Uncertainty quantification of photovoltaic power generation

Faranak Golestaneh

2017

Faranak Golestaneh. (2017). Uncertainty quantification of photovoltaic power generation.
Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/69658>

<https://doi.org/10.32657/10356/69658>

UNCERTAINTY QUANTIFICATION OF PHOTOVOLTAIC POWER GENERATION

FARANAK GOLESTANEH

School of Electrical and Electronic Engineering

A thesis submitted to Nanyang Technological University
in partial fulfillment of the requirements for the degree of
PhD of Engineering

2017

The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present.

by Paul Saffo

Acknowledgements

This thesis would not have been possible without the inspiration and support of a number of wonderful individuals. My thanks and appreciation to all of them for being part of this journey and making this thesis possible.

I would like to express my special appreciation and thanks to my supervisor Associate Professor Gooi Hoay Beng, for his professional guidance and patience offered during this period. He has always encouraged my research and allowed me to find my area of interest and grow as an independent researcher.

I owe my deepest gratitude to Professor Pierre Pinson for his encouragement and support during my study. His professionalism and enthusiasm in research has always been a constant source of inspiration for spurring me on.

I would like to thank Professor Saifur Rahman and Virginia Tech (VT) University for giving me the opportunity to be a member of VT as a visiting scholar. Thanks to Professor Saifur Rahman for the constant encouragement and interactive discussions he offered me.

I am also thankful to the technical staff of Clean Energy Research Laboratory for the assistance and support provided by Mdm. Chia-Nge Tak Heng.

This research work is supported by the Agency for Science, Technology and Research (A*STAR). I am grateful to A*STAR and Nanyang Technological University (NTU) for awarding me the research scholarship and giving me the opportunity to pursue a fulfilling graduate study.

My time at NTU was made enjoyable due to many friends and groups that became a part of my life. I am grateful for all my friends who made my stay in Singapore memorable.

A special thanks to my family. Words cannot express how grateful I am to my family for all of the sacrifices that they have made on my behalf. I know I always have my family to count

on when times are rough.

And finally to Mehdi, his support has been unconditional all these years. He has been by my side throughout this PhD, living every single minute of it even when I was irritable and depressed. This journey became possible because he never stopped having faith in me.

Table of Contents

Acknowledgement	iii
Table of Contents	iv
List of Figures	viii
List of Tables	xi
List of Acronyms	xii
Abstract	xiii
1 Introduction	1
1.1 Introduction and Motivation	1
1.2 Contributions	5
1.2.1 Point forecasting contributions	6
1.2.2 Probabilistic forecasting contributions	6
1.2.3 Spatio-temporal forecasting contributions	7
1.3 Objectives	8
1.4 Organization of the Report	9
2 Literature Review	12
2.1 Introduction	12
2.2 Uncertainty versus Variability	13
2.2.1 Variability	13
2.2.2 Uncertainty	14
2.3 Data Used for Forecasting	15
2.3.1 Numerical Weather Prediction (NWP)	16
2.3.2 Ensemble forecasts	17
2.4 Key Factors in Forecast Accuracy and Relevance	17
2.4.1 Geographical aggregation and uncertainty diversity	17

2.4.2	Forecast horizon	18
2.4.3	Regional climate	18
2.5	Solar Power Point Forecasts	19
2.5.1	Physical approaches	19
2.5.2	Statistical approaches	20
2.5.3	Performance evaluation	22
2.6	Probabilistic Forecasts	22
2.6.1	A decision-making example on the application of probabilistic forecasts .	23
2.6.2	Two groups of probabilistic forecasts	25
2.6.2.1	Parametric approaches	26
2.6.2.2	Nonparametric approaches	29
2.6.3	Prediction intervals versus confidence intervals	31
2.6.4	Performance evaluation	32
2.6.5	A review on state-of-the-art on probabilistic forecasts of PV Power	34
2.7	Spatio-temporal Correlation of PV Generation	35
2.7.1	Performance evaluation	38
2.7.2	A review on state-of-the-art on spatio-temporal analysis of PV power . .	38
2.8	Conclusions	39
3	Batch and Sequential Forecast Models for Photovoltaic Generation	41
3.1	Introduction	41
3.2	Extreme Learning Machine (ELM)	41
3.3	Online Sequential Learning Algorithm	43
3.4	Online and Adaptive PV Forecast Model	46
3.5	Empirical Investigations	48
3.6	Conclusions	53
4	Very Short-term Nonparametric Probabilistic Forecasting of Photovoltaic Generation	55
4.1	Introduction	55
4.2	Experimental Framework	56
4.3	Non-Parametric Probabilistic Forecasts	57
4.4	Performance Evaluation	58
4.4.1	Reliability	58
4.4.2	Skill score	58

4.5	From Point Predictions to Nonparametric Probabilistic Forecasts	59
4.5.1	Extreme Learning Machine (ELM)	60
4.5.2	Optimization formulation	60
4.5.3	Optimization procedure	62
4.6	Benchmark Methods	64
4.6.1	HIA	64
4.6.2	BELM	64
4.6.3	Probabilistic persistence benchmark	66
4.6.4	Probabilistic climatology benchmark	66
4.7	Empirical Investigation	68
4.7.1	ELM tuning	68
4.7.2	Analysis of forecast error distributions	69
4.7.3	Nonparametric quantile forecast results	72
4.7.4	Normalization with respect to clear sky PV power	79
4.8	Conclusions	82
5	Space-Time Trajectories of Photovoltaic Power	83
5.1	Introduction	83
5.2	Experimental Data Description	84
5.3	Deriving and Evaluating Predictive Marginal Distribution	85
5.4	Space-time Dependency Modelling	86
5.4.1	Multivariate distributions	86
5.4.2	Multivariate Normal distribution with adaptive correlation matrix	88
5.4.3	Trajectory generation	89
5.5	Multivariate Scoring Rules	89
5.5.1	Energy Score	90
5.5.2	Variogram-based Score	90
5.5.3	Event-based Score	91
5.5.4	Probability Integral Transformation (PIT)	94
5.6	Results	94
5.6.1	Benchmark methods	94
5.6.1.1	Naive method	94
5.6.1.2	Independent method	95
5.6.1.3	Independent Gaussian	96

5.6.1.4	Gaussian copula with full empirical covariance	96
5.6.2	Time grid	96
5.6.3	Marginal distributions	97
5.6.4	The goodness of fit of Gaussian Copula	99
5.6.5	Multivariate analysis	102
5.7	Conclusions	110
6	Conclusion and Future Work	112
6.1	Conclusion	112
6.2	The Applications of the Developed Forecasting Frameworks	114
6.2.1	Forecast horizon	115
6.2.2	Forecast type	115
6.3	Future Works	117
6.3.1	Finding the right multivariate copula for modelling spatio-temporal de- pendency in PV generation	117
6.3.2	Estimation of multivariate prediction regions for random generation . . .	118
6.3.3	Regime switching models for PV generation	121
6.3.4	Sky imaging	122
6.3.5	The impacts of PV power fluctuations on power market and demand response management	122
	List of Publications	123
	References	124
	Appendix A Kernel Density Estimate (KDE)	135
	Appendix B Cross Validation	137
B.1	<i>K</i> -Fold Cross Validation	137
	Appendix C Clear Sky Solar Irradiance	139

List of Figures

2.1	Clear sky PV power along with observed PV output for a typical day from the case study (<i>Site A</i>) described in Section ??	14
2.2	Variability vs. Uncertainty: PV power measurements along day-ahead point forecasts for four typical successive days.	15
2.3	A typical CDF for wind/solar production.	24
2.4	Empirical distribution and four fitted parametric distributions for PV generations at 11:00 am.	27
2.5	Two typical predictive densities along with their corresponding 80 % prediction intervals.	34
2.6	Space correlations between PV generations at three PV zones described in Section ??	36
2.7	Time correlations between PV generations at PV zone 1 described in Section ??	37
3.1	The flowchart of ELM forecasting method	44
3.2	The flowchart of OSELM forecasting method.	47
3.3	The flowchart of the proposed adaptive forecast model.	48
3.4	10-minute ahead point forecasts for 6 successive days, Site <i>A</i>	50
3.5	10-minute ahead point forecasts for 6 successive days, Site <i>B</i>	50
3.6	1-24 hour ahead point forecasts for 8 successive days. Forecasts are performed at 00:00 each day, Site <i>C</i>	51
3.7	Histogram of the RMSE values provided by ELM in 500 independent runs, Site <i>C</i>	53
4.1	PV power variability for a typical day at Site <i>A</i> in March 2013	56
4.2	The flowchart of ELM forecasting method.	63
4.3	Kernel smoothing density estimate of distribution of errors in forecasted values (<i>Case 1</i>).	70
4.4	Kernel smoothing density estimate of distribution of errors in forecasted values (<i>Case 2</i>).	71

4.5	Top: the overall kernel density of bins altogether versus fitted <i>Normal</i> distribution to the same data for <i>Case 1</i> and <i>Case 2</i> , respectively. Bottom: Empirical CDF versus fitted <i>Normal</i> CDF to the same data	72
4.6	Predictive densities as a set of prediction intervals for five successive days in January 2013 (<i>Case 1</i>)	73
4.7	Predictive densities as a set of prediction intervals for five successive days in August 2013 (<i>Case 2</i>)	73
4.8	Predictive densities as a set of prediction intervals for two successive days in January 2013 (<i>Case 3</i>)	74
4.9	BELM predictive densities as a set of prediction intervals for one day in March 2013 (<i>Case 3</i>)	76
4.10	Sharpness diagrams for 9 central prediction intervals with nominal coverage rates increasing from 10% to 90% for <i>Case 1</i> , (a) Spring, (b) Summer, (c) Fall, (d) winter	76
4.11	Reliability diagrams for 18 quantiles with 5% increasing nominal proportion, <i>Case 1</i> (a) Spring, (b) Summer, (c) Fall, (d) Winter	77
4.12	PV power measurements (red colour curve), physical clear sky power (green colour curve), normalized PV power with respect to the clear sky values (black colour curve) and normalized values with respect to 85% quantile (cyan colour curve) for 6 randomly selected days from the dataset for <i>Case 2</i> . The diagram covers values for 5 am to 7 pm. Counting minutes of day is started from 5 am. Power values more than 1.5 pu are not shown.	80
5.1	All pairwise scatter plots of PV measurements in three zones	95
5.2	PV observations (yellow colour curves) along with 49 prediction intervals with coverage ranging from 0.02 to 0.98 by 0.02 increments (from the darkest to the lightest)	97
5.3	Reliability evaluation of 99 quantiles with 0.01 increasing nominal proportion generated by quantile regression for zone 1 and all forecast horizons.	98
5.4	Skill evaluation of the probabilistic forecasts of PV generation as a function of forecast horizon	99
5.5	Comparison of real measurement and the simulated Gaussian process for the dependency between PV power at 15:00 and 16:00 over the all data available for Zone 1.	100
5.6	Comparison of empirical and Gaussian copulas for the dependency between PV power at 15:00 and 16:00 over all data available for zone 1.	101
5.7	PV observations (dark black colour curves) along with 20 generated space-time trajectories (grey colour curves)	103
5.8	Time trajectories: evaluation of the probabilistic correctness of generated scenarios with a PIT histogram.	104

5.9	Space-time trajectories: evaluation of the probabilistic correctness of generated scenarios with a PIT histogram.	105
5.10	Temporal trajectories: the scores for multivariate normal distribution with recursive covariance and five other benchmarks. 5 different sizes for the trajectories set are considered.	106
5.11	Space-time trajectories: the scores for multivariate normal distribution with recursive covariance and five other benchmarks. 5 different sizes for the trajectories set are considered.	107
5.12	Visualization of correlation matrix for the transformed measurements on the time grid, the matrix is of size of 45×45 . Red to white colours represent the lowest to the highest correlations	108
5.13	Space-time trajectories: BS2 and BS3 scores for MVN, naive and Independent methods for five different values of ξ . For BS2 $k = 11$ and $h = 2$ and for BS3 $k = 11$ and $h = 4$	109
6.1	Calibration of the simultaneous prediction intervals.	120
6.2	PV power observations (green colour curves) along with 19 simultaneous prediction intervals with nominal coverage ranging from 0.05 to 0.95 by 0.05 increments (from the darkest to the lightest), for 3 randomly selected days from the evaluation set. t refers to day number in the year.	121
A.1	Kernel density estimate for an artificial dataset	136
B.1	A typical diagram for 5-fold cross validation	137

List of Tables

2.1	Potential power system operational problem affected by various intermittency timescales	18
2.2	Kernel functions for different random variables [1]	31
3.1	Data resolution, forecast horizon and predictors in three case studies	49
3.2	Comparative results for point predictions (accuracy metrics in percent (%)) . . .	52
4.1	Scores, maximum and sum of probability deviation of quantiles, <i>Case 1</i>	74
4.2	Scores, maximum and sum of probability deviation of quantiles, <i>Case 2</i>	75
4.3	Scores, maximum and sum of probability deviation of quantiles, <i>Case 3</i>	75
4.4	Point forecasts accuracy by season, <i>Case 1</i>	78
5.1	Description of the three PV zones	84
5.2	Comparison of BIC and AIC values for various copulas	102

List of Acronyms

ANN	Artificial Neural Network
AR	AutoRegressive
ARIMA	Autoregressive Integrated Moving Average (ARIMA)
ARMA	Autoregressive Moving Average
BELM	Extreme Learning Machine (ELM)-Bootstrap
BS	Brier Score
CDF	Continuous Rank Probability Score
CRPS	Cumulative Distribution Function
CV	Cross Validation
ECMWF	European Centre for Medium-Range Weather Forecasts
ELM	Extreme Learning Machine
ES	Energy Score
ESM	Exponential Smoothing Method
FFBPG	Feed-forward Back Propagation Neural Network
GARCH	Generalized Autoregressive with Conditional Heteroscedasticity
GEFCom	Global Energy Forecasting Competition
GHI	Global Horizontal Irradiance
HIA	Hybrid Intelligent Algorithm
ISO	Independent System Operator
KDE	Kernel Density Estimate
<i>MAE</i>	Mean Absolute Error
MSAR	Markov-switching Autoregressive
MVN	Multivariate Normal Distribution
NTU	Nanyang Technological University
NWP	Numerical Weather Prediction
OSELN	Online Sequential Extreme Learning Machine
PDF	Probability Density Function
PE	Ellipsoidal Prediction regions
PIT	Probability Integral Transformation
PSO	Particle Swarm Optimization
PV	Photovoltaic
QR	Quantile Regression
<i>rRMSE</i>	Relative Root Mean Square Error
<i>RMSE</i>	Root Mean Square Error
SLFN	Single-Hidden Layer Feed-forward Neural Network
VS	Variogram-based Score

Abstract

Environmental concerns and global energy crises have caused a rapid proliferation of renewable energy sources worldwide. For Singapore, due to its geographical location within the tropical Sun Belt, PV energy is the most propitious kind of renewables that can be deployed. Two important issues related to intermittency of PV generation are variability and uncertainty. Variable power with uncertain nature has several consequences in power systems such as the need for rising ancillary service, increase in voltage and frequency fluctuations and adverse effects on power quality and economics. What exacerbates the situation is the lack of inertia in PV systems which leads to abrupt changes in voltages and power outputs.

Improving PV forecasting methods, as an indispensable approach to mitigate PV power intermittency, has attracted researchers recently. Forecasts can be done as point or probabilistic predictions. Nearly all previous works focused on single-valued (or point) forecasts, similar to the case of wind power forecasting. However, point forecasts can only be helpful when no significant uncertainty is involved, since it fails to dispense a full picture of all potential future outcomes.

On the other hand, the aim of probabilistic predictions is to provide decision-making under uncertainty with the full information. In recent years, there has been a surge of interest in stochastic optimization approaches to cover different uncertainties in power systems. It is often assumed that the random variables involved have known parametric distributions. However, even in cases where observations form a known and well-behaved marginal distribution, there is no guarantee that conditional predictive densities (or distributions of forecast errors) follow that same distribution. Wrong distributional assumptions may directly yield biases in analyses and results. Stochastic optimization therefore calls for a thorough design and evaluation of probabilistic forecasting approaches.

Although valuable studies have been conducted on probabilistic wind forecasting, one can hardly find published investigations on probabilistic solar energy forecasts. In spite of the similarities in wind and PV power forecasts, there are significant differences such as the influential

variables and the relationship between meteorological variables and the available PV power which may result in different statistical features. This, indeed, necessitates a more focused research on solar power statistics. This thesis aims at providing practical methods for both point and probabilistic forecasts of PV generation.

If probabilistic forecasts are properly employed, they can serve as a decision-aiding tool to alleviate challenges attached to stochastic generation. However, despite of the benefits of probabilistic forecasts over point forecasts, they fail to capture development of forecast errors through successive lead-times, interdependent generation in contiguous locations or negatively correlated generation levels in diverse geographic areas. The reason is that they treat random variables for each lead-time and each location individually and separately while PV generations are stochastic processes with spatially spread and time interdependent infeeds. Therefore, in multi-stage decision-making problems such as unit-commitment or optimal power flow, it is an integral requirement to estimate aggregated uncertainties in the system and model space-time stochasticity of intermittent resources. This issue is addressed here by analysing and capturing spatio-temporal dependencies in PV generation. Multivariate predictive distributions are modelled and space-time trajectories describing the potential evolution of forecast errors through successive lead-times and locations are generated.

Chapter 1

Introduction

1.1 Introduction and Motivation

Increasing concerns about environmental issues and the global energy crisis have led to a rapid increase of interest in renewable energy sources, such as solar energy. However solar energy is mainly exploited in regions located outside the Sunbelt area (within 35 degrees north and 35 degrees south) with less solar radiation. In 2009, only 9% of the installed Photovoltaic (PV) capacity was located in Sunbelt countries [2]. Singapore being located almost on the equator, PV is one of the most propitious kinds of renewable energy generation techniques that can be deployed.

Although in recent years, there has been a surge of interest in PV power, its non-dispatchable and intermittent nature is still a challenging barrier against high penetration of solar energy into the grid. The lack of inertia in PV systems, high ramp rates due to cloud passages, and switching in power generation dynamics pose challenges at different time scales. Very short-term fluctuations in PV output have an impact on power quality (e.g., voltage flicker), generation-load imbalance, and regulation cost. Since increasing regulation reserve to compensate PV fluctuations is a costly option, using high-quality forecasts to preventively deal with potential imbalances is indispensable [3, 4]. For instance, a part of imbalances can be adjusted through re-dispatch rather than limited and expensive frequency regulation control. This motivates us to focus on very short-term and high frequency forecasting, in a probabilistic forecasting framework. Here, “very to short-term” is referred to lead times up to one hour ahead. Day-ahead forecasting also possesses a high economic value for power system operation because the

majority of conventional generation units are scheduled based on the output of the day-ahead market. Therefore, 1 to 24 hour-ahead predictions are also covered in this work.

