *European Energy Market (EEM), 2014 11th International Conference on the*, 2014, pp. 1-5: IEEE.
- [116] K. Maciejowska, J. Nowotarski, and R. Weron, "Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging," *International Journal of Forecasting*, vol. 32, no. 3, pp. 957-965, 2016.
- [117] K. Maciejowska and J. Nowotarski, "A hybrid model for GEFCom2014 probabilistic electricity price forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1051-1056, 2016.
- [118] P. Gaillard, Y. Goude, and R. Nedellec, "Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting," *International Journal of forecasting*, vol. 32, no. 3, pp. 1038-1050, 2016.
- [119] A. Khosravi, S. Nahavandi, and D. Creighton, "A neural network-garch-based method for construction of prediction intervals," *Electric Power Systems Research*, vol. 96, pp. 185-193, 2013.
- [120] C. Wan, Z. Xu, J. Østergaard, Z. Y. Dong, and K. P. Wong, "Discussion of "Combined Nonparametric Prediction Intervals for Wind Power Generation"," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 1021-1021, 2014.
- [121] J. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 748-757, 1997.
- [122] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference On*, 1994, vol. 1, pp. 55-60: IEEE.
- [123] D. J. MacKay, "The evidence framework applied to classification networks," *Neural computation*, vol. 4, no. 5, pp. 720-736, 1992.
- [124] S. Tewari, C. J. Geyer, and N. Mohan, "A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2031-2039, 2011.

- [125] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1496-1503, 2010.
- [126] A. Khosravi and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 849-856, 2013.
- [127] W. Pedrycz, *Granular computing: analysis and design of intelligent systems*. CRC Press, 2013.
- [128] Y. Yiyu, "Perspectives of granular computing," in *Granular Computing, 2005 IEEE International Conference on*, 2005, vol. 1, pp. 85-90 Vol. 1.
- [129] J. Han and T. Y. Lin, "Granular computing: Models and applications," *International Journal of Intelligent Systems*, vol. 25, no. 2, pp. 111-117, 2010.
- [130] M. G. C. A. Cimino, B. Lazzerini, F. Marcelloni, and W. Pedrycz, "Genetic interval neural networks for granular data regression," *Information Sciences*, vol. 257, no. 0, pp. 313-330, 2/1/ 2014.
- [131] R. Al-Hmouz, W. Pedrycz, and A. Balamash, "Description and prediction of time series: A general framework of Granular Computing," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4830-4839, 6/15/ 2015.
- [132] W. Pedrycz, B. Russo, and G. Succi, "Knowledge transfer in system modeling and its realization through an optimal allocation of information granularity," *Applied Soft Computing*, vol. 12, no. 8, pp. 1985-1995, 2012.
- [133] D. P. Larson, L. Nonnenmacher, and C. F. Coimbra, "Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest," *Renewable Energy*, vol. 91, pp. 11-20, 2016.
- [134] F. Ziel, R. Steinert, and S. Husmann, "Efficient modeling and forecasting of electricity spot prices," *Energy Economics*, vol. 47, pp. 98-111, 2015.
- [135] C.R. Rao, S.K. Mitra, *Generalized Inverse of Matrices and its Applications*, Wiley, New York, 1971
- [136] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 2, pp. 513-529, 2012.
- [137] G.-B. Huang, X. Ding, and H. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing*, vol. 74, no. 1, pp. 155-163, 2010.
- [138] D. P. Loucks, E. Van Beek, J. R. Stedinger, J. P. Dijkman, and M. T. Villars, *Water resources systems planning and management: an introduction to methods, models and applications*. Paris: Unesco, 2005.
- [139] J. C. Refsgaard, J. P. Van der Sluijs, J. Brown, and P. Van der Keur, "A framework for dealing with uncertainty due to model structure error," *Advances in Water Resources*, vol. 29, no. 11, pp. 1586-1597, 2006.
- [140] E. Zio and T. Aven, "Uncertainties in smart grids behavior and modeling: What are the risks and vulnerabilities? How to analyze them?," *Energy Policy*, vol. 39, no. 10, pp. 6308-6320, 2011.
- [141] H. Agarwal, J. E. Renaud, E. L. Preston, and D. Padmanabhan, "Uncertainty quantification using evidence theory in multidisciplinary design optimization," *Reliability Engineering & System Safety*, vol. 85, no. 1-3, pp. 281-294, 2004.