In general, a more accurate forecast can bring several benefits to the power systems. Better prediction of solar power generation will help system operators to better manage the dispatchable generators and the generation-load balance such that system production cost is minimized to the best extent possible. Moreover, accurate forecasts help firm up the frequency regulation and reserve markets by reducing the uncertainty associated with variable generation resources. This will avoid paying for extra generation in the frequency regulation market that will not be needed with more accurate forecast from solar units.

Predictions of intermittent generations including PV energy can be performed in a number of different manners. Depending on the operational problems and the inherent stochastic process, the preference for the forecast and its presentation may vary. However, one should note that any type of forecast is based on the crucial assumption that the existing patterns and the identified dynamics will continue in the future.

Developing a forecast structure and fitting it using historical/predicted data make the forecasting conditional [5]. A conditional forecast is prediction of an event given that other events occurred. One might think of the probability that future PV power will be between 10 kW and 25 kW if the current temperature is 30 °C, or if the solar power at the day before, the same hour has been 22 kW.

The most commonly known type of forecast is the deterministic forecast in which a single value is predicted for each lead time. For instance, a forecaster may predict that PV power generated by a particular PV site is expected to be 20 kW the day after today at 3 pm.

In the last few years, several researchers have focused on developing point forecast techniques for solar irradiance or PV power [6–14]. The online short-term PV forecasting using a clear sky model described in [15] comprises a notable example of a relevant approach. A recent review of the state of the art can be found in [6].

In nearly all practices carried out in the area of PV power (energy) forecasting, the emphasis has been placed on point or single-valued forecasts, similar to the case of wind power forecasting [16]. Although point forecasts are easily understandable to the end-users, they are always subject to errors. If the consequences of relying on inaccurate forecasts are significant, it is required to provide a full picture of the estimated error. However, point forecasts are incapable of doing so [17]. For instance, in the above example, the probability that 20 kW

forecast comes precisely true is close to zero. In other words, the forecast is always precisely wrong but approximately right. In the above example, what forecaster means by 20 kW is that the predicted average of all potential power generations is 20 kW. However, in many applications, decision-makers need to know how much uncertainty is associated with forecasts. Error measure criteria like *RMSE* can only inform about historical performance of a prediction technique not an estimation of the uncertainty associated with each given forecast [18]. This is the aim of probabilistic predictions to provide decision-making under uncertainty with the full information.

Although probabilistic forecasting still is not commonly used in energy applications, its applications are becoming more widespread in different decision-making problems under uncertainty such as weather predictions [19], economics and finance [20]. As an example, the bank of England provides official probabilistic forecasts of Gross Domestic Product (GDP) every quarter. In [21], it has been shown that for a large family of decision-making problems, e.g. in economics and finance or weather and climate prediction among other areas, relying on probabilistic forecast results in more optimal decisions.

For the case of energy analysis, due to the inherent uncertainty involved in renewable energy forecasting, uncertainty quantification is a key input to maintain acceptable levels of reliability and profitability in power system operation. Indeed in recent years, there has been a surge of interest in stochastic optimization approaches to cover different uncertainties in power systems and electricity markets [5]. Various problems have been analysed, e.g., stochastic security-constrained unit commitment [22], probabilistic optimal power flow [23], stochastic feeder reconfiguration [24] and market participation for renewables [25]. For a recent overview on uncertainty quantification and its integration in various decision-making problems related to electricity markets, the reader is referred to [18].

In stochastic decision-making, uncertainty estimation is often modelled based on questionable assumptions. For example, for the case of density estimation, usually parametric distributions such as Gaussian or Beta are considered. Also, for uncertainty sets in the form of intervals, often it is assumed that intervals are centred on point forecasts and have “assumed” widths. However, the fact is that, even if the historical observations of a random variable can be characterized by a specific parametric distribution, there is no guarantee that the probability density of each single forecast follows that same distribution. When relying on uncertainty estimation in decision-making, any erroneous assumption can lead to biases in analyses and

undermines the benefits of stochastic programming.

Probabilistic forecast context is more evolved in the realm of wind energy. As relevant parallel studies, there have been comprehensive works carried out for the case of probabilistic forecasting of wind power generation. These studies can be categorized as quantile [26, 27], interval [28] and density forecasting [29], risk indices [30], and scenario generation [31]. In contrast to the vast amount of work on probabilistic wind power forecasting, one can hardly find proposals and empirical investigations for the solar energy prediction. Although solar energy forecasting can benefit from developments in wind forecasting, it has its own unique challenges such as higher ramp rates due to lack of inertia and movements of clouds, or higher share of distributed and small scale generation. This has motivated us to investigate the potential benefits that uncertainty forecasts of the PV power generation can bring for either PV system owners or national networks.

Although probabilistic forecasts dispense considerable information about the inherent uncertainty in the stochastic process of interest, they fail to inform about interdependence structure of uncertainty at different locations and/or different lead times. The reason is that marginal densities in the probabilistic forecast context are generated on per-site, per lead time basis.

Therefore, if operational problems are spatially and/or temporally coupled, both point forecasts and predictive marginal distributions given by probabilistic forecast 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.

When forecast errors are correlated in space, if a large forecast error occurs in site A, it is expected to observe large forecast errors in the neighboring sites as well. Similarly, if forecast errors are correlated in time, a large forecast error at time $t + k$, with t as the current time and k as the lead time, most probably will be followed by another large error in the next time step $(t + k + 1)$ [32]. This impacts on overall overestimation and underestimation of net load in the control zone as well as the required ramp-up and ramp-down flexibility provided by controllable generators in successive hours.

The foundation of the few present studies [33–37] on the spatio-temporal PV power forecasting is on deploying measurements or meteorological data from the neighbouring locations into the forecasting process of the site of interest (spatial analyses). Moreover, as the developed methods use lagged data, they are categorized as temporal investigations. Here, spatio-temporal

correlations of PV power are leveraged in a different way. In contrast with [33–37] in which dependencies are founded on the base of point forecasts, here the dependencies are modeled based on probabilistic forecasts. The goal in this study is to provide more informative forecasts for probabilistic decision-making.

Space-time dependencies can be described by multivariate densities. The copula concept can be used to characterize joint density with predictive marginal densities as inputs [38]. Copulas allow us to specify the dependency between random variables independently of their marginal distribution functions. This is important because generation of marginal density for each individual site and lead time, is already a challenging task.

Space-time trajectories as realizations (samples) of joint density can be generated using copulas. When those trajectories are deployed in operational problems, they should be considered as equally likely samples of the multivariate stochastic process of interest.

1.2 Contributions

This work aims at proposing original solutions for uncertainty estimation and quantification of the PV generation. Three complementary solutions have been developed. Depending on the decision-making problem, the objective and the risk tolerance of the decision-maker, the most proper one can be deployed.

The first part of this work concentrates on developing point forecast frameworks for both intra-hour and day ahead PV generation prediction. The fact is that point forecasts always are subject to errors. If the level of the error is acceptable for the decision-maker, then point forecasting can be a sufficient approach. However, if the associated risk and error are noticeable and the likely consequences can be significant, then the decision-maker requires an estimation of the uncertainty attached to forecasts. In this case, probabilistic forecasts would be essential.

In case the decision-making processes are temporally and spatially coupled, the integral requirement would be to estimate aggregated uncertainties in the system and model space-time stochasticity of uncertain variables. In order to deal with this issue, in this thesis, spatio-temporal stochasticity of PV generation is analysed and a framework to model multivariate predictive distributions is developed.

1.2.1 Point forecasting contributions

In point forecasting, depending on the application, the system operator may need forecast models which are trained based on batch or sequential approaches. In this thesis, ELM is used for batch training and forecasts for more than one hour lead time. Moreover, an adaptive and ensemble model based on Online Sequential Learning Algorithm (OSELM) [39] is proposed for very short-term forecast horizons. It should be noted that hereafter very short-term is refereed to intra-hour forecasts.

One of the pragmatic approaches to improve the accuracy of the forecasting methods is to develop ensemble models instead of using only a single learner [40]. This is especially important for PV generation forecasting due to its regime switching aspects in minute and second time scales. Therefore, to enhance the capability of the forecast model in following different dynamics and ramp-rates in PV generation time series, an ensemble adaptive forecast models is developed which works based on the OSELM learning theory. The proposed method is called nOSELM.

For forecasts with up to 6 hours lead time, statistical approach provide the best accuracy while for more than six hours lead time predictions, Numerical Weather Prediction (NWP) is essential. Generally, for lead times more than six hours, the best approach would be to combine NWP and statistical models [41]. Therefore, in this study, for intra-hour PV prediction, ELM and nOSELM models are developed with historical data as their inputs. However, for day-ahead predictions, NWP is used as the input to provide more accurate forecasts.

Three datasets with one minute and one hour temporal resolutions are used for evaluation of the methods. The performance of the proposed methods is compared with that of some the relevant approaches in terms of the commonly used accuracy metrics.

1.2.2 Probabilistic forecasting contributions

Most of the previous works on PV forecasting focused on point prediction only, hence researching on conditional expectation of PV power output. Here, a systematic framework is proposed to generate probabilistic forecasts of PV power generation.

The probabilistic forecast work in this thesis was started by investigating and characterizing the empirical distributions of point forecast errors. According to proposed empirical investigation, where cloud passages forecast is not available (a common situation), forecast errors do not obey any known parametric distributions. Hence, in order to avoid restrictive assumptions on

the shape of predictive distributions, a framework to generate nonparametric density forecasts is proposed. Density forecasts are assembled by predicting their quantiles for various nominal levels, in a quantile regression framework. The ELM [39] is used as a regression solver while a gradient-free optimization algorithm serves for parameter estimation, based on a suitable objective function that accounts for feasibility, reliability and, sharpness of predictive densities.

The empirical results in this thesis show that the proposed method is able to efficiently provide reliable and sharp marginal predictive densities for the very short-term (10-minute and one-hour lead times) PV generation.

1.2.3 Spatio-temporal forecasting contributions

In this part of the research, the aim is to investigate and analyse spatio-temporal dependency of PV generations. In this part of work, the following contributions are developed.

Multivariate predictive distributions of PV generation describing space-time dependency and time dependency are modelled.

In this work, the discriminating capability of the relevant scoring rules on performance assessment of space-time trajectories of PV generation is studied. The investigations cover the Variogram-based score as a recently proposed score with high discrimination ability and some former ones such as energy score. The scores which originally have been proposed for the case of time-dependency are modified and used for space-time dependency. Three events are proposed for the specific case of PV generations where PV power measurements for each time of day are compared with the maximum expected power (clear sky PV power) for the same time.

In order to track time dependency of PV generations for successive lead times while taking into account seasonally variations of sunrise and sunset times, observations and marginal distributions are transferred to a time grid and dependency modelling is carried out on this grid. Eventually, the generating trajectories are transformed back to the original space.

Investigations are performed to evaluate the goodness of fit of Gaussian copula in describing the joint characterization of PV power generation in time and space. The goodness of fit of Gaussian copula is compared with a number of Archimedean copulas namely Frank, Joe, Clayton and Gumbel in terms of Bayesian information criterion and Akaike information criterion.

The quality of generated trajectories from multivariate distributions are compared with those drawn from predictive densities, normal distributions centered on point predictions and

a generalization of climatology forecasts. The comparative performances of the forecast series generated by the mentioned methods are very important in operational problems which are sensitive to dependency information. For example, normal distributions centered on point predictions is the common practice in the literature to model the uncertainty in PV generation. The analyses conducted in the thesis can lead the reader to find out the extent to which systematic dependency modelling can be more informative and relevant for operational problems.

1.3 Objectives

The cornerstone of this research is to develop pragmatic approaches to deal with intermitency in PV generation more effectively. The primary objective here is to provide end-users with estimations of the future PV power to contribute to a secure and profitable power system operation. Because depending on the intended application, the preferred forecasting methods and the desired outcomes are different, our objectives here include contributions in three complementary contexts, namely, point and probabilistic forecasts and spatio-temporal analysis. The objectives of this thesis are summarized as,

- 1 Developing a statistical point forecast model which caters for short-term forecasts (one to 24 hours ahead). Forecasts of 1 to 24-hour ahead are very frequently used in many power systems operational problems like day-ahead and intra-day markets.
- 2 Devising an online statistical point forecast model for very short-term forecasts (less than one hour) and with high forecast periodicity. The predictions provided by this model can be used for frequent update and imbalance correction in order to alleviate frequency regulation costs in the power market and the need for increasing regulation reserves in power systems.
- 3 Because point forecasts can only inform about the conditional expectation of random variables, the main objective here is to forecast distributions (instead of single-valued prediction) for the likely future behaviour and realizations of PV generation. One can hardly find any thorough and exploratory statistical analysis and characterization of prediction errors for the case of PV generation. Therefore, in this thesis, as the fundamental part of uncertainty analysis, empirical distributions of PV generation prediction errors are estimated and analyzed.

- 4 Based on the outcome of the previous step, the type and family of the predictive distributions should be decided. If the empirical distributions show strange and nonparametric behaviour, then to avoid restrictive assumptions on the shape of predictive distributions, the objective would be to develop a nonparametric probabilistic model. Otherwise, the parameters of the decided parametric distribution (Gaussian, student-t, Beta, ...) should be predicted.
- 5 Benchmarks should be developed for performance evaluation of predictive densities. The evaluation criteria should allow for objective assessment of the quality and predictive capability of each predictive density.
- 6 An exploratory statistical analysis should be conducted for characterization of spatio-temporal correlation of PV generation. The strength of correlations should be explored to discourage or reinforce deploying space-time trajectory generation in decision-making.
- 7 It is aimed to develop a framework for modelling multivariate predictive distribution of PV generation and then generate space-time trajectories from that.
- 8 New scoring rules should be developed which are specially tailored to the PV generation case.

1.4 Organization of the Report

This thesis comprises six chapters and is organized as follows:

1. Chapter 1 provides a description of the problems related to the integration of the intermittent PV energy into the grid. The motivations for the research conducted in this thesis are also discussed. The objectives and the contributions of this research study have been introduced in the later part of this chapter. The contributions are categorized into three distinctive while related areas which are point and probabilistic forecasts and space-time correlation investigations.
2. Chapter 2 reports a brief literature review of the works related to this study. The concept of the point, probabilistic and uncertainty forecasts are introduced in this chapter. Examples are provided to justify the necessity of uncertainty forecasts. The young history of spatio-temporal analysis of renewable generations are reviewed and discussed.

3. Chapter 3 presents two point forecasting models for PV generation. ELM as the building block of the models is summarized first. A batch-learning-based forecasting model is presented to cater for several hours-ahead prediction. OSELM as the sequential learning edition of the ELM is reviewed. An online ensemble forecasting model based on OSELM is developed and presented. This is followed by empirical investigations to evaluate the efficiency and accuracy of the proposed models. Three case studies are introduced and the results are analysed.
4. Chapter 4 proposes nonparametric forecasting model for PV generation. The chapter starts with giving a brief introduction of the shortcomings of point forecasting and introducing uncertainty forecasts as a viable solution. It is followed by reviewing the fundamental concepts in probabilistic forecast context. The setup for nonparametric quantile and interval forecasts is discussed later on. The performance evaluation criteria for quantile forecasts are described. The formulation of the proposed quantile forecast approach is presented and it is followed by introducing four relevant benchmarks for comparative performance evaluation. In the last section of this chapter, the focus in one of the subsections is given to analysing the empirical distributions of the point forecast errors. The empirical distributions found by Kernel Density Estimation (KDE) are depicted and their stochastic features are explored. After justification of the nonparametric quantile forecasts, the empirical results for the proposed nonparametric framework as well as the benchmark approaches are provided and discussed. This chapter concludes by examining the potential benefits that normalizing PV generation with respect to the clear sky power can bring about.
5. Chapter 5 presents a framework for generation and evaluation of spatio-temporal trajectories of PV generation. The chapter starts by mentioning the decision-making problems which requires information about the development of forecast errors through successive lead times and the interdependent generation in contiguous locations. The experimental data for investigations is described. The derivation and evaluation of marginal predictive densities as the inputs of multivariate analysis are summarized. The frameworks for deriving the multivariate predictive density and generating space-time trajectories for the case of PV generation are proposed. The most relevant scoring rules for evaluation of the trajectories are introduced and three new event-based scores are proposed for the specific case of PV generation. Four most relevant benchmark approaches are developed in

the last section for comparative evaluation of the proposed methodology. The empirical results are analysed and discussed in the same section.

6. Chapter 6 concludes this thesis, and provides direction for future research.
7. List of authors publications, Bibliography and Appendices related to this research work are presented at the end of this thesis.

Chapter 2

Literature Review

2.1 Introduction

The fast growing integration of renewable energies with uncontrollable nature has increased the level of intermittency and uncertainty in power systems.

Deployment of solar energy as one of the highly intermittent types of energies is accelerating worldwide. However, high integration of solar energy technologies can endanger reliability of power grids unless their variable and uncertain nature is actively managed. Solar energy forecasting is a vital decision-aiding tool for either balancing authorities or market participants. Feeding high quality forecasts into scheduling and dispatch facilitates the operation of regulation services and ancillary markets [4].

Wind power forecasting is more developed than solar energy forecasting. It has taken almost a decade for researchers to obtain fairly reasonable forecast accuracy for wind power while there are many issues yet to be resolved. Solar forecasting, however, is still in early stages of research and market development, while it carries its own unique challenges. For instance the ramp rates in the output of PV installations are potentially much higher than those in the case of wind power. The reason is the lack of inertia in PV generation system while rotary components make some inertia for wind generation. In addition, a great share of PV power comes from distributed rooftop panels which are behind the meter and invisible to the ISOs. Also, the current rooftop PV installations do not have curtailment capability [4, 42].

In this chapter, it is tried to give an overview of the current status of PV power forecasting. The key factors in the relevance and accuracy of forecasting are discussed. To explain the

fundamental concepts and developments in the areas of probabilistic forecast and space-time correlation analysis, the available practices for wind energy are cited. The reason is that uncertainty forecast for PV power is currently at a very early stage. However, the state-of-the-art studies for either solar irradiance or PV power prediction are reviewed where available.

2.2 Uncertainty versus Variability

Uncertainty and variability are two commonly used terminologies in the area of forecasting. However, it is important to distinguish their true definitions. When a decision-maker ask for a range for a quantity like PV generation, it is essential for the forecaster to know whether a range for variability should be provided or a range for uncertainty [43].

2.2.1 Variability

Variability is related to heterogeneity or diversity of values a random variable has. It is “a quantitative description of the range or spread of a set of values [44]”. In simple words, variability refers to the case when a quantity is not homogeneous and has various values at different times, locations or instances.

For the case of PV generation, variability refers to the maximum generation limit in days with clear sky. The generation level of PV sites inevitably change throughout the day, the seasons and the year as the consequence of earth rotation around its axis. Maximum limit for output power of PV plants is estimated based on the clear sky solar irradiance. Clear sky solar irradiance has its maximum around noon and its minimum close to sunrise and sunset. The variations in the relative location of solar panels and sun, due to rotation of the earth, regularly causes 10-13% changes in the generation output of the PV sites within a period of 15-minute for a single-axis system [45]. Fig. 2.1 illustrates the clear sky PV power and observed PV output for a typical day from the case study described in Section 3.5. Due to reflections from the clouds and varying level of water vapor in atmosphere, in some instances the clear sky output power can be less than the observed ones [15].

Variability cannot be represented by a single value. It is generally quantified by the statistical moments such as mean, standard deviation and skewness. Variability also cannot be reduced by research because it describes the inherent diversity in a quantity. For instance, PV power generation is dependent on clear sky solar irradiance which we have no control over it, with the

Figure 2.1: Clear sky PV power along with observed PV output for a typical day from the case study (*Site A*) described in Section 3.5.

current technology available.

2.2.2 Uncertainty

Uncertainty refers to the lack of knowledge, incomplete understanding or non-precise measurements of a stochastic quantity of interest. Therefore, uncertainty can be reduced by research if it leads to more precise measurements or better characterization and estimation of the underlying stochastic behavior of a quantity. Uncertainty is represented mainly by a probability distribution which describes the degree of our knowledge about the probability of what the true value of a quantity is [45].

For the case of PV power, uncertainty is mainly caused by cloud passages. The critical challenges regarded high integration of renewable energy sources is mainly related to their uncertainty not variability. Cloud passages can cause even more than 60% drop of the peak isolation within few seconds. The duration which takes for the clouds to shade a PV site entirely depends on the PV site size and its geographical spread, cloud speed, cloud height and type, etc. For a PV plant with nominal capacity of 100 MW, the time is in the order of minutes while for the smaller plants it is in the order of seconds.

In Fig. 2.2, the difference between variability and uncertainty of PV power generation for a PV zone is demonstrated.

Figure 2.2: Variability vs. Uncertainty: PV power measurements along day-ahead point forecasts for four typical successive days.

2.3 Data Used for Forecasting

Generally, two different categories of data can be used in every type of forecasting. The first category includes the measurements of related variables provided by the sensors installed at the site(s) of interest. Denote ψ_t as the set of observations used as the predictors in forecasting for time $t + k$ at time t . For instance, for the case of PV power forecasting ψ_t can be shown as,

$$\psi_t = \{y_t, y_{t-1}, \dots, y_{t-g}, H_t, H_{t-1}, \dots, H_{t-g}, \mathcal{T}_t, \mathcal{T}_{t-1}, \dots, \mathcal{T}_{t-g}, \dots\} \quad (2.1)$$

with g as the number of lags (past observations), y , H and $\mathcal{T}$ as the power measurements, humidity and temperate up to lag g . ψ_t can include as many as different relevant variables available like total precipitation, cloud cover, wind speed.

The second category involves the forecasted meteorological variables provided by either private or public weather forecasting agencies. The relevance of each category to be used as the input to forecasting depends on temporal and spatial scale of forecasting. Usually NWP are used as the meteorological forecasts.

The kind of data used for forecasting is directly related to forecast horizon and required temporal and spatial resolutions. For hour ahead and intra-hour (very short-time) PV power

forecasts ground measurements are necessary. Ground measurements of GHI are required for solar farms because PV power is primarily a function of GHI. In addition, air temperature, wind speed are required for cell temperature forecasts. Persistence method can use the PV power plant measurements or radiometer output to generate forecasts for seconds to few minutes. Statistical methods are modeled to use ground measurements to perform intra-hour and hour ahead forecasts. Ground sky imaging is one of the recently developed methods to provide data to improve intra-hour forecasts. This method works based on image processing and cloud tracking techniques. In ground sky imaging, irradiance is forecasted for the current cloud shadow and then the cloud shadow is moved forward in time based on cloud velocity and direction.

For forecast horizons of 1 hour to 6 hour ahead (short-time), the data provided by satellite imagery is considered as the best forecasting information possible. However, satellite imagery has lower spatial and temporal resolution in comparison with ground sky imaging. Therefore, satellite imagery is not suitable for intra-hour forecasts. For forecast horizons more than 6 hours, NWP always has been the most promising forecasting method.

2.3.1 Numerical Weather Prediction (NWP)

NWPs are generated based on meteorological and physical equations which forecast the transition of atmosphere from one state to another state, given the initial condition.

The accuracy of the initial condition plays a vital role in the accuracy of resultant NWP, because the atmospheric models, in fact, predict how the initial condition will evolve at various time steps. The initial values are derived from ground measurements, lower and higher atmosphere observations, radar and satellite data, lightening and oceanographic data [46].

Because global models are computationally extensive, currently there are only 14 providers for NWP worldwide. In order to make the computational burden of NWPs affordable, the resolution of global NWP is relatively low, with spatial resolution in the order of 40 to 90 km. NWP models run four or two times per day, depending on the provider. For example at 00 and 12 UTC (universal time coordinated). The outcome of models provide predictions for most of the meteorological parameters such as surface solar radiation and total precipitation with time steps of 3 to 12 hours for relatively long-time horizon of up to 360 hours. For forecast hours more than 6 hours, NWPs exhibit better accuracy than statistical methods [41, 47, 48].

2.3.2 Ensemble forecasts

One of the potential approaches to improve PV power forecasting is to benefit from NWP in an ensemble-based framework. Ensemble forecasts are generated by multiple runs of NWP models, which differ in the physical parametrizations and initial conditions to address the uncertainty in measurements, initial conditions and model misspecification.

It has been reported that utilizing ensemble forecast can considerably improve the predictive performance of wind forecasting [49]. For the case of solar energy, among very few works available on meteorological ensemble predictions, in [50] small improvements in forecasting accuracy have been reported. The improvements are obtained by averaging the outcome of two or three NWP models.

2.4 Key Factors in Forecast Accuracy and Relevance

2.4.1 Geographical aggregation and uncertainty diversity

There has been always a share of uncertainty and variability in power systems due to non-constant load, outage of generators and lines, etc. However, recently the integration of renewables like wind and solar energy has dramatically increased the degree of intermittency and the need for flexible generation. A higher diversity in uncertainty and variability in the system makes managing the variability and uncertainty easier and cheaper. For instance, the aggregated generation profile of a number of PV plants spread over a large control zone is much smoother than that of an individual PV plant. The reason is that the intermittent variations in outputs of those plants can cancel out each other. Similarly, owing to the diversity of electricity usage among different consumers, the consumption profile of an residential costumer exhibits higher variability than the profile of aggregated load in the grid. In [51], it is shown that aggregation of wind generation over a wide geographical region improves forecast accuracy by 30%-50%. In [52], the aggregated irradiance for a fleet of 16 installations of Global Horizontal Irradiance (GHI) sensors is found to be much lower than that of a single location. It has been shown that the total relative power output variability drops quasi-exponentially as a function of dispersion factor of the PV fleets and it can be further decreased if the fleets are optimally spaced. This implies that a larger balancing area comparing to a smaller one can accommodate a higher amount of variability and uncertainty and it requires relatively less balancing services.

Both generation pools and generation diversity reduce the required reserve services to deal with net load fluctuations [51].

2.4.2 Forecast horizon

As it is expected when temporal resolution is the same, the forecasts accuracy for shorter look-ahead times is higher than that for the longer forecast horizons. In [3], the forecast performance for a 5-day period using persistence cloudiness for 5-minute and 60-minute look-ahead times are compared. It is reported that standard deviation of prediction errors for the 5-minute lead time is 1.1% of the nameplate while that for a 60-minute lead time is 3% of the nameplate. In [53], intra-hour solar irradiance forecasting is performed by predicting cloud motions and locations. Cloud locations are predicted for 5-minute ahead and simulations showed that for coastal areas, the cloud forecast error increases as the forecast horizon increases. Therefore, power markets which work closer to real-time can benefit from higher skilled predictions. Intra-day and real-time markets provide higher potential to accommodate intermittent generation. To gain an insight on the required flexibility and generation mix to deal with high integration of renewable energies, it is required to study the variability and uncertainty of net load in various time frames of interest, e.g. second, minute, hour [51]. Table 2.1 summarizes multiple intermittency timescales and the corresponding affected operational problems [4].

Table 2.1: Potential power system operational problem affected by various intermittency timescales

Timescale	Affected Operational Problem
Seconds	Power quality (e.g. voltage flicker)
Minutes	Regulation reserves
Minutes to hours	Load following
Hours to days	Unit commitment and scheduling

2.4.3 Regional climate

We should not expect to obtain the same PV power or solar irradiance forecast accuracy for locations with different climate types, or for different weather conditions at a particular location. Investigations carried out in [54] for 26 ground measurement stations in Spain over the course of a year (2005), have shown that the forecast accuracy is significantly dependent

on the local climate. Based on the results, much lower forecast accuracy should be expected for the solar installations at the northern Spain than the central or southern regions. In line with that, [55] has reported that *RMSE* values associated with forecasts for central European stations are higher than those reported for Spanish stations.

Generally, forecasting is less challenging in regions with more sunny days [54, 56]. In [56], it has been discussed that solar elevation is also a key factor in determining the forecast accuracy, with a lower solar angle leading to a lower accuracy [54].

2.5 Solar Power Point Forecasts

When a forecast $\hat{y}_{t+k|t}$ issued at time t for time $t+k$ is provided in the form of a single value, it is referred to as point forecast. $\hat{y}_{t+k|t}$ is the conditional expectation of the random process Y_{t+k} . In other words, $\hat{y}_{t+k|t}$ is the average of all values that may happen for Y_{t+k} . Therefore, $\hat{y}_{t+k|t}$ is represented as [18]:

$$\hat{y}_{t+k|t} = \mathbb{E}[Y_{t+k}|g, \Theta_t, \Omega_t] \quad (2.2)$$

with Θ_t as the model parameters, g as the model structure, and Ω_t as the set of predictors.

Here, PV generation point forecasting is briefly reviewed.

When deciding the most proper point forecasting technique, three important factors should be considered. Those are forecast horizon, the geographical spreading (a single PV site or available aggregation of a large number of PV zones) and available data for model specification.

Solar energy point forecasting can broadly be divided into two groups, physical and statistical. In [57, 58], physical and statistical techniques are compared and the statistical methods are found to be slightly superior. In practice, it has been suggested that statistical approaches tend to outperform physical models for short-term forecasts with lead times less than 6 hours.

2.5.1 Physical approaches

Physical approaches generally model physical and dynamical models of atmospheric flows using NWP models or sky images. Physical models deploy weather data such as irradiance in the plane of the array, temperature at the back of the cell, tilt angle and manufacture

specifications [59, 60].

Neither of the cell temperature or irradiance in the plan of the array is given by weather forecast agencies. Therefore, the relevant weather forecasts, e.g. global solar irradiance and ambient temperature forecasts can be used along with the module manufacture specifications to find the required information for physical forecasting. Forecasts are generally provided for a single PV system. In case that forecasts for a large aggregation is required, upscaling methods should be deployed [60].

2.5.2 Statistical approaches

The pure statistical techniques mainly rely on using data (measurements and/or NWP) to train a prediction model. In fact, the statistical methods do not rely on PV system models and their specifications. Time-series, regression and artificial intelligence are the mostly used tools to identify the patterns in data and provide statistical predictions.

The persistence method is the most basic and well-known statistical method. In persistence method, the current generation level is extrapolated into the future with only minor adjustment due to variations of sun angle. While the persistence method is competitive for very short forecast horizons, its accuracy degrades significantly with the increase in forecast horizon.

For forecast horizons of less than few hours, statistical techniques are very competitive, even without using exogenous data and by relying on power output measurements only [61]. Their predictive performance can greatly improve when using exogenous data such as NWP, sky imagery and satellite data. Deploying NWPs as the explanatory variables can improve the performance of statistical methods for longer look-ahead times, e.g. more than 6 hours [41].

Autoregressive (AR), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) as conventional regression, k-nearest neighbours, Artificial Neural Networks (ANN) as the pattern recognition techniques and hybrid method (combining ANN and optimization) have been widely used for solar energy prediction. A comprehensive review on their history and performance is provided in [6, 62].

Lorenz *et al.* proposed an approach for regional PV prediction in Germany based on forecast up to three days ahead provided by the European Centre of Medium-Range Weather Forecasts (ECMWF) [48]. The results have shown a 36% *rRMSE* for the first forecast day for a single PV site. The *rRMSE* increases to 46.3% for the third day. Interestingly, taking account a larger

area shows much higher forecast accuracy, around 13.4%, for the first day. This happens as a result of smoothing effects of spatial averaging due to uncorrelated solar irradiance attenuation in sparse locations.

In [63], two different types of fuzzy systems are proposed for PV power prediction and the results are compared with AR and ARIMA methods. This paper also discusses the uncertainty bounds around point predictions which in fact are the expected values of PV power. The uncertainty bounds can be considered as confidence intervals for conditional expectation of PV generations.

ANN and fuzzy theory have been employed in [43] to forecast hourly insolation with 24 hours lead time. The amount of clouds in the sky, humidity and temperature are used as predictors. Also, to alleviate forecast errors, a correction factor for insolation forecast is calculated using historical data. For the specific database used in this work, the forecast errors have increased in the afternoon.

An online solar irradiance forecasting method is proposed in [64] where Radial Basis Function (RBF) networks along with growing and pruning strategies of the minimal resource allocation technique have been used to predict hourly solar irradiance. Hourly forecasted values for weather variables, the number of the day in the year, the hour of the day and temperature have been considered as explanatory variables.

The ANN is employed in [65] to estimate solar potential based on geographical and meteorological data such as latitude, longitude, mean temperature and sunshine duration.

To predict 24-hour ahead solar irradiance, Mellit *et al.* proposed a multilayer perceptron model with the average of solar irradiance and air temperature as inputs [66]. The authors have employed K -fold cross validation to estimate the generalization capability of their proposed method over an independent dataset.

The above mentioned works mainly have focused on gradient-based algorithms such as back-propagation. However, these methods suffer from local optima and iterative learning process which make them slow and computationally expensive. The theory of Extreme Learning Machine (ELM) has been recently proposed and attracted a lot of research interests due to its capability to overcome the mentioned shortcomings [67]. The performance of the ELM is compared with some of the most efficient learning methods such as support vector machines in [67] and [68]. It has been shown that ELM can provide good generalization performance with several times faster learning mechanism comparing to SVM or BP-based methods.

Physical and statistical methods can be combined to get a better predictive performance. Ref. [69] is an example where a neural network-based reforecasting method is developed to improve forecast accuracy of (a) a physical model based on cloud tracking, (b) auto regressive moving average and (c) k -th nearest neighbor methods for forecast horizons of 15 minutes and less.

2.5.3 Performance evaluation

For point forecasting, ultimately the forecast accuracy is judged based on the difference between the predictions and the true measurements over a specified time. Here, four standard error metrics for point forecasting are concerned [41].

- *MAPE*: Mean absolute percentage error

$$\text{MAPE} = \frac{1}{N_T} \sum_{t=1}^{N_T} \left| \frac{y_t - \hat{y}_t}{\bar{y}_t} \right| \quad (2.3)$$

- *MSE*: Mean square error

$$\text{MSE} = \frac{1}{N_T} \sum_{t=1}^{N_T} (y_t - \hat{y}_t)^2 \quad (2.4)$$

- *RMSE*: Root-mean-square deviation

$$\text{RMSE} = \sqrt{\frac{1}{N_T} \sum_{t=1}^{N_T} (y_t - \hat{y}_t)^2} \quad (2.5)$$

- *MAE*: Mean absolute error

$$\text{MAE} = \frac{1}{N_T} \sum_{t=1}^{N_T} |y_t - \hat{y}_t| \quad (2.6)$$

with N_T as the number of measurements/forecasts pairs, y_t and $\hat{y}_t$ the observation and forecast for time t , respectively.

2.6 Probabilistic Forecasts

Probabilistic forecasts are the most general form of describing uncertainty and can be represented in the form of predictive distributions or a more straightforward form, prediction

intervals. Probabilistic forecasts provide quantitative information on the uncertainty associated with each single forecast for the future. This can provide substantial information for the decision-maker comparing to point forecasts which are only the conditional expectation of the random variable of interest. Probabilistic forecasts are critical in cases where the forecast errors are large and/or the likely consequences of making forecast errors are significant [5].

2.6.1 A decision-making example on the application of probabilistic forecasts

The emphasize in this thesis is on issuing probabilistic forecasts rather than point forecasts. In this section, a simple example is discussed to demonstrate the necessity of probabilistic forecast in decision-making [1, 25, 70]. For solar/wind power owners, finding the most optimal contract level to manage the associated risk with uncontrollable generation is highly challenging [71]. In this example, it is shown that the optimal contract level is one of the quantiles of predictive distribution of wind/solar power.

To make the example straightforward, it is assumed that the owner of wind/solar plant is price-taker in the forward market. This is a reasonable assumption owing to much smaller capacity of wind/solar plants comparing to the total capacity of the market. It is assumed that the imbalance prices can be represented by their expectations, e.g. their historical averages. ISO is responsible to clear the day-ahead and real-time markets. Renewable power owners participate in the market under the same regulations like the other participants.

Renewable owners offer to provide a constant power P_{t+k}^b at price π_{t+k} (\$/MWh) continuously over an interval of one hour starting from $t+k$ in the day-ahead market. Here, k represents the look-ahead time. In the PJM day-ahead market, participants have to submit their bids for the period starting from 00:00 am to 12:00 pm of the next day, before 12:00 pm every day. Therefore, for wind/solar energy owners, 12- to 36-hour ahead predictions play a vital role in setting out their bidding strategies. All participants have to take responsibility for their upward or downward deviations from their submitted bids. The total revenue Π for a wind/solar owner over an one-hour period can be formulated as [1],

$$\Pi(P_{t+k}^b, P_{t+k}^a) = \begin{cases} P_{t+k}^a \pi_{t+k} - (P_{t+k}^b - P_{t+k}^a) \hat{\pi}_{t+k}^-, & P_{t+k}^b \geq P_{t+k}^a \\ P_{t+k}^a \pi_{t+k} - (P_{t+k}^a - P_{t+k}^b) \hat{\pi}_{t+k}^+, & P_{t+k}^b < P_{t+k}^a \end{cases} \quad (2.7)$$

with $\hat{\pi}_{t+k}^-$ and $\hat{\pi}_{t+k}^+$ as expectations of the ex-post settlement prices for negative and positive

imbalances. P_{t+k}^a is the actual power production while P_{t+k}^b is the bidding amount. Solar/wind generation is considered a random variable with F_{t+k} as its cumulative distribution function (CDF) for time $t+k$. A typical CDF for generation at time $t+k$ is shown in Fig. 2.3 and an arbitrary value for bidding amount is selected and marked. Fig. 2.3 informs that $\Pr(P_{t+k}^b \geq P_{t+k}^a) = \alpha$, hence, $\Pr(P_{t+k}^b < P_{t+k}^a) = (1 - \alpha)$.