- [142] R. Ak, V. Vitelli, and E. Zio, "An Interval-Valued Neural Network Approach for Uncertainty Quantification in Short-Term Wind Speed Prediction," 2015.
- [143] C. W. Chow *et al.*, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881-2893, 2011.
- [144] W. Pedrycz and W. Homenda, "Building the fundamentals of granular computing: a principle of justifiable granularity," *Applied Soft Computing*, vol. 13, no. 10, pp. 4209-4218, 2013.
- [145] J. Ruan, X. Wang, and Y. Shi, "Developing fast predictors for large-scale time series using fuzzy granular support vector machines," *Applied Soft Computing*, vol. 13, no. 9, pp. 3981-4000, 9// 2013.
- [146] C. Wan, "Uncertainty analysis, modelling and prognosis for power system operation and planning," The Hong Kong Polytechnic University, 2015.
- [147] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*: Springer, 2011, pp. 760-766.
- [148] A. Roque, C. Maté, J. Arroyo, and Á. Sarabia, "iMLP: Applying Multi-Layer Perceptrons to Interval-Valued Data," (in English), *Neural Processing Letters*, vol. 25, no. 2, pp. 157-169, 2007/04/01 2007.
- [149] M. G. C. A. Cimino, B. Lazzarini, F. Marcelloni, and W. Pedrycz, "Genetic interval neural networks for granular data regression," *Information Sciences*, vol. 257, pp. 313-330, 2/1/ 2014.
- [150] D. Chetwynd, K. Worden, and G. Manson, "An application of interval-valued neural networks to a regression problem," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 2006, vol. 462, no. 2074, pp. 3097-3114: The Royal Society.
- [151] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, "Non- parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497-516, 2007.
- [152] T. J. Hastie, "Generalized additive models," in *Statistical models in S*: Routledge, 2017, pp. 249-307.
- [153] W. Can, X. Zhao, and P. Pinson, "Direct Interval Forecasting of Wind Power," *Power Systems, IEEE Transactions on*, vol. 28, no. 4, pp. 4877-4878, 2013.
- [154] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Solar Energy*, vol. 84, no. 5, pp. 807-821, 2010.
- [155] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, "Interval type-2 fuzzy logic systems for load forecasting: A comparative study," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1274-1282, 2012.
- [156] J. P. Catalão, *Electric power systems: advanced forecasting techniques and optimal generation scheduling*. CRC Press, 2012.
- [157] C. Wan, J. Lin, Y. Song, Z. Xu, and G. Yang, "Probabilistic forecasting of photovoltaic generation: an efficient statistical approach," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2471-2472, 2017.
- [158] C. Wan, M. Niu, Y. Song, and Z. Xu, "Pareto optimal prediction intervals of electricity price," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 817-819, 2017.

- [159] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using PSO-tuned support vector machines," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 2, pp. 322-331, 2015.
- [160] T. Jónsson, P. Pinson, H. Madsen, and H. A. Nielsen, "Predictive densities for day-ahead electricity prices using time-adaptive quantile regression," *Energies*, vol. 7, no. 9, pp. 5523-5547, 2014.
- [161] Ö. F. Ertuğrul and Y. Kaya, "A detailed analysis on extreme learning machine and novel approaches based on ELM," *American Journal of Computer Science and Engineering*, vol. 1, no. 5, pp. 43-50, 2014.
- [162] H. Farhangi, "The path of the smart grid," *IEEE power and energy magazine*, vol. 8, no. 1, 2010.
- [163] V. Fonti and E. Belitser, "Feature Selection using LASSO," ed, 2017.
- [164] N. Ludwig, S. Feuerriegel, and D. Neumann, "Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests," *Journal of Decision Systems*, vol. 24, no. 1, pp. 19-36, 2015.
- [165] G. Udny Yule and M. G. Kendall, "An introduction to the theory of statistics," *An introduction to the theory of statistics.*, no. 14th ed, 1950.
- [166] F. Ziel, "Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4977-4987, 2016.
- [167] B. Uniejewski, J. Nowotarski, and R. Weron, "Automated variable selection and shrinkage for day-ahead electricity price forecasting," *Energies*, vol. 9, no. 8, p. 621, 2016.
- [168] Q. Yu, Y. Miche, E. Eirola, M. Van Heeswijk, E. SéVerin, and A. Lendasse, "Regularized extreme learning machine for regression with missing data," *Neurocomputing*, vol. 102, pp. 45-51, 2013.
- [169] S. Baran and S. Lerch, "Log- normal distribution based Ensemble Model Output Statistics models for probabilistic wind- speed forecasting," *Quarterly Journal of the Royal Meteorological Society*, vol. 141, no. 691, pp. 2289-2299, 2015.
- [170] M. Scheuerer and D. Möller, "Probabilistic wind speed forecasting on a grid based on ensemble model output statistics," *The Annals of Applied Statistics*, vol. 9, no. 3, pp. 1328-1349, 2015.
- [171] N. Balakrishnan, *Handbook of the logistic distribution*. CRC Press, 2013.
- [172] J. Prokosch, "Bivariate Bayesian Model Averaging and Ensemble Model Output Statistics: With a Case Study of Ensemble Temperature Forecasts in Trondheim," 2013.
- [173] J.-F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media, 2006.
- [174] X.-Y. Ma, Y.-Z. Sun, and H.-L. Fang, "Scenario generation of wind power based on statistical uncertainty and variability," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 894-904, 2013.
- [175] D. Singhal and K. Swarup, "Electricity price forecasting using artificial neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 3, pp. 550-555, 2011.

- [176] F. Ziel and R. Steinert, "Electricity price forecasting using sale and purchase curves: The X-Model," *Energy Economics*, vol. 59, pp. 435-454, 2016.
- [177] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & economic statistics*, vol. 20, no. 1, pp. 134-144, 2002.
- [178] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986.
- [179] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189-1232, 2001.
- [180] N. M. Razali and A. Hashim, "Backward reduction application for minimizing wind power scenarios in stochastic programming," in *Power Engineering and Optimization Conference (PEOCO), 2010 4th International*, 2010, pp. 430-434: IEEE.
- [181] H. Zareipour, C. A. Canizares, and K. Bhattacharya, "Economic impact of electricity market price forecasting errors: a demand-side analysis," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 254-262, 2010.
- [182] T. Hong, "Crystal ball lessons in predictive analytics," *EnergyBiz Mag*, pp. 35-37, 2015.
- [183] P. K. Shivaswamy, C. Bhattacharyya, and A. J. Smola, "Second order cone programming approaches for handling missing and uncertain data," *Journal of Machine Learning Research*, vol. 7, no. Jul, pp. 1283-1314, 2006.
- [184] C. Bhattacharyya, P. K. Shivaswamy, and A. J. Smola, "A second order cone programming formulation for classifying missing data," in *Advances in neural information processing systems*, 2005, pp. 153-160.