Figure 2.3: A typical CDF for wind/solar production.

If the contract level is decided based on point forecasts, then the maximum revenue is obtained in case of perfect forecast with zero error, i.e., $P_{t+k}^b = P_{t+k}^a = \hat{P}_{t+k}^a$, with $\hat{P}_{t+k}^a$ as the point forecast available. However, the probability of getting perfect forecasts is close to zero, because point forecasts are always subject to an error. Therefore, a rational strategy for the solar/wind owner is to maximize its “expected” revenue, considering the existing uncertainties. This can be written as

$$P_{t+k}^{*,b} = \arg_{P_{t+k}^b} \max \mathbb{E}[\Pi(P_{t+k}^b, P_{t+k}^a)] \quad (2.8)$$

where $P_{t+k}^{*,b}$ is the optimal contract level.

Maximizing the utility of wind/solar owners (II) is equivalent to minimizing the expectation of imbalance loss. The loss function, L , is defined as,

$$L(P_{t+k}^b, P_{t+k}^a) = \begin{cases} (P_{t+k}^b - P_{t+k}^a) \hat{\pi}_{t+k}^-, & P_{t+k}^b - P_{t+k}^a \geq 0 \\ (P_{t+k}^a - P_{t+k}^b) \hat{\pi}_{t+k}^+, & P_{t+k}^b - P_{t+k}^a < 0 \end{cases} \quad (2.9)$$

The expectation of loss function can be calculated as

$$\mathbb{E}[L(P_{t+k}^b, P_{t+k}^a)] = \int_0^1 L(x - P_{t+k}^b) F_{t+k}(x) dx \quad (2.10)$$

Hence, the optimal bidding amount is the solution of the following stochastic optimization problem:

$$P_{t+k}^{*,b} = \arg_{P_{t+k}^b} \min \mathbb{E}[L(P_{t+k}^b, P_{t+k}^a)] \quad (2.11)$$

Solving (2.11) gives the closed-form solution for the optimal contract level as (for a proof, see [26, 72])

$$P_{t+k}^{*,b} = F_{t+k}^{-1}\left(\frac{\hat{\pi}_{t+k}^+}{\hat{\pi}_{t+k}^+ + \hat{\pi}_{t+k}^-}\right) \quad (2.12)$$

which is a quantile of the true CDF of wind/solar generation at time $t + k$. The nominal proportion of this quantile is dependent on the imbalance prices. The aim of the probabilistic forecast is to provide an appropriate estimation of F_{t+k} . It is to be noted that point forecasts give an estimation of the mean value of F_{t+k} which not necessarily is equal to $F_{t+k}^{-1}\left(\frac{\hat{\pi}_{t+k}^+}{\hat{\pi}_{t+k}^+ + \hat{\pi}_{t+k}^-}\right)$.

Similarly, Pinson *et al.* demonstrated that employing uncertainty quantification in the decision-making process of bidding strategies in the power market can increase the revenue [25]. In their work, the uncertainty information is presented in the form of the predictive distributions and the methodology is flexible enough to cater for various utility functions of the market participant. This approach is tested on a multi-MW wind farm in the Dutch electricity pool with the objective of revenue-maximization. For a more comprehensive formulation, one can refer to [70] where the uncertainty in imbalance prices is also discussed.

2.6.2 Two groups of probabilistic forecasts

The probabilistic forecasts can be classified into two groups as parametric and non-parametric approaches.

2.6.2.1 Parametric approaches

In parametric approaches, a predefined shape is assumed for the predictive densities. The parametric density can be identified by a few parameters, depending on the type of density selected. For example, a Gaussian distribution can be characterized by two parameters known as mean and standard deviation.

A fewer number of parameters to be estimated makes the estimation of parametric models simple and computationally inexpensive. A parametric distribution of a random process Y at time $t + k$ is defined as

$$Y_{t+k} \sim f(y; \theta_{t+k}) \quad (2.13)$$

with f as the selected parametric density and θ_{t+k} as the set of parameters to be estimated. Let $\hat{\theta}_{t+k|t}$ as the estimated parameters given information available up to time t . A probabilistic forecast is described as

$$\hat{f}_{t+k|t} \sim f(y; \hat{\theta}_{t+k|t}) \quad (2.14)$$

For parametric approaches, researchers have focused on two aspects, which are making assumption on the shape of density and estimation of parameters of density of choice.

- ***Assumption on the Shape of Density***

The assumption on the shape of density is the most crucial part of parametric approaches. The choice of assumption depends on many factors such as forecast range, horizon, geographical spreading. For renewable generation, Gaussian distribution has been the mostly used one while in many cases it is not the best assumption. For example, for the case of wind and PV generations, a Gaussian distribution cannot describe their double-bounded natures, allowing for limited generation levels between zero and the maximum capacity of the installation [5, 73].

Different shape assumption leads to significantly different outcome for the potential range of the quantity of interest. In Fig. 2.4, the empirical distribution of PV generation at 11:00 am is illustrated. One year worth of data from dataset described in Section 5.2, is used for this illustration. In addition to the empirical density, Normal, Beta, Log-normal, Weibull distributions are fitted to the data available. Maximum-likelihood fitting is used to estimate the parameters of the assumed distributions. All the mentioned distributions

are characterized by only two parameters, allowing for very fast estimation. As can be seen in this figure, the assumption of the density type is a fundamental part of the parametric approaches because different densities result in completely different estimation of the stochastic behavior of the random variable of interest, e.g. renewable generation. In addition, as one can easily deduce, none of the parametric densities illustrated in Fig. 2.4 can appropriately describe the uncertainty in PV generation for 11:00 am because it exhibits highly nonparametric statistics as it is characterized by the empirical density. A further study is required to investigate proposing mixtures of parametric distributions in the future, based on a hidden regime sequences.

Figure 2.4: Empirical distribution and four fitted parametric distributions for PV generations at 11:00 am.

Although not a particular preferred distribution is reported for solar energy, in the literature, one can find useful investigations for parametric estimation of other types of renewable energies specially wind energy. Ref. [74] can be mentioned as one of the basic works in this area in which it is assumed that errors follow a Gaussian distribution with zero mean and a given standard deviation. Gaussian distribution also has been used in [75] to describe wind speed forecast errors.

A statistical analysis of wind generation forecast errors is presented in [76] and the results are then used to find optimal rates of energy storage systems. In this work, it is argued that distribution of wind power forecasting errors exhibit large kurtosis and have fat-tails. Therefore, Beta distribution can be a more appropriate parametric distribution than Gaussian distribution both for its large kurtosis and its limited range between zero and one [77, 78].

- ***Estimation of Parameters of Density of Choice***

The first parameter of parametric distribution is usually called location parameter. In the area of renewable energy forecasting, the location parameter usually is considered equal to point forecasts, especially for the case of Gaussian distribution.

The second parameter of the predictive distribution usually is called the scale parameter. Traditionally, in the AR model, the scale parameter is considered fix and equal to the variance of the white noise term [1]. However, due to complex atmospheric dynamics, both wind and solar energies are indeed non-linear and non-stationary stochastic processes. In order to avoid the assumption of constant variance, Autoregression-Generalized Autoregressive with Conditional Heteroscedasticity (AR-GARCH) has been proposed in which an autoregressive moving average model is utilized to estimate the variance of the white noise [1]. In [79], an ARIMA-GRACH approach is devised to predict multistep density forecasts for wind power generation as a non-Gaussian stochastic process. Forecasts for 15 minutes to 24 hours ahead are generated. In [80], a bivariate vector autoregressive moving average-GARCH framework with t-distribution is developed to characterize the underling uncertainty in wind speed and wind direction. Ref [81] has focused on Normal and student-t distributions to describe the highly volatile dynamics of offshore wind power for one-step ahead predictions of wind power measurements. Markov-switching Autoregressive (MSAR) model with GARCH is used to characterize predictive densities capturing autocorrelation and heteroscedasticity of offshore wind power time series. Arti-

ficial intelligence is another alternative approach to estimate the parameters of predictive densities [1].

2.6.2.2 Nonparametric approaches

Although parametric approaches are straightforward and require low computation cost, in many cases, they provide very poor estimation of the stochastic process of interest, as shown in Fig. 2.4.

In contrast with parametric approaches, non-parametric predictive distributions are generated without making any assumptions about their shape or moments. Therefore, they are referred to as distribution-free forecasts. When there is no pre-defined distribution type, non-parametric predictive densities have to fully characterize the estimated stochastic features instead of being represented by only few parameters. This makes them computationally much more expensive than their counterparts. In general, non-parametric predictive densities are composed of a set of quantiles, each with a pre-assigned nominal proportion.

Let Y as the stochastic process to be specified by a non-parametric CDF F as $Y_{t+k} \sim F_{t+k}$. F can be represented by

$$F_{t+k} = \{q_{t+k}^{\alpha_i}; 0 \leq \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m \leq 1\} \quad (2.15)$$

with $q_{t+k}^{\alpha_i}$ as the true quantile of the true distribution F_{t+k} at time $t+k$.

The aim of probabilistic forecasts is to provide an appropriate estimation of F_{t+k} , denoted by $\hat{F}_{t+k}$ which is constructed from a set of m predictive quantiles.

Although for PV power or solar irradiance, one can barely find reported practices addressing probabilistic forecasts, there is a comprehensive share of research available for non-parametric probabilistic forecasts of wind energy.

The adaptive resampling method proposed in [82] is one of the well-known nonparametric approaches which provide a framework to generate conditional prediction intervals of wind generation. The methodology is developed based on an empirical and nonparametric approach called adaptive resampling. Using this method, prediction intervals can be formed based on observed accuracy of the available point forecast method. Quantile regression, Kernel Density Estimate (KDE), and artificial intelligence-based methods are among the commonly used non-

parametric techniques. For quantile regression method typically NWP are among the mostly used explanatory variables [83].

- ***Quantile Regression***

The application of the quantile regression for prediction of quantiles is discussed in [84]. Local quantile regression is used in [83] to estimate quantiles of wind power directly. This method is called local because it tries to find a linear regression in the neighborhood of the predictors. This is performed by assigning a higher weight to the historical data which is closer and more similar to the current predictors.

In [85], a time adaptive quantile regression is proposed which works based on the simplex algorithm. The primary aim of this algorithm is to reduce the computational time of static quantile regression models. This has been pursued by developing an algorithm to generate new solutions based on old ones. With the aim of monitoring the impact of changing weather regimes, a set of variability indices are developed in [27] to be used as predictors for short-term probabilistic forecast of wind power. These indices are extracted from the wind power time series.

Generally, there are two drawbacks with quantile regression. First, for each quantile, a separate model should be trained. This leads to very high computational burden if a large number of quantiles is required. Crossing quantile is the other problem with quantile regression.

- ***Kernel Density Estimate***

KDE has been widely used for estimating predictive PDF of wind power. Through KDE, a smooth histogram of density of a random quantity can be found. The formulation of KDE is given in Appendix A. When deploying KDE, there are three challenges to be addressed. The first one is how to find the most relevant explanatory variables which exhibit high correlations with the quantity of interest. The second issue is related to find the proper kernel function for each explanatory variable. The last crucial issue is how to determine a proper bandwidth to avoid over fitting or under fitting [86].

Usually those variables provided by NWP which show high correlation with the generation level (wind or PV) are chosen as explanatory variables. The choice of kernel function is also dependent on the type of explanatory variables. In [87–89] four types of functions are proposed for wind energy, depending on whether the corresponding variables are bounded,

unlimited or periodic. Table 2.2 summarizes the recommended kernel functions. Further investigations are required to check the relevance of functions given in this table for the variables related to solar energy.

The accuracy of KDE is highly dependent on the optimality of the bandwidth. A too large bandwidth results in an over-smooth density while a small bandwidth causes under fitting. In [86], an approach to find the optimal bandwidth based on wind speed data is proposed. Mean Integrated Squared Error (MISE) is employed in [87] as the optimality criterion to find an appropriate bandwidth for KDE of the case of annual variation of wind.

Table 2.2: Kernel functions for different random variables [1]

Stochastic Process	Example	Kernel Function
Bounded between 0 and 1	Normalized wind/PV power	Beta
Bounded between 0 and $+\infty$	Wind speed	Gamma
Unbounded	Temperature	Gaussian or bi-weight
Circular or periodic	Wind direction	Von Mises

- *Artificial Intelligence*

The training criterion in ANN can be designed such that it enables the ANN to generate a set of predictive quantiles. In [90], a hybrid intelligent algorithm (named HIA) is proposed to find nonparametric prediction intervals for wind generation. However, the evaluation criteria and objective function proposed by this paper verify only the coverage and the width of the prediction intervals. The quality of quantiles which is essential for any reliable probabilistic forecast is ignored. An Extreme Learning Machine (ELM)-Bootstrap (named BELM) method is developed in [91] to estimate the statistical properties of the wind generation and predict intervals for three forecast horizons. This method stands on the normality assumption for distribution of forecast errors. In applications where the distribution of forecast errors does not have a normal shape, its accuracy should be verified.

2.6.3 Prediction intervals versus confidence intervals

In the area of statistical analysis of power systems, the term “confidence intervals” has been used for many years. “Prediction interval”, though, is more recent. While there is a clear

distinction between these two, they are often confused.

In the most general case, a confidence interval describes a range for the mean of a variable quantity, e.g. PV generation. For instance, suppose there is one year worth of data available for PV generation at 9:00 am. If one randomly samples from data and calculates the 95% confidence interval to be between 30 MW to 40 MW, it indicates that one can be 95% confident that the mean of PV generation lies within this range. When it comes to PV power forecasting, a confidence interval associated with a point forecast $\hat{y}_{t+k|t}$ determines a range for the mean of probability distribution of the expected PV generation at time $t + k$. Hence, it does not mean that 95% of the future PV measurements will lie within that interval.

Alternatively, a prediction interval determines a range that an individual new measurement will fall within it with an specific probability. For example, a 95% prediction interval for time $t + k$ gives a range that one can be 95% confident that the measured (true) PV generation at time $t + k$ will fall in.

A prediction interval is necessarily wider than its corresponding confidence interval due to additional uncertainty associated with prediction of a single measurement versus prediction of the average of likely values [92].

In this thesis, the target in probabilistic forecasting is to generate quantile forecasts not prediction intervals directly. However, central prediction intervals are the natural by-product of the quantile forecasts. Central prediction intervals refer to prediction intervals centered on the median of the predictive distribution. For a central prediction interval, there is equal probability for a non-covered measurement to be below or upper the interval limits. Therefore, for instance, a 80% prediction interval consists of quantile 10% and quantile 90% as its lower and upper bounds. It is to be noted that central prediction intervals are different than the traditional prediction intervals which were assumed to be centered on point predictions [5].

2.6.4 Performance evaluation

Two required properties associated with probabilistic forecasts are reliability (also referred to as calibration) and sharpness. For a calibrated quantile the observed coverage should be close to its nominal coverage. As discussed in Subsection 2.6.1, in the optimal bidding problem as one of the important decision-making problems in power systems, the optimal contract level is a quantile of the distribution of generation. Therefore, any miscalibration in the predictive

distribution leads to sub-optimal solutions because it gives an inaccurate quantile for a desired nominal coverage ($\alpha = \frac{\hat{\pi}_{t+k}^+}{\hat{\pi}_{t+k}^+ + \hat{\pi}_{t+k}^-}$). Sharpness also is a crucial requirement for high quality forecasts and it is referred to the width of predictive densities. In [93], a comprehensive study of assessment criteria and required properties for non-parametric probabilistic wind generation forecast is presented.

- **Reliability**

Reliability of a predictive density is generally decided based on reliability of its quantiles. Perfect reliability is the case when the empirical coverages of quantiles are equal to their corresponding nominal coverage [94]. The empirical coverage of a quantile forecast is calculated by comparing the measurements with the quantile value to find the proportion of measurements which have value less than that specified quantile. Similarly, the empirical coverage of a prediction interval is calculated based on the proportion of the observations which falls in that interval. In Fig. 2.5, two 80% prediction intervals from two different predictive densities are illustrated. Both prediction intervals are perfectly calibrated as they envelop eight out of ten observations. In practice, however, ideal calibration can hardly be obtained. Therefore, the goal is to generate prediction intervals with observed coverage as close as possible to the desired coverage. Reliability can be visualized by Q-Q diagram or probability Integral Transform (PIT) [93,94].

- **Sharpness**

Sharpness determines the concentration of a predictive density. In practice, narrower densities provide more information for decision-making. The reason is that making decisions based on a smaller range is more straightforward. Comparing two distributions illustrated in Fig. 2.5 can give a better understanding of the concept and importance of sharpness. In Fig. 2.5 while both distributions have ideal reliability, 80% prediction interval in one is wider than the other one. Decision-making is going to be more challenging by giving a wider range for the same reliability level. Sharpness can be examined by diagnostic tools such as box plots or by calculating the average width of the prediction intervals [26,93,94]. However, in theory and practice, it is more common to verify sharpness along with calibration through probabilistic skill scores [94].

Figure 2.5: Two typical predictive densities along with their corresponding 80 % prediction intervals.

2.6.5 A review on state-of-the-art on probabilistic forecasts of PV Power

Improving PV power forecasting methods, as an essential component to mitigate PV power intermittency, has attracted increased attention over the last few years. The online short-term PV forecasting using a clear sky model described in [15] comprises a notable example of a relevant approach.

In contrast to this vast amount of work on probabilistic wind power forecasting, one can hardly find proposals and empirical investigations for the solar energy case. Lorenz *et al.* [48] proposed a method to obtain 95% prediction interval 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 their work, standard errors follow normal distributions and prediction intervals are central on point forecasts.

In [95], a hybrid method is developed to provide prediction intervals for one-minute average direct normal irradiance for three intra-hour forecast horizons. Prediction intervals with three

coverage rates are obtained and compared with those given by bootstrap-ANN and persistence methods. A heuristic formulation is used to evaluate and compare prediction intervals. Moreover, prediction intervals are obtained independent of central quantiles and the quality of the quantiles is not checked and concerned.

A proposal for a geostrophic wind dependent probabilistic solar irradiance forecasts can be found in [96]. By post-processing NWP and comparing the North American Mesoscale model with the California Irrigation Management Information System observations and SolarAnywhere data, the model generates intervals for global horizontal solar irradiance. In the mentioned paper, only 95% prediction interval is sought and is compared with few methods including the method proposed in [48] to check if it can provide narrower intervals. Probabilistic skill scores are not calculated and the evaluation of intervals includes width and reliability metrics separately where the widths of the intervals are compared visually.

2.7 Spatio-temporal Correlation of PV Generation

Probabilistic forecasts are generally generated on per site, per look-ahead time basis and so their resultant predicative distributions are called marginal distributions. Although marginal distributions carry substantial information about the random process at each location or time individually, they do not inform about the interdependence structure of forecast errors at different locations or the evaluation of forecast errors in time. Hence, in those operational problems which are temporally or specifically coupled, probabilistic forecasts by neglecting the dependency structure lead to suboptimal solutions.

Owing to the complex meteorological mechanisms, PV generations exhibit high temporal and spatial dependency. Figs. 2.6 and 2.7 give examples of space and temporal correlations of PV generations. The figures are plotted using the data described in Section 5.2 for PV generations from 9:00 am to 5:00 pm. One year worth of data from April 2012 to April 2013 is used for illustration. Fig. 2.6 reveals high correlations in space for PV generations. Also, as can be seen in Fig. 2.7 there are noticeable correlations between PV generations in successive hours. Moreover, the shorter interval between two times is, a higher correlation is expected.

To describe PV generations at a number of locations and lead times, joint or multivariate distributions can be constructed. Joint densities are often characterized by spatio-temporal trajectories, also referred to as scenarios or ensembles. The reason behind preferring trajec-

Figure 2.6: Space correlations between PV generations at three PV zones described in Section 5.2.

tories over full multivariate distributions is that in many cases those densities do not have a straightforward analytical structure and deploying them in decision-making is challenging [29]. In contrast, trajectories are preferred by the forecast users because they can be easily incorporated in traditional optimization frameworks. Trajectories are generated by random draw from the joint predictive distributions.

The cornerstone of available practices on multivariate analysis of renewable energies is to use marginal distributions given by probabilistic forecasts as infeed and couple them using copulas to form a multivariate distribution [97].

In [31], an approach is proposed to capture and model the time dependent structure of wind power generation for successive hours using marginal densities. With an Gaussian copula assumption, the covariance matrix carries information about temporal dependency. Due to

Figure 2.7: Time correlations between PV generations at PV zone 1 described in Section 5.2.

non-stationary characteristics of wind power, an adaptive and recursive method is proposed to track dependency and estimate the covariance matrix for each time. The developed idea is then tested on a multi-MW wind farm with historical data available for a course of two years.

In [98], under Gaussian assumption, space-time dependency of wind generation is described by precision matrix when boundary points and direction-dependent cross-correlations are taken into account. In the same reference, by taking advantage of sparsity of precision matrices, a method for parametrization of high dimensional precision matrices is proposed. The method is then evaluated on a database including 15 areas and 43 lead times.

In [97], a basic framework is discussed for using copulas to model dependency and relative behaviour of random variables in power systems where the focus is given to Normal Copula. It has been shown that even with the same marginal distributions, different dependency structures result in distinct aggregation of uncertainties in the system.

2.7.1 Performance evaluation

The predictive performance is generally verified by scoring rules. Each scoring rule is aimed to assess a number of quality aspects and provide a numerical score representing the overall performance of forecasts. The scoring rules examine the ability of forecasts in imitating the real stochastic process of interest. The scoring rules are expected to be proper, meaning that they give the best score to the perfect forecasts.

While a range of quantitative assessment frameworks for univariate quantities and independently generated marginal densities exists [94], only few frameworks for the case of multivariate quantity evaluation can be found in the literature. Energy score is the most commonly used scoring rule to evaluate multivariate densities which are described by a finite number of samples. However, this score does not discriminate the misspecified dependency structure between components of a multivariate quantity [99, 100]. A variogram-based score is a recently proposed scoring rule and it is claimed to be more sensitive to misspecified mean, variance and correlations [100].

An event-based scoring rule is proposed in [99] as a diagnostic approach to assess the correspondence of trajectories generated on a multivariate base and related measurements. Frameworks are provided to evaluate time-dependent trajectories in predicting gradient and long-lasting events.

2.7.2 A review on state-of-the-art on spatio-temporal analysis of PV power

Just recently, few practices have studied space-time correlations of PV power (or solar irradiance) and tried to benefit from them in point forecasting. Gueymard and Wilcox [101] have presented a general investigation on long-term variability of solar resources in the United States. Yang *et al.* have proposed a statistical approach to obtain temporal and spatial stationarity for solar irradiance time series logged in few numbers of monitoring stations in Singapore [102]. To do so, solar irradiance time series at individual sites are detrended to get temporal stationarity.

Spatial stationarity also is obtained by coordinate transformation. Using time forward kriging and with respect to the persistence method, 25 % improvement in *RMSE* is reported.

In [33], spatial cross-correlations between all the PV sites under consideration are reported to be more than 0.85. In [34], almost the same results are shown where the cross-correlations between each pairs of 12 PV zones with distance up to 1500 km in Sweden are found to be more than 0.8. These findings have supported the idea of using data from neighbouring sites as additional explanatory variables for PV generation forecasting at the target location [35].

Chen Yang *et al.* have looked into space-time correlations of solar irradiance to devise forecast models for seven locations in California with one-, two- and three-hour forecast horizons [35]. To do so, firstly, the regions which present high correlations between satellite-derived data and ground data are determined. Then, the data for those areas are employed as exogenous variables to predict global solar irradiance at the point of interest. To determine the most optimal time lags for local and exogenous variables, genetic algorithm is used.

In [36], to benefit from spatial-temporal dependencies of PV power, a vector autoregression based method is proposed. Using past observations from the neighbouring locations and for forecast horizons less than 4 hours, up to 10% improvements in *RMSE* values are achieved. In [37], also to enhance predictability of PV power, measurements from the adjacent PV sites are used as exogenous variables.

As reviewed above, the foundation of the few present studies on the spatio-temporal PV power forecasting is on deploying measurements or meteorological data from the neighbouring locations in forecasting process of the site of interest (spatial analyses). Moreover, as the developed methods use lagged data, they are categorized as temporal investigations [103]. In this thesis, though, spatio-temporal correlations of PV power are leveraged in a different way. In contrast with reviewed works in which dependencies are founded on the base of point forecasts, here, the dependencies are modelled based on probabilistic forecasts. The goal of this study is to provide more informative forecasts for probabilistic decision-making.

2.8 Conclusions

As reviewed in this chapter, enhancing forecasting approaches is a crucial measure which should be taken to tackle with intermittent generation. Therefore, in recent years several researchers have tried to develop forecast models with improved predictive performance. In the

realm of PV generation, the focus in most of the practices has been placed on point forecasting. Due to larger integration of wind in comparison with solar power, a larger body of literature has focused on wind energy rather than solar energy. This is more pronounced in the uncertainty forecast area. It is tried to benefit from the bulk of research available for wind energy to explain the required concepts.

In this chapter, the most important factors in forecasting are reviewed and summarized. The impacts of geographical diversity, regional climate and forecast horizon on the expected forecasts accuracy are explained. It is tried to explain the common confusions in the area of forecasting such as the confusion between variability and uncertainty, and between prediction intervals and confidence intervals.

A decision-making example is provided to clarify why it is argued here that probabilistic forecasts are more optimal inputs for decision-making than point forecasts. The fundamental concepts in point forecasting, probabilistic forecasting and spatio-temporal analysis are summarized and the state-of-the-art practices for the case of PV generation are reviewed.

Chapter 3

Batch and Sequential Forecast Models for Photovoltaic Generation

3.1 Introduction

The worldwide increase in the integration of photovoltaic generation has necessitated improvements in the forecasting approaches. In this chapter, two models are proposed to cater for PV generation forecasts for a few minutes to several hours look-ahead times. A very fast and accurate prediction model based on extreme learning machine is deployed for day-ahead prediction and is presented in this chapter. Moreover, an adaptive and sequential model is introduced for intra-hour and high frequency forecasting in the PV sites with highly intermittent generation and without availability of clouds forecasts. The efficiency of the proposed models are verified on three sites with different generation patterns and sampling resolutions.

Three datasets with one minute and one hour temporal resolutions are used for evaluation of the methods. The performances of the proposed methods are compared with those of some the relevant approaches in terms of the commonly used accuracy metrics.

3.2 Extreme Learning Machine (ELM)

The ELM is a recently introduced learning algorithm for single-hidden layer feed-forward neural network (SLFN) [67]. It provides a good generalization performance with several times faster learning mechanism in comparison with traditional algorithms such as backpropaga-

tion [39].

The reason of fast learning mechanism of ELM is that, in ELM, parameters for the hidden layer are randomly assigned while they do not need to be tuned through the training process. The output weights are analytically calculated at once without going through a lengthy iterative training.

Given a dataset of N samples $\{(\mathbf{x}_j, \mathbf{g}_j)\}_{j=1}^N$ with $\mathbf{x}_j \in R_n$ as input and $\mathbf{g}_j \in R_m$ as output, in a SLFN with $\tilde{N}$ hidden neurons and $\varphi(\cdot)$ as activation function, the relationship between inputs and outputs is defined by

$$\sum_{i=1}^{\tilde{N}} \beta_i \varphi(\mathbf{a}_i \cdot \mathbf{x}_j + b_i) = \mathbf{g}_j, \quad j = 1, \dots, N \quad (3.1)$$

where $\mathbf{a}_i = [a_{i1}, a_{i2}, \dots, a_{in}]^\top$ is the weight vector which connects input neurons to the neurons in the hidden layer; $\mathbf{a}_i \cdot \mathbf{x}_j$ represents the inner product of $\mathbf{a}_i$ and $\mathbf{x}_j$ and $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^\top$ denotes the vector of weights of output nodes.

The set of equations in (3.1) can be compacted as

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

where

$$\mathbf{H}(\mathbf{a}_1, \dots, \mathbf{a}_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, \mathbf{x}_1, \dots, \mathbf{x}_N) = \begin{bmatrix} \varphi(\mathbf{a}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & \varphi(\mathbf{a}_{\tilde{N}} \cdot \mathbf{x}_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ \varphi(\mathbf{a}_1 \cdot \mathbf{x}_N + b_1) & \cdots & \varphi(\mathbf{a}_{\tilde{N}} \cdot \mathbf{x}_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad (3.3)$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^\top \\ \vdots \\ \beta_{\tilde{N}}^\top \end{bmatrix}_{\tilde{N} \times m} \quad \mathbf{G} = \begin{bmatrix} \mathbf{g}_1^\top \\ \vdots \\ \mathbf{g}_N^\top \end{bmatrix}_{N \times m} \quad (3.4)$$

Given the assigned input weights $\mathbf{a}_i$ and biases b_i , training a SLFN can be translated to obtaining a least square solution for output weights $\hat{\boldsymbol{\beta}}$ which is the solution of $\mathbf{H}\boldsymbol{\beta} = \mathbf{G}$ as

$$\begin{aligned} & \| \mathbf{H}(\mathbf{a}_1, \dots, \mathbf{a}_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, \mathbf{x}_1, \dots, \mathbf{x}_N) \hat{\boldsymbol{\beta}} - \mathbf{G} \| \\ &= \min_{\boldsymbol{\beta}} \| \mathbf{H}(\mathbf{a}_1, \dots, \mathbf{a}_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, \mathbf{x}_1, \dots, \mathbf{x}_N) \boldsymbol{\beta} - \mathbf{G} \| \end{aligned} \quad (3.5)$$

In (3.5), if $N = \tilde{N}$ and the activation function is infinitely differentiable, $\mathbf{H}$ is a square matrix and invertible (N sample data should be distinct). The ELM can train the SLFN with zero error [67]. However, in most of the practical applications $N \gg \tilde{N}$, therefore, $\mathbf{H}$ is not a square matrix any more, and there may not exist a_i, b_i and β_i ($i = 1, \dots, \tilde{N}$) such that $\mathbf{H}\boldsymbol{\beta} = \mathbf{G}$. In this case, it can be proven that the minimum training error can be achieved by

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

where $\mathbf{H}^\dagger$ is the Moore-Penrose generalized inverse of $\mathbf{H}$ and it is generally derived by singular value decomposition [67]. Moreover, it can be shown that $\hat{\boldsymbol{\beta}}$ is also the minimum norm least square solution of the linear system in (3.2).

The flowchart of the ELM algorithm is given in Fig. 3.1.

3.3 Online Sequential Learning Algorithm

Although the ELM is a powerful batch learning algorithm, for PV generation forecasts with minutes or seconds temporal resolutions, an adaptive method with sequential updating capability is required to follow the chaotic and non-stationary PV power patterns. Therefore, a sequential learning algorithm named OSELM [39] is deployed. This algorithm has been developed based on the idea of ELM and like ELM needs no pre-defined control parameters except the number of neurons in the hidden layer. The OSELM can train with one-by-one or block-by-block sequential data. Moreover, the size of the sequential blocks can be fixed or variable. It takes advantage of the fast and accurate mechanism of ELM and caters for online applications.

In order to train a SLFN by the OSELM, the model should be initialized with a small block of training samples, N_0 where $N_0 \geq \tilde{N}$. The first phase of training in OSELM is similar to ELM-based batch learning with $\mathfrak{N}_0 = \{(x_j, \mathbf{g}_j)\}_{j=1}^{N_0}$ as the initial training set, and the algorithm tries to minimize $\|\mathbf{H}_0 \hat{\boldsymbol{\beta}} - \mathbf{G}_0\|$ with

$$\mathbf{H}_0 = \begin{bmatrix} \varphi(a_1.x_1 + b_1) & \cdots & \varphi(a_{\tilde{N}}.x_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ \varphi(a_1.x_{N_0} + b_1) & \cdots & \varphi(a_{\tilde{N}}.x_{N_0} + b_{\tilde{N}}) \end{bmatrix}_{N_0 \times \tilde{N}} \quad \mathbf{G} = \begin{bmatrix} \mathbf{g}_1^\top \\ \vdots \\ \mathbf{g}_{N_0}^\top \end{bmatrix}_{N_0 \times m} \quad (3.7)$$

Figure 3.1: The flowchart of ELM forecasting method

The initial output weights can be calculated by

$$\beta_0 = \mathbf{K}_0^{-1} \mathbf{H}_0^\top \mathbf{G}_0 \quad (3.8)$$

where $\mathbf{K}_0^{-1} = \mathbf{H}_0^\top \mathbf{H}_0$.

When a new block of data arrives, the new training set would be $\aleph_1 = \{(x_j, \mathbf{g}_j)\}_{j=N_0+1}^{N_0+N_1}$ with N_1 as the number of recent observations in the newly arrived block. N_1 can take any value

greater or equal to zero. The updated training problem is as follows

$$\left\| \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \boldsymbol{\beta} - \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \end{bmatrix} \right\| \quad (3.9)$$

where

$$\mathbf{H}_1 = \begin{bmatrix} \varphi(a_1 \cdot x_{N_0+1} + b_1) & \cdots & \varphi(a_{\tilde{N}} \cdot x_{N_0+1} + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ \varphi(a_1 \cdot x_{N_0+N_1} + b_1) & \cdots & \varphi(a_{\tilde{N}} \cdot x_{N_0+N_1} + b_{\tilde{N}}) \end{bmatrix}_{N_1 \times \tilde{N}} \quad \mathbf{G} = \begin{bmatrix} \mathbf{g}_{N_0+1}^\top \\ \vdots \\ \mathbf{g}_{N_0+N_1}^\top \end{bmatrix}_{N_1 \times m} \quad (3.10)$$

The updated $\boldsymbol{\beta}$ can be calculated by

$$\boldsymbol{\beta}_1 = \mathbf{K}_1^{-1} \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix}^\top \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \end{bmatrix} \quad (3.11)$$

where

$$\mathbf{K}_1 = \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix}^\top \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \quad (3.12)$$

By manipulating (3.12), we can get $\mathbf{K}_1 = \mathbf{K}_0 + \mathbf{H}_1^\top \mathbf{H}_1$ and following that

$$\begin{aligned} \boldsymbol{\beta}_1 &= \mathbf{K}_1^{-1} \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix}^\top \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \end{bmatrix} \\ &= \mathbf{K}_1^{-1} (\mathbf{K}_1 \boldsymbol{\beta}_0 - \mathbf{H}_1^\top \mathbf{H}_1 \boldsymbol{\beta}_0 + \mathbf{H}_1^\top \mathbf{G}_1) \\ &= \boldsymbol{\beta}_0 + \mathbf{K}_1^{-1} \mathbf{H}_1^\top (\mathbf{G}_1 - \mathbf{H}_1 \boldsymbol{\beta}_0) \end{aligned} \quad (3.13)$$

To generalize the procedure when $(k+1)^{\text{th}}$ chunk of data arrives, the training dataset can be shown by

$$\mathfrak{N}_{k+1} = \{(x_j, \mathbf{g}_j)\}_{j=(\sum_{i=0}^k N_i)+1}^{\sum_{i=0}^{k+1} N_i} \quad (3.14)$$

where N_{k+1} denotes the number of data items in the $(k+1)^{\text{th}}$ chunk and $N_{k+1} \geq 0$. The generalized updating formula for $\mathbf{K}$ and $\boldsymbol{\beta}$ are given by

$$\mathbf{K}_{k+1} = \mathbf{K}_k + \mathbf{H}_{k+1}^\top \mathbf{H}_{k+1} \quad (3.15)$$

$$\beta_{k+1} = \beta_k + \mathbf{K}_{k+1}^{-1} \mathbf{H}_{k+1}^\top (\mathbf{G}_{k+1} - \mathbf{H}_{k+1} \beta_k) \quad (3.16)$$

where β_{k+1} and $\mathbf{G}_{k+1}$ are as follows

$$\mathbf{G}_{k+1} = \begin{bmatrix} \mathbf{g}_k^\top \\ (\sum_{i=0}^k N_i) + 1 \\ \vdots \\ \mathbf{g}_{k+1}^\top \\ \sum_{i=0}^{k+1} N_i \end{bmatrix}_{N_{k+1} \times m} \quad (3.17)$$

$$\mathbf{H}_{k+1} = \begin{bmatrix} \varphi \left(\mathbf{a}_1 \cdot \mathbf{x}_{(\sum_{i=0}^k N_i) + 1} + b_1 \right) & \cdots & \varphi \left(\mathbf{a}_{\tilde{N}} \cdot \mathbf{x}_{(\sum_{i=0}^k N_i) + 1} + b_{\tilde{N}} \right) \\ \vdots & \dots & \vdots \\ \varphi \left(\mathbf{a}_1 \cdot \mathbf{x}_{\sum_{i=0}^{k+1} N_i} + b_1 \right) & \cdots & \varphi \left(\mathbf{a}_{\tilde{N}} \cdot \mathbf{x}_{\sum_{i=0}^{k+1} N_i} + b_{\tilde{N}} \right) \end{bmatrix}_{N_{k+1} \times \tilde{N}} \quad (3.18)$$

The flowchart of the OSELM algorithm is given in Fig. 3.2.

3.4 Online and Adaptive PV Forecast Model

Fig. 3.3 shows a diagram on the proposed adaptive forecast model. In the designed model (nOSELM), due to inherent online and very fast learning mechanism of OSELM, n number of OSELMs are working in parallel in a competitive and adaptive environment. It is expected that using multiple learning models in parallel offers a better predictive performance than that could be obtained from any of the constituent learning models individually [40].

In the forecast method, all parallel OSELM models are initialized with all the available historical data. The initialization for each of the parallel models can be performed by using a different subset of available dataset to make each of parallel models potentially suitable for a specific regime in the fluctuations.

After initialization phase, the model can start to generate forecasts and at the same time update itself based on the specific kinds of dynamic and fluctuations existing in that time. The parallel models will be updated by the M most recent data items, where M is the number of data items in the most recent measured block. The data collection rate can be set at the periodicity of once per minute or less. Once updating is done, the training accuracy of all parallel OSELMs

Figure 3.2: The flowchart of OSELM forecasting method.

are checked over the considered data block and the one with the best performance over the updating block will be responsible for performing PV power prediction for N -minute ahead of time. Different accuracy metrics such as $RMSE$ or MAE can be used to rate the models and select the best one. Any time, a chunk of newly measured data arrives, this process will be repeated.

The block size can vary from only one to several thousands and updating will be done in very few seconds. Notably, apart from this selection policy, one may consider the average or the median of the outputs of the parallel models as the final output for the ensemble model. Based on the performed experiments for the case of PV generation with data of minute temporal

Figure 3.3: The flowchart of the proposed adaptive forecast model.

resolution and where spot measurements (not average values) are taken into account, selecting the final output based on the minimum $RMSE$ results in more accurate predictions.

Furthermore, to deal with PV generation in newly established PV sites with not long archived data available for pre-training, online learning feature of the proposed model is a great asset since it can start prediction with only one day worth of data and continuously updates itself with the real-time measured data. The continuous update, in turn, makes the model suitable for the inherent non-stationary feature of PV generations.

3.5 Empirical Investigations

To verify the performance of the proposed method, time series of solar power generation at three sites with different climatic regions are used. The temporal resolution of the datasets for the first two sites is of one minute and that of the third site is of one hour.

Site A: This is a 433 kWp PV system at the St Lucia, the University of Queensland. To have a consistent forecasting duration for all days throughout a year, daytime data from 6:30 am to 5 pm is considered. There are 630 data points per day and 229,950 data points per year [104].

Site B: This is a 1 kWp PV system at Nanyang Technological University (NTU), Singapore. NTU is located in a tropical region close to the equator. The diurnal time for this case is assumed to be 7 am to 6:50 pm. Therefore, each day contains 711 data points.

Site C: It contains three adjacent solar farms. For this site there are 12 independent variables from the output of the NWP given by European Centre for Medium-Range Weather Forecasts (ECMWF). These variables are total column liquid water, total column ice water, surface pressure, relative humidity at 1000 mbar, total cloud cover, 2 metre temperature, surface solar

radiation down, surface thermal radiation down, net solar radiation at the top of the atmosphere, total precipitation, 10-metre U wind component, 10-metre V wind component [105].

For all sites, the data for 2012 is used for training, while that for 2013 is used for genuine forecast evaluation. Apart from seasonal and diurnal variations, at small temporal resolution and lead times, cloud passages are the main reason of PV power fluctuations, commonly causing drops of 60% of the installed capacity [4]. For Sites *A* and *B*, 10-minutes lead time with the forecast frequency of one minute is considered while in Site *C*, 1-24 hours lead times with 1-hour resolution are defined. For Site *C*, for each day, forecasts are performed at 00:00 midnight. Therefore, for example, the forecasts for 8:00 am have a lead time of 8 hours while those for 2:00 pm have a lead time of 14 hours and so on. Table 3.1 summarizes resolution of data, forecast horizon and predictors for each site.

Table 3.1: Data resolution, forecast horizon and predictors in three case studies

	Database Details	
<i>Site A</i>	Data resolution	1 minute
	Forecast horizon	10 minutes
	Predictors	Time series of PV power generation
<i>Site B</i>	Data resolution	1 minute
	Forecast horizon	10 minutes
	Predictors	Time series of PV power generation
<i>Site C</i>	Data resolution	1 hour
	Forecast horizon	1-24 hours
	Predictors	12 independent variables from the output of the NWP

Figs. 3.4, 3.5 and 3.6 show the point forecasts versus measured values for successive days, Sites *A*, *B* and *C*, respectively. As can be seen, there is an acceptable agreement between measured values and the predictions.

Table 3.2 gives the comparative results of ELM and nOSELN versus feed-forward backpropagation neural network (FFBPG) and persistence methods. The number of training samples for all three sites are given in the table. The number of training samples includes 87, 75 and 196 days worth of data for Sites *A*, *B* and *C*, respectively. Moreover, the same number of data items are used for genuine testing.

Figure 3.4: 10-minute ahead point forecasts for 6 successive days, Site *A* Figure 3.5: 10-minute ahead point forecasts for 6 successive days, Site *B*

The persistence method is chosen here as a benchmark because it is very hard to be surpassed when considering very short lead times [106]. For the persistence and the nOSELM, as online learning forecast models, simulation times are not mentioned in the table because for these methods, the training and testing phases are mixed. However, it is worth to note that the training time of the nOSELM is quite comparable with that of the ELM as it takes advantage of the similar learning theory. For Site *C*, as it needs day-ahead predictions, an offline batch training-based method can be used. Moreover, as the persistence method shows its best performance only for very short-term lead times, it is also excluded for Site *C*.

The 10-fold cross validation (refer to Appendix B.1) approach is deployed to tune the control parameters of all the models. To find the best number of hidden neurons for the FFBP, ELM

Figure 3.6: 1-24 hour ahead point forecasts for 8 successive days. Forecasts are performed at 00:00 each day, Site *C*

and nOSELN, different values ranging from 20 to 400 with 10-step increments have been checked by the cross validation. The number of neurons in the hidden layer for are selected equal to 200 and 50 for the ELM and OSELN, respectively. For FFBPG, Matlab ANN toolbox is used. The predictors for Site *A* and Site *B* are the past observations up to lag 10.

Moreover, to choose the best number of parallel models in nOSELN, values between 2 to 20 with 2 steps increments have been investigated. Based on the output of the cross validation and by compromising between accuracy and computational cost, 5 OSELN models are chosen to work in parallel. For each future time instances, the forecasted value given by the one which provides the lowest *RMSE* for the updating block is selected as the final forecasted value of the ensemble model.

Referring to Table 3.2, the accuracy of ELM is slightly better than that of the FFBPG and more importantly its required simulation time for the same number of training samples is about 1/60 of that of the FFBPG. This supports the argument regarding the very fast learning mechanism of ELM with respect to the iterative learning algorithms such as FFBPG. Comparing ELM with nOSELN, nOSELN outperforms ELM for both Sites *A* and *B*.

Interestingly, from Table 3.2, the proposed nOSELN method outperforms the persistence methods for both Site *A* and Site *B* in terms of all the mentioned accuracy metrics. The improvements in *RMSE* provided by nOSELN with respect to ELM and the persistence method is around 14% for Site *A*. For Site *B*, the improvements provided by nOSELN for the *MAE*

Table 3.2: Comparative results for point predictions (accuracy metrics in percent (%))

	Methods	MAPE	MSE	RMSE	MAE	Training Time (sec)	Number of Training Data
<i>Site A</i>	nOSELm	7.66	1.52	12.33	7.55	-	-
	ELM	9.11	2.05	14.33	9.06	2.6988	54984
	Persistence	8.57	2.07	14.39	8.49	-	-
	FFBPG	8.90	2.06	14.34	8.86	133.786	54984
<i>Site B</i>	nOSELm	5.63	0.84	9.14	5.58	-	-
	ELM	6.97	1.32	11.49	6.90	2.6208	53325
	Persistence	6.62	0.99	9.95	6.56	-	-
	FFBPG	7.21	1.33	11.52	7.14	120.2	53325
<i>Site C</i>	ELM	7.41	1.65	12.84	6.68	0.7800	14118
	FFBPG	7.72	1.78	13.33	7.53	66.7996	14118

metric are 15.5%, 15% and 21.9% in comparison with the ELM, the persistence and the FFBPG, respectively. By adding the clouds forecasts to the predictors' set used in the forecast model, it is expected that the proposed nOSELm model surpasses its current version pronouncedly. In the current case, with no availability of clouds forecasts, the learning model does not have enough clue about the behaviour, the features and the future locations of the clouds to update the model more precisely.

It should be emphasized that in all simulations where the temporal resolution of data is of one minute, the instantaneous measurements, not the average values, are taken into account. Needless to say, temporal averaging has smoothing effect on variability and if used it is expected to result in forecasts with higher accuracy (lower *RMSE*).

Generating random input weights and biases in ELM in some cases has been considered as an issue that may cause unstable performance. If this randomness leads to unstable performance in a particular application, the ensemble based model developed for online and adaptive PV forecasts in Section 3.4, can be a solution. The reason is that in this ensemble model a number of OSELm models which have different initial values (different input weight and biases) are working in parallel to provide reliable forecasts. The performances of all models are observed and evaluated continuously to find their shares in generating final forecasts.

In this thesis, in order to investigate the impact of random parametrization on the performance of ELM in the proposed PV forecasting applications, simulations are performed to generate forecasts using ELM for Site C in 500 independent runs. The *RMSE* values for all

Figure 3.7: Histogram of the RMSE values provided by ELM in 500 independent runs, Site *C*

independent runs are compared. The average and standard deviation of *RMSE* values are calculated as 12.664 and 0.273, respectively. Fig. 3.7 shows a histogram of calculated *RMSE* values. Looking at the histogram and taking into account the relatively low standard deviation allows us to conclude that the random parametrization has not led to an unreliable performance of ELM in the applications covered in this chapter.

3.6 Conclusions

In this chapter, two PV generation forecast models are successfully implemented and verified.

In this study, to enhance the capability of the forecast model in following different dynamics and ramp-rates in PV generation time series, an ensemble adaptive forecast models is proposed which works based on the OSELM learning theory. In the proposed model n number of OSELMs are working in parallel in a competitive and adaptive environment. It is expected that using multiple learning models in parallel offers a better predictive performance than that could be obtained from any of the constituent learning models individually. The predictive performance

of each member of the proposed ensemble model is observed and evaluated continuously to determine the share of each member in producing the final forecasts. In this chapter, a batch learning forecast model also is implemented for day ahead forecasting. This model uses ELM as the learning algorithm and NWP as the predictors.

If point forecasts are of interest, depending on the application in hand, ELM or nOSELN can be used. The nOSELN is preferred for online and real-time training, while ELM caters for longer lead times. Based on the results and discussions, the ELM algorithm can train the forecast model several times faster than traditional FFBPG and it offers a slightly better accuracy.

The results showed that the proposed nOSELN method outperforms the other verified methods including the persistence method which is a highly skilled prediction model for very short forecast horizons. It has been shown that the proposed nOSELN model not only has very fast learning mechanism but also can provide more accurate forecasts with betterments from approximately 15% to 22% in terms of different accuracy metrics.

Chapter 4

Very Short-term Nonparametric Probabilistic Forecasting of Photovoltaic Generation

4.1 Introduction

Due to the inherent uncertainty involved in renewable energy forecasting, uncertainty quantification is a key input to maintain acceptable levels of reliability and profitability in power system operation. A proposal is formulated and evaluated here for the case of solar power generation, when only power and meteorological measurements are available, without sky-imaging and information about cloud passages.

The proposed empirical investigation reveals that the distribution of forecast errors do not follow any of the common parametric densities. This therefore motivates the proposal of a nonparametric approach to generate very short-term predictive densities, i.e., for lead times between a few minutes to one hour ahead, with fast frequency updates. The idea is to rely on an Extreme Learning Machine (ELM) as a fast regression model, trained in varied ways to obtain both point and quantile forecasts of solar power generation. Four probabilistic methods are implemented as benchmarks.

Rival approaches are evaluated based on a number of test cases for two solar power generation sites in different climatic regions, allowing us to show that the proposed approach results in

generation of skilful and reliable probabilistic forecasts in a computationally efficient manner.

4.2 Experimental Framework

As a basis for the empirical investigation in this study, the time series with solar power generation at two sites with different climatic regions are used. Their temporal resolution is of one minute. Site *A* is the St Lucia PV site at the University of Queensland (433 kWp), with latitude and longitude of 27.498 and 153.013 respectively and 28 m elevation above the sea level. To have a consistent forecasting duration for all days throughout a year, daytime data from 5:00 am to 7:00 pm is considered.

Site *B* is a 1 kWp PV system at Nanyang Technological University (NTU), Singapore, with latitude and longitude of 1.345 and 103.681 respectively and 41 m elevation above the sea level. NTU is located in a tropical region with mostly cloudy days all year round. The diurnal time for this case is assumed to be 7:00 am to 6:50 pm. Therefore, each day contains 710 data points. For both sites, the data from 2012 is used for training, while that for 2013 is used for genuine forecast evaluation. Fig. 4.1 illustrates PV generation for a typical day at Site *A*, showing its high variability and consequent forecast uncertainty. At such a small temporal resolution and lead times, cloud passages are the main reason for PV power fluctuations, commonly causing 60% drop in the installed capacity [4].

Figure 4.1: PV power variability for a typical day at Site *A* in March 2013

4.3 Non-Parametric Probabilistic Forecasts

Let Y_t be the random variable for PV power generation at time t . f_t and F_t are its PDF and CDF. The quantile $q_t^{\alpha_i}$ with level α_i can be defined as

$$q_t^{\alpha_i} = F_t^{-1}(\alpha_i) \quad (4.1)$$

A quantile forecast $\hat{q}_{t+k|t}^{\alpha_i}$ is consequently the estimate of $q_{t+k}^{\alpha_i}$ issued at time t for lead time $t+k$ based on information available at that time. A single quantile forecast provides only limited information on forecast uncertainties. This motivates the joint consideration of m number of quantile predictions, with increasing nominal levels α_i , $i = 1, \dots, m$, then yielding a nonparametric density forecast $\hat{f}_{t+k|t}$, i.e.,

$$\hat{f}_{t+k|t} = \{\hat{q}_{t+k|t}^{\alpha_i}; 0 \leq \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m \leq 1\} \quad (4.2)$$

with the full continuous density obtained by linear or nonlinear interpolation through these defined quantile forecasts.

Once quantile forecasts and predictive densities are generated, prediction intervals are a natural by-product allowing easing visualization of forecast uncertainties. This is because they represent a range of potential values in which the future observation y_{t+k} may lie within a certain probability. Considering two quantile forecasts whose nominal proportions are $\bar{\alpha}$ and $\underline{\alpha}$, central prediction intervals with nominal coverage $(1 - \gamma)$ are defined by

$$\hat{I}_{t+k|t}^{1-\gamma} = [\hat{q}_{t+k|t}^{\bar{\alpha}}, \hat{q}_{t+k|t}^{\underline{\alpha}}] \quad (4.3)$$

where nominal levels of the quantile forecasts are centered around the median, i.e.

$$\bar{\alpha} - \underline{\alpha} = 1 - \gamma \quad (4.4)$$

The central prediction interval will serve as a basis for the illustration of the proposed nonparametric probabilistic forecasts of solar power generation.

4.4 Performance Evaluation

4.4.1 Reliability

High reliability is one of the major requirements for probabilistic forecasts, as lack of reliability will most likely result in a systematic bias in decision-making. Since $\hat{f}_{t+k|t}$ is composed of a set of quantile forecasts, the evaluation of forecast reliability also focuses on these quantiles. As a basis, let $\xi_{t,k}^{(\cdot)}$ be an indicator variable

$$\xi_{t,k}^{\alpha_i} = \mathbf{1}_{\{y_{t+k|t} \leq \hat{q}_{t+k|t}^{\alpha_i}\}} = \begin{cases} 1, & \text{if } y_{t+k|t} \leq \hat{q}_{t+k|t}^{\alpha_i} \\ 0, & \text{otherwise} \end{cases} \quad (4.5)$$

The empirical level $\hat{\alpha}_i$ of a quantile forecast with nominal proportion α_i is then

$$\hat{\alpha}_i = \frac{1}{N} \sum_{t=1}^N \xi_{t,k}^{\alpha_i} \quad (4.6)$$

Reliability is eventually assessed visually through the use of reliability diagrams, showing empirical versus nominal levels of the defining quantile forecasts, $i = 1, \dots, m$ [107]. Besides, the absolute deviation (denoted *Dev*) from the nominal proportion is a factor to assess overall reliability: quantile forecasts with a lower *Dev* are to be seen as more reliable. The metric *Dev* can be formulated as

$$Dev_i = |\alpha_i - \hat{\alpha}_i| \quad (4.7)$$

4.4.2 Skill score

As for point prediction, forecast users are keen on using a unique score which is able to summarize the performance of rival probabilistic forecasting while accounting for their reliability and sharpness. Reliability evaluation alone does not allow us to do so, since probabilistic forecasts with very bad skill, can easily be made probabilistically reliable.

At time $t + k$, for a single forecast-observation pair, the score

$$Sc(\hat{f}_{t+k|t}, y_{t+k|t}) = - \sum_{i=1}^m \left(\xi_{t,k}^{\alpha_i} - \alpha_i \right) \left(y_{t+k|t} - \hat{q}_{t+k|t}^{\alpha_i} \right) \quad (4.8)$$

is to be seen as a proper skill score [108], hence allowing to objectively sort rival forecasting

approaches. However, both reliability and sharpness should also be examined separately and visually via reliability diagrams and quantiles' plots to make sure that there is a reasonable balance between these two attributes. Skill score values for a given lead time are obtained by averaging the score in (4.8) over all forecast-observation pairs. Score values are positive, with lower score values meaning higher skill.

Probabilistic forecasts are a crucial input for decision-making problems which involve uncertainty. In such problems, traditionally point forecasts have been deployed. More recently in stochastic optimization, several assumptions are made to build a distribution for uncertain variables. For example, in interval optimization, intervals of the random variables are mostly assumed to be a certain percentages lower or upper the point predictions. In Monte Carlo or scenario based methods, the scenarios are extracted from an assumed distribution, usually normal. Therefore, the accuracy of the probabilistic approaches is highly dependent on how much those assumptions match the reality. In this proposal, instead of making assumptions to define the uncertainty of a random variable, a forecast density is predicted for each time point in the future. This can help the decision maker to get a better insight of the uncertainties involved in prediction and make more pragmatic decisions. In [1], as an example to demonstrate the benefits and necessities of the probabilistic forecasts, a bidding strategy in a pool based electricity market is investigated and mathematically it has been proved that the optimal bidding amount is equal to a special quantile produced by wind power probabilistic forecasts.

4.5 From Point Predictions to Nonparametric Probabilistic Forecasts

The ELM is employed as a regression model which can allow to generate both point and quantile forecasts, depending on the training criterion. For point predictions, the goal is to minimize the error between predictions and measurements. This can be translated to a least-square solution for parameters of the regression model. However, in probabilistic forecasting, the targets of the model are quantiles which should cover a certain proportion of the measurements over a defined time. In this case, a pure regression framework which tries to mimic the measurements based on least-square methods cannot be employed. Therefore, new criteria and formulation should be devised to train the model. In the following subsections, after introducing the overall ELM approach, the procedure for designing ELMs to directly generate predictive

densities is explained.

4.5.1 Extreme Learning Machine (ELM)

As explained in Section 3.2, for an ELM, parameters for the hidden layer are randomly selected while they do not need to be tuned through the training process. The output weights are analytically calculated at once without going through a lengthy iterative training [109]. Due to high scalability and very low computational burden, ELM not only unifies different widely used learning methods but also offers a unified approach for different practical applications. The formulation of ELM is described in Section 3.2.

4.5.2 Optimization formulation

The optimization methodology proposed in this thesis aims at adjusting the ELM parameters β (refer to Section 3.2) in order to make them optimal in view of the skill score given in Section 4.4, for the evaluation of predictive densities while satisfying the problem's constraints. The predicted quantiles should be non-crossing and lie in a feasible range of the PV generation. The proposed objective function is

$$\begin{aligned} \min F = & \frac{-1}{N_T m} \sum_{t=1}^{N_T} \sum_{i=1}^m \left(\gamma_1 \times \left((\xi_{t,k}^{\alpha_i} - \alpha_i)(y_{t+k|t} - \hat{q}_{t,k}^{\alpha_i}) \right) \right. \\ & \left. + \gamma_2 \times \xi_{t,k}'^{\alpha_i} + \gamma_3 \times \xi_{t,k}''^{\alpha_i} \right) \end{aligned} \quad (4.9)$$

where N_T is the number of samples over the training period and m is the number of quantiles. γ_1 is a coefficient which represents the importance of the skill score in the objective function. γ_2 and γ_3 are penalty factors for violations from the problem constraints i.e. contravening the feasible range and crossing quantiles respectively. $\xi_{t,k}'^{\alpha_i}$ is an indicator variable which examines if a solution has violated from the feasible range $0 \leq [G]_{N \times m} \leq 1$ and is defined by

$$\xi_{t,k}'^{\alpha_i} = \begin{cases} 1, & \text{if } (\hat{q}_{t,k}^{\alpha_i} > 1) \text{ or } (\hat{q}_{t,k}^{\alpha_i} < 0) \\ 0, & \text{otherwise} \end{cases} \quad (4.10)$$

To ensure that a quantile with a higher ratio encapsulates the one with a lower ratio, $\xi_{t,k}''^{\alpha_i}$ as defined in (4.11) is added to the objective function which checks if a violation has happened.

$$\xi_{t,k}''^{\alpha_i} = \begin{cases} 1, & \text{if } (\alpha_i > \alpha_m) \text{ and } (\hat{q}_{t,k}^{\alpha_i} < \hat{q}_{t,k}^{\alpha_m}) \\ 0, & \text{otherwise} \end{cases} \quad (4.11)$$

In order to keep the values of three terms of the objective function in a roughly similar range, firstly γ_1 , γ_2 and γ_3 are set to 1 and the problem is solved. Secondly, the calculated values for $c_2 = \frac{-1}{N_T m} \sum_{t=1}^{N_T} \sum_{i=1}^m \xi_{t,k}'^{\alpha_i}$ and $c_3 = \frac{-1}{N_T m} \sum_{t=1}^{N_T} \sum_{i=1}^m \xi_{t,k}''^{\alpha_i}$ are compared with those for $c_1 = \frac{-1}{N_T m} \sum_{t=1}^{N_T} \sum_{i=1}^m (\xi_{t,k}^{\alpha_i} - \alpha_i)(y_{t+k|t} - \hat{q}_{t,k}^{\alpha_i})$. Then, γ_1 is kept equal to 1 and γ_j ($j = 2, 3$) is set to $1/a_j$ where $a_j = c_j/c_1$ ($j = 2, 3$).

An optimization algorithm is developed and customized to optimize the proposed objective function. In the modelled optimization algorithm, the optimization procedure is performed through two updating methods. The first one is inspired by the Particle Swarm Optimization (PSO) and is defined by

$$\beta_\chi^1 = \beta_\chi'^1 + W \times \Omega_\chi + C_1 \times (\Omega_\chi - P_\chi^b) + C_2 \times (\Omega_\chi - P^g) \quad (4.12)$$

where χ is the particles index; P^g is the global best solution found by the population and P_χ^b is the best solution found by the particle χ so far. $\beta_\chi'^1$ is the solution found by particle χ in the previous iteration and β_χ^1 is the new solution for particle χ following the updating method in (4.12). Ω is a repository which keeps a portion of the best solutions found by the whole population using two updating methods (4.12) and (4.13). Its size has been considered two times the population size. W , C_1 and C_2 are three random numbers between 0 and 1.

The second updating formula tries to increase diversity among solutions and is defined by

$$\beta_\chi^2 = \Omega_{n_1} + r_1 \times (\Omega_{n_2} - \Omega_{n_3}) + r_2 \times (\Omega_{n_4} - \Omega_{n_5}) \quad (4.13)$$

where β_χ^2 is the updated solution using (4.13) and r_1 and r_2 are two random numbers between 0 and 1. Ω_i ($i = 1, \dots, 5$) are five randomly selected members from the repository.

For this specific optimization problem, if instead of searching within the repository, one searches the unconstrained search space; as the number of iterations increases, the number of infeasible solutions increases. This happens due to the fact that in this problem, the design

variables are the output weight of the ELM ($[\beta]_{\tilde{N} \times m}$) while the constraints are on the output of the ELM ($[G]_{N \times m}$) as the quantiles of PV generations. There is no straightforward method to limit the search space of $[\beta]_{\tilde{N} \times m}$ based on the $[G]$ limitations.

The above optimization algorithm is developed and called Repository-based PSO or RPSO, because several optimization functions in R programming and Matlab failed to provide satisfactory results for the problem formulated in this thesis. “constrOptim”, “optim” and “optimx” and “GenSA” in R and “fmincon”, “fminunc” and “simulannealbnd” in Matlab are tested. The mentioned methods resulted in long simulation time, trapping in local optima or infeasible solutions.

4.5.3 Optimization procedure

The optimization procedure can be described as follows:

Step 1) To initialize the optimal forecast density model, a set of training samples is defined as $\{(\mathbf{x}_{t,j}, \mathbf{g}_{t+k|t,j})\}_{j=1}^{N_T}$ where $\mathbf{g}_{t+k|t,j}$ is composed of a set of m quantiles as

$$\mathbf{g}_{t+k|t,j} = \{\hat{q}_{t+k|t,j}^{\alpha_1}, \dots, \hat{q}_{t+k|t,j}^{\alpha_i}, \dots, \hat{q}_{t+k|t,j}^{\alpha_m}\} \quad (4.14)$$

for $j = 1, \dots, N_T$, and

$$\hat{q}_{t+k|t,j}^{\alpha_i} = (\alpha_i + \text{round}(\alpha_i))y_{t+k} \quad (4.15)$$

where $\text{round}(\alpha_i)$ is a function which rounds α_i to the nearest integer (0 or 1). Other forms of initialization could be seen as relevant, like using the climatology method.

Initial values of β are computed by following (3.6). The subsequent optimization phase then explores the search space to find better values for β , i.e., yielding a set of quantile forecasts with a better skill score value.

Step 2) Initialize the population by randomly generating N_P particles around the initial β .

Step 3) Once the N_P output weights β are positioned using the proposed optimisation approach in subsection 4.5.2, their corresponding estimated quantiles and fitness values can be calculated according to (3.2) and (4.9), respectively.

Step 4) The iterative optimization procedure is carried out until the maximum number of iterations is reached or the convergence criterion is met. At each iteration, for each potential solution $\{\beta_p\}_{p=1}^{N_P}$, the outputs of ELM are calculated by (3.2) and the corresponding objective

function is computed using (4.9).

Step 5) The final β is retained, and then used to generate predictive densities out of the samples.

A flowchart of the proposed framework is given in Fig. 4.2.

Figure 4.2: The flowchart of ELM forecasting method.

4.6 Benchmark Methods

4.6.1 HIA

HIA is formulated in [90] to directly find prediction intervals. Prediction intervals with associated confidence levels are estimated through an optimization procedure when the ELM serves as the regression model. The objective function of the optimization approach takes care of the coverage and sharpness of intervals only, regardless of the quality of quantile forecasts. In HIA, calibration of intervals is evaluated by

$$A_t^{1-\gamma_i} = \hat{P}C_t^{1-\gamma_i} - (1 - \gamma_i) \quad (4.16)$$

where $\hat{P}C_t^{1-\gamma_i}$ denotes the proportion of the observations lying within the interval with nominal coverage $1 - \gamma_i$. The objective function is defined as

$$\min F = \sum_{i=1}^{m'} |A_t^{1-\gamma_i}| + |S_t^{1-\gamma_i}| \quad (4.17)$$

where $S_t^{1-\gamma_i}$ measures the width of the intervals and m' is the number of intervals.

HIA is compared with the climatology, constant, persistence, exponential smoothing method (ESM), and quantile regression (QR) approaches and found to be superior for datasets used in [90].

4.6.2 BELM

A bootstrap-based method is proposed in [91] to predict uncertainties involved in prediction. ELM serves as the regression model and a number of ELMs are trained with resampled training data. Two kinds of uncertainties are modelled, uncertainty related to the noise in data σ_ϵ^2 and uncertainty in the ELM model $\sigma_{\hat{y}}^2$. These two uncertainties are assumed to be Gaussian, independent of each other and dependent on the inputs of the regression model. The total variance of prediction for time t and $\mathbf{x}_j$ denoted by $\sigma_t^2(\mathbf{x}_j)$ is calculated by

$$\sigma_t^2(\mathbf{x}_j) = \sigma_{\hat{y}}^2(\mathbf{x}_j) + \sigma_\epsilon^2(\mathbf{x}_j) \quad (4.18)$$

B_M training datasets are generated by resampling with replacement from the original train-

ing dataset $\{(\mathbf{x}_j, \mathbf{g}_j)\}_{j=1}^N$ and a separate ELM is trained for each of the bootstrapped datasets. With $\hat{y}_l(\mathbf{x}_j)$ as the output of the l^{th} bootstrapped ELM, the mean of the distribution of the predictions is calculated by

$$\hat{y}(\mathbf{x}_j) = \frac{1}{B_M} \sum_{l=1}^{B_M} \hat{y}_l(\mathbf{x}_j) \quad (4.19)$$

In (4.18), $\sigma_{\hat{y}}^2$ is regarded as the variance of the outputs of B_M bootstrapped ELMs $\{\hat{y}_l(\mathbf{x}_j)\}_{l=1}^{B_M}$.

To estimate uncertainty related to noise of data, by replacing the targets $\{\mathbf{g}_j\}_{j=1}^N$ with $\{(\hat{y}(\mathbf{x}_j) - \mathbf{g}_j)\}_{j=1}^N$, a transformed dataset is designed as

$$D_\varepsilon = \{\mathbf{x}_j, (\hat{y}(\mathbf{x}_j) - \mathbf{g}_j)\}_{j=1}^N \quad (4.20)$$

Resampling with replacements is carried out to generate B_N number of resampled datasets from D_ε . The output of the l^{th} bootstrapped ELM is denoted by $\hat{r}_l(\mathbf{x}_j)$. For the transformed training data, again a share of uncertainty is related to the ELM model. Estimated variance of noise $\hat{\sigma}_\varepsilon^2(\mathbf{x}_j)$ and variance attached to the ELM model $\sigma_{\hat{r}}^2(\mathbf{x}_j)$ can be calculated by

$$\hat{\sigma}_\varepsilon^2(\mathbf{x}_j) = \hat{r}(\mathbf{x}_j) = \frac{1}{B_N} \sum_{l=1}^{B_N} \hat{r}_l(\mathbf{x}_j) \quad (4.21)$$

$$\sigma_{\hat{r}}^2(\mathbf{x}_j) = \frac{1}{1 - B_N} \sum_{l=1}^{B_N} (\hat{r}_l(\mathbf{x}_j) - \hat{r}(\mathbf{x}_j))^2 \quad (4.22)$$

The total variance of data noise can be obtained by

$$\sigma_\varepsilon^2(\mathbf{x}_j) = \hat{\sigma}_\varepsilon^2(\mathbf{x}_j) + \sigma_{\hat{r}}^2(\mathbf{x}_j) \quad (4.23)$$

Finally, the lower and upper bounds of interval predictions are obtained by

$$\hat{q}_{t+k|t}^{\bar{\alpha}} = \hat{y}(\mathbf{x}_j) + z_{1-\underline{\alpha}/2} \sqrt{\sigma_t^2(\mathbf{x}_j)} \quad (4.24)$$

$$\hat{q}_{t+k|t}^{\alpha} = \hat{y}(\mathbf{x}_j) - z_{1-\underline{\alpha}/2} \sqrt{\sigma_t^2(\mathbf{x}_j)} \quad (4.25)$$

with $z_{1-\underline{\alpha}/2}$ as the critical value of standard Gaussian distribution dependent on the coverage

level of prediction interval $(1 - \gamma)$.

4.6.3 Probabilistic persistence benchmark

Persistence is a benchmark that is difficult to outperform when considering very short lead times, especially if issuing point forecasts only [106]. A probabilistic generalization is considered here. The persistence forecast density $\hat{f}_{t+k|t}^p$, issued at time t for lead time $t + k$ relies on the sample characteristics of the past n_p observations. The optimal size of this training window is determined using a 10-fold cross validation approach (refer to Appendix B.1) with the aim of minimizing the previously described skill score. $\hat{f}_{t+k|t}^p$ is a Gaussian predictive density, whose quantile $\hat{q}_{t+k|t}^{p, \alpha_i}$ with nominal level α_i is computed as

$$\hat{q}_{t+k|t}^{p, \alpha_i} = \Phi^{-1}(\alpha_i | y_t, \sigma_t) \quad (4.26)$$

where Φ^{-1} is the inverse Gaussian CDF. y_t is the latest measured power value, defining the mean of the Gaussian distribution, and finally with σ_t the sample standard deviation of the n_p past observations in the training window.

4.6.4 Probabilistic climatology benchmark

Climatology is the most widely used benchmark in probabilistic forecasting for weather and weather-related processes. It is not a dynamic forecast because it overlooks recent power values. It is based on all available power measurements, though not making any distributional assumption (like Gaussian) for predictive densities. Climatology is obtained as a function of the time of the day (e.g. every minute in datasets deployed in this study), based on KDE. The idea is to estimate the climatology distribution $\hat{f}_t^c$ for each time t based on the available data for the same time of all days in the same season, as well as neighboring time of day values. Let us define an operator which represents absolute calendar date and is a function of season (s), day (d), and minute of a day (τ) as

$$\varrho(t) = \{s, d, \tau\}, \quad t = \varrho^{-1}(\{s, d, \tau\}) \quad (4.27)$$

To build such a climatology forecast, first of all, a kernel is assigned to each of the available past measurements y_t as

$$K_\sigma(y - y_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{(y - y_t)^2}{2\sigma^2} \quad (4.28)$$

where σ defines the bandwidth of the Gaussian kernel.

Due to the fact that PV power values must be greater or equal to zero, the kernels in Eq. (4.28) should be modified to censored kernels,

$$K_\sigma^+(y - y_t) = \begin{cases} K_\sigma(y - y_t), & y > 0 \\ \Phi(-\frac{y_t}{\sigma}), & y = 0 \end{cases} \quad (4.29)$$

with Φ as the cumulative distribution function of a standard normal distribution.

Then, $\hat{f}_t^c$ is generated as a mixture of the kernels for the same time of day τ (as well as neighbouring ones) for all days of the same season in the dataset. The n_c minutes before and after each τ are considered. Moreover, in order to discount the impact of times further from τ , a weight is assigned to each of the $2n_c + 1$ neighbouring time instances, also using a Gaussian kernel,

$$w_j = K_{\sigma_\tau}(j), \quad j = -n_c, \dots, n_c \quad (4.30)$$

with bandwidth σ_τ . Here, for instance, the case of $j = 1$ denotes the following time instance in the same day, while $j = -1$ refers for the previous time instance. According to Eq. (4.30), as $|j|$ increases, w_j decreases.

Now, $\hat{f}_t^c$ for a given season s and time of day τ can be written as a weighted summation over all the kernels, with appropriate weighting,

$$\hat{f}_t^c(s, \tau) = \frac{1}{N_t \sum_j w_j} \sum_{i=1}^{N_t} \sum_j w_j K_\sigma^+(y - y_{t_{ij}}) \quad (4.31)$$

where N_t is the number of days used to generate $\hat{f}_t^c$, while t_{ij} is a compact notation for $\varrho^{-1}(s, d + i, \tau + j)$.

Once climatology distributions are obtained, the quantiles $\hat{q}_{t+k|t}^{c, \alpha_i}$ for the various nominal levels $\alpha_i (i = 1, \dots, m)$ can be readily computed, as a function of the time of day that the lead time $t + k$ is pointing at, irrespective of the forecast issue time t .

4.7 Empirical Investigation

The empirical investigation in this study aims at evaluating the feasibility of the proposed approach for the solar power application, while evaluating it against a number of relevant benchmarks, namely the persistence [90], climatology [110], HIA [90] and BELM [91].

HIA [90] and BELM [91] are implemented as state-of-the-art methods to verify the comparative performance of the proposed method.

4.7.1 ELM tuning

For the proposed ELM-based predictive densities, three case studies are considered:

Case 1: Focus is given to Site *A* and the lead time is of 10 minutes ahead. The predictors for this case are past solar power observations up to lag n time intervals, clear sky solar irradiance [111], [112], lag values for air temperature, wind speed, humidity, and solar insolation. Therefore, in this case the number of predictors is $n + 5$. Air temperature, wind speed, humidity are measured by Vaisala WXT520 Weather Transmitter and solar irradiance by Kipp & Zonen CMP11. The optimal value for n is determined using 10-fold cross validation and forecasts frequency and resolution is of one minute.

Case 2: Focus is given to Site *B*, while the model relies on past solar power observations up to lag n time intervals, with the same setup as for *Case 1*.

Case 3: Focus is given to Site *A* except that the lead time of one hour is considered.

For all forecasting approaches, the relevant training window size n for the proposed method, n_p for the persistence and n_c for the climatology are between 10 and 120 time steps (minutes), with 10-step increments. As an outcome of the cross validation procedure, n in *Cases 1* and *2* is chosen to be 10 and for *Case 3*, $n = 30$. Besides, n_c is chosen equal to 30, while the best value found for n_p in *Cases 1* and *2* is 60, and finally 120 for *Case 3*. The population size and maximum number of iterations for the optimization phase are set to 70 and 60 respectively and γ_2 and γ_3 are set to one-tenth of γ_1 .

The pattern of PV generation varies significantly daily and seasonally. In order to assess the effectiveness and applicability of the proposed method in an computationally efficient manner and present results rationally, the dataset for each year is divided into quarters or so-called

seasons. For *Case 1* and *Case 3*, the definition of seasons in Southern hemisphere is used and for *Case 2*, the Northern hemisphere seasons are considered. For each season, a separate model is trained by using the archived data of the previous year for the same season. It should be noted that the diurnal hours are as described in Section 4.2 and the night-time hours are excluded in the study. For example, in *Case 1* approximately 3 (months) $\times$ 30 (days) $\times$ 840 (5 am to 7 pm) = $75,600$ time instances are taken into account for each season. All PV generation data is normalized by the nominal capacity of their respective sites.

The simulation results for two forecast horizons, 10-minute (*Cases 1* and *2*) and 1-hour (*Case 3*) lead times are analysed. The 10-minute look-ahead time is one of the most important forecast horizons due to its significant influence on imbalances in power systems. Moreover, the 1-hour lead time forecast has a notable role to play in several operation problems and PV power trading. One-minute forecast is performed to ensure the applicability and effectiveness of the method for applications which require forecasts with high frequency and time resolution. System operators can deploy the proposed methodologies with their most preferred forecast frequency for their specific applications.

4.7.2 Analysis of forecast error distributions

To obtain the empirical PDF of the forecast errors, they are considered conditional on their corresponding predicted values. In principle, like for wind power, the conditional mean (i.e., the point forecasts) is expected to be the most important variable conditioning the prediction intervals or conditional PDFs [82].

The deterministic ELM trained with quadratic criterion is used to generate forecast time series. The data is paired as [forecast, error] where errors $\varepsilon_{t+k|t}$ for a given time and lead time are calculated as

$$\varepsilon_{t+k|t} = y_{t+k} - \hat{y}_{t+k|t} \quad (4.32)$$

with $\hat{y}_{t+k|t}$ the PV power forecast for lead time $t + k$ and y_{t+k} the corresponding observation.

The pairs are then sorted by their predicted values and divided into 20 bins with 0.05 pu width each. Figs. 4.3 and 4.4 show kernel density estimates of forecast errors in all 20 bins for the predicted power values in *Case 1* and *Case 2* respectively. Each bin defines a range for point predictions of solar power values, for which all forecast errors are collected to illustrate their conditional empirical distributions. Kernel density estimate approach is explained in Appendix

Figure 4.3: Kernel smoothing density estimate of distribution of errors in forecasted values (*Case 1*).

A. From the figures, it is clear that PV forecast errors cannot be assumed to be *Gaussian*, *Beta* or following common parametric distributions. The same experiment was conducted for forecast errors for the persistence benchmark, yielding similar conclusions. Without using additional explanatory variables, e.g., from sky imaging, it is unlikely that these distributions could look much different.

Fig. 4.5 illustrates the overall kernel density of bins altogether versus fitted *Normal* distribution to the same data in *Case 1* and *Case 2* respectively. In the same figure, empirical CDF as well as fitted *Normal* CDF to the all forecasted errors are shown. From the figure, it is obvious that the overall distribution of the errors is not *Normal*. To support this argument, One-sample Kolmogorov-Smirnov (KS) test is conducted. Null hypothesis is defined as “the forecasted errors come from a *Normal* distribution” where mean and standard deviation of that distribution are calculated by fitting a normal distribution to the data. For both *Case 1* and *Case 2*, KS test rejects the null hypothesis at the 5% significance level. The test was repeated for *t-location-scale* and *Logistic* distributions and again KS test rejects the null hypothesis. *Rayleigh*, *Beta*, *Gamma* and *Weibull* distributions are ignored as they support only positive values. KS test also is conducted for each of the 40 bins illustrated in Figs. 4.3 and 4.4 with

Figure 4.4: Kernel smoothing density estimate of distribution of errors in forecasted values (*Case 2*).

null hypotheses explained above. In all cases, the hypothesis is rejected.

The solar power generation process is nonlinear and bounded between zero and nominal capacity. Furthermore, its upper bound is actually given by the clear sky model for each and every time (for instance at night, the upper bound is zero). Similar to the case of wind power, this very nature of the process leads to the observed characteristics of the PDFs. When predicting power close to the upper bound, the potential realization can only get lower (hence a long lower tail), and when predicting close to zero, the power output can only get higher (hence a long upper tail). Besides, the bimodality is due to cloud passages, that can be regularly missed by the forecasting approach. When predicting, the method could inform that there are two “most probable states” for future power generation: one if no cloud is coming, and another one if cloud is coming. The presence of this bimodality depends on the local climate, and how much cloud passages may influence short-term variations in power production.

One could possibly envisage proposing mixtures of parametric distributions in the future, based on a hidden regime sequences. Here, emphasis is placed on obtaining highly-skilled nonparametric predictive distributions that accommodate these peculiar characteristics.

Figure 4.5: Top: the overall kernel density of bins altogether versus fitted *Normal* distribution to the same data for *Case 1* and *Case 2*, respectively. Bottom: Empirical CDF versus fitted *Normal* CDF to the same data

4.7.3 Nonparametric quantile forecast results

To describe predictive densities, 18 quantiles with nominal levels from 0.05 to 0.95, at increments of 0.05, except for the median (0.5), are issued. These quantiles form a set of 9 central prediction intervals, with nominal coverage rates from 10 to 90%. All quantiles in this study are censored such that PV generation is bounded between zero and the nominal capacity of PV at the respective sites.

It is to be noted that the most important attributes of good probabilistic forecasts are reliability and sharpness. The reliability diagram gives information on how well the predicted probabilities of an event correspond to their observed frequencies. However, the probabilistic forecasts models cannot be judged only on their reliability. Considering both sharpness and reliability are required to benchmark methods. Those can be offered by the discussed skill score in subsection 4.4.2 which examines both the mentioned attributes. Both reliability and sharpness should also be examined separately and visually via reliability and sharpness diagrams

as well as the quantiles' plots. This helps to check if there is a reasonable balance between these two attributes. Some methods tend to provide too sharp quantiles that may result in good skill scores but if the reliability is too low, the efficiency is questionable.

Figs. 4.6, 4.7 and 4.8 depict example episodes with probabilistic forecasts generated for *Cases* 1, 2 and 3, respectively. They cover five successive days in *Cases* 1, 2 and two days (the same last two days shown in Fig. 4.6) in *Cases* 3 to provide a view of the high diurnal diversity in PV generation patterns. These plots illustrate how the proposed forecasting approach reacts to the high intermittency and variability in PV generation for different days. At first sight, the coverage of the prediction intervals appears reasonable. Besides, these intervals are not symmetric around the point forecasts. This is to be expected since point predictions correspond to the mean value of the distributions whereas prediction intervals are centered on the median.

Figure 4.6: Predictive densities as a set of prediction intervals for five successive days in January 2013 (*Case* 1)

Figure 4.7: Predictive densities as a set of prediction intervals for five successive days in August 2013 (*Case* 2)

Figure 4.8: Predictive densities as a set of prediction intervals for two successive days in January 2013 (*Case 3*)

Table 4.1: Scores, maximum and sum of probability deviation of quantiles, *Case 1*

<i>Case 1</i>	Proposed Method			Climatology			Persistence			HIA			BELM		
	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum
Spring	2.92	4.52	28.71	4.63	10.47	75.87	2.96	6.77	57.07	4.51	10.23	82.41	2.34	19.88	127.55
Summer	3.30	5.99	35.81	5	4.82	36.8	3.63	8.8	53.71	4.21	8.43	61.21	3.31	6.21	61.51
Fall	3.12	3.74	26.94	3.79	11.08	93.34	3.67	9.92	79.48	4.35	9.54	67.54	2.78	13.77	111.46
Winter	2.10	3.44	30.98	2.67	9.26	95.01	2.23	8.38	49.81	2.91	11.21	75.43	1.53	19.81	182.82

Scores, maximum absolute deviation from nominal proportion, and the sum of the absolute deviations for all 18 quantiles are given in Tables 4.1, 4.2 and 4.3 for *Cases 1, 2 and 3* respectively. The values are given in percent form (%). From the tables, the persistence and the BELM quantiles show acceptable performance in terms of skill score but they have very low reliability. The maximum deviations from perfect reliability using the BELM can be as high as 20%. This undermines its efficiency and reliability in practical applications. On the other hand, climatology provides the worst overall performance. This is of course reasonable since climatology is an unconditional forecast and ignores the heteroscedasticity of PV generation series.

The HIA method also presents poor performance in terms of both reliability and sharpness. The poorly calibrated quantiles given by the HIA are the consequence of the evaluation criteria defined for this method in which only the calibration of intervals is taken into account and the quality of quantiles is neglected.

The BELM method tends to provide very narrow and sharp quantiles which result in good skill scores. However, there should be a reasonable balance between sharpness and calibration. By looking at large deviations from the nominal proportions of the BELM quantiles, it can be concluded that the BELM gives more weight to sharpness which leads to unreliable quantiles.

Table 4.2: Scores, maximum and sum of probability deviation of quantiles, *Case 2*

<i>Case 2</i>	Proposed Method			Climatology			Persistence			HIA			BELM		
	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum
Spring	2.39	3.71	14.22	4.7	4.62	36.49	2.66	10.05	85.79	2.97	9.18	82.67	2.98	13.81	152.23
Summer	2.41	3.80	22.10	4.6	6.60	50.93	2.6	10.23	66.98	3.45	7.52	63.45	2.62	8.93	79.93
Fall	2.78	8.00	56.93	4.9	4.77	57.08	3.13	10.17	66.13	3.21	6.33	69.52	2.53	9.71	88.44
Winter	2.51	5.36	41.03	4.68	13.76	167.6	3.23	9.87	82.67	3.15	8.42	78.43	2.48	13.03	113.68

Table 4.3: Scores, maximum and sum of probability deviation of quantiles, *Case 3*

<i>Case 3</i>	Proposed Method			Climatology			Persistence			HIA			BELM		
	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum	Score	Max	Sum
Spring	3.52	5.32	48.51	4.63	10.47	75.87	5.85	6.04	48.53	4.34	8.24	70.43	3.23	13.83	157.48
Summer	4.24	5.28	41.25	5.00	4.82	36.8	5.77	6.21	36.58	4.92	11.86	84.91	4.40	13.81	127.07
Fall	3.48	4.41	37.21	3.79	11.08	93.34	5.91	12.21	88.28	4.39	8.32	69.47	3.62	11.13	92.61
Winter	2.23	4.32	23.5	2.67	9.26	95.01	4.58	14.72	92.95	3.42	7.53	61.42	2.15	11.33	98.71

Overly narrow quantiles work in case of clear sky when no significant uncertainty is present and PV generation is predictable with good accuracy. However, they fail to cover most of the ramp-ups and ramp-downs in cloudy days. This can be checked in Fig. 4.9 where BELM quantiles for the first day depicted in Fig. 4.8 are illustrated. The rational behind too narrow prediction intervals given by BELM is related to its formulation where the distribution of the uncertainty is assumed to be Gaussian. However, according to investigation discussed in subsection 4.7.2, for the case of PV generations, forecast errors are not distributed Gaussian. The variances among the output of different sets of bootstrapped ELMs are small and cannot represent the level of uncertainty in the measurements. Looking at Fig. 4.9, the centres of the prediction intervals are very close to the point predictions. As the prediction intervals according to (4.25) are centered on the average of the outputs of the bootstrapped ELMs, it can be concluded that the averages of forecasts given by bootstrapped ELMs are very close to the point forecasts. It is to be noted that point predictions are generated using ELM [113] when the original dataset is treated as the input.

To get a better understanding of the sharpness of quantile forecasts generated by the various methods discussed, Fig. 4.10 is provided. Here, sharpness is examined by assessing the average width of central prediction intervals regardless of their similarity to the measurements. For each interval prediction, the distance between its two boundaries are calculated for each time and then the obtained values are averaged over the evaluation dataset. In Fig. 4.10, the nominal coverage rates of the intervals versus their average widths are depicted for four seasons of 2013, *Case 2*. According to this figure, BELM prediction intervals have the lowest width, followed by the persistence intervals. Although for the intervals with the lower coverage, the widths of

Figure 4.9: BELM predictive densities as a set of prediction intervals for one day in March 2013 (*Case 3*)

Figure 4.10: Sharpness diagrams for 9 central prediction intervals with nominal coverage rates increasing from 10% to 90% for *Case 1*, (a) Spring, (b) Summer, (c) Fall, (d) winter

the intervals given by the proposed method are close to those given by BELM, for the coverage rates of 70% and more, they are wider than those of BELM intervals. The reason for that is the inclusion of the calibration criterion in the proposed framework which leads to wider intervals to meet the calibration requirements.

It should be noted that in all cases, the temporal resolution of forecasting is of one minute and the instantaneous measurements, not the average values, are taken into account. Needless to say, temporal averaging has smoothing effect on variability and if pursued it is expected to result in sharper quantiles.

Reliability diagrams in Fig. 4.11, for four seasons of 2013 in *Case 1*, allow for a more comprehensive evaluation of the reliability of predictive densities. The red solid line represents

Figure 4.11: Reliability diagrams for 18 quantiles with 5% increasing nominal proportion, *Case 1* (a) Spring, (b) Summer, (c) Fall, (d) Winter

perfect reliability and reliability curves for the predictive densities should be as close as possible to that of the ideal case. Here, the proposed method yields satisfactory reliability, with the observed coverage close to the nominal proportions for most nominal levels.

The BELM tends to underestimate quantiles with nominal levels below 50% while overestimating those greater the median. Among all methods mentioned in Fig. 4.11, the BELM shows the worst reliability and this can be a consequence of the normality assumption in this method as well as high intermittent datasets used in this study. This shows that the BELM is incapable of covering sharp ramp rates in PV generations.

As can be seen in Fig. 4.11, the reliability performance for the summer season is better than that of the other seasons. This also can be noticed by comparing the sum of the deviations from perfect reliability in Table 4.1. Table 4.4 provides point forecasts accuracy by season for *Case 1* in terms of *RMSE* and *MAE*. Because the same predictors are used for all seasons, the lower *RMSE* can be interpreted as more predictability for the corresponding season. Looking at the table, the summer season has the lowest predictability among others which of course is expected. Thunderstorms are a normal part of Brisbane weather condition during December to March (summer season) and can be accompanied by hail and strong wind gusts, making the weather very unstable. PV generation experiences very intermittent and sharp ramps in the summer season. This will force the climatology, BELM and persistence methods to provide wider prediction intervals.

The reason is that all the mentioned methods work based on observed prediction errors or PV power variations. When the ramps are sharper, these methods tend to widen the intervals to envelop the variations. The low sharpness of the quantiles given by the mentioned methods can also be concluded by comparing the scores in Table 4.1 for different seasons. For the summer

Table 4.4: Point forecasts accuracy by season, *Case 1*

Season\accuracy	<i>RMSE</i>	<i>MAE</i>
Spring	13.26%	8.08%
Summer	16.43%	10.73%
Fall	13.85%	8.48%
Winter	7.67%	4.16%

season, the worst score is obtained. As skill scores represent both sharpness and reliability, a better reliability with a worse skill score is an evidence of lower sharpness or wider quantiles. For the winter, when PV generation is more predictable, the best sharpness and the lowest reliability are obtained.

The proposed approach makes a compromise between sharpness and reliability, permitting us to obtain the best or close to best score values while maintaining reliability in an acceptable level. Analysing Tables 4.1-4.3 for *Case 1*, improvements with respect to the persistence benchmark are within 4-14%. Compared to the climatology, improvements reach 53%. Similarly from Table 4.2, improvements with respect to the persistence range between 7 and 22% depending on the season of the year. The best improvements obtained by comparing against the climatology are up to 40%. Comparing the proposed method with the BELM, it is outperformed by the BELM for a couple of seasons, however, the deviations of the BELM quantiles are too high which undermine their reliabilities. Poorly reliable quantiles can cause systematic biases in decision-making process.

Complementing the previous discussion on reliability assessment, the results in the various tables confirm the substantially better calibration of the proposed approach. As an example, in Spring and for *Case 2*, the sum of deviations from the nominal levels for the proposed approach is one-tenth of that for the BELM benchmark. The ratio is one-fifth in Fall for *Case 1* and one-third in Summer for *Case 3*.

As some of the operational problems such as state estimation requires quantiles with high coverages [114], additional tests were conducted to evaluate the capability of both persistence benchmark and the proposed approach to generate quantiles with 99.5%, 99% and 97.5% proportions for *Case 2*. Similarly to less extreme quantiles, it was noticed that the proposed approach yielded score values improvements of 67%, 53% and 47%, for nominal levels of 99.5%, 99% and 97.5% respectively.

In order to investigate the performance of the methodology for times close to sunset, for *Case 2*, point prediction is performed for data values between 6:05 pm to 7:05 pm using ELM methods. All data for 2012 is used as training data and all data for 2013 as a test set. The yearly *RMSE* and *MAE* are found to be 2.6% and 1.6% respectively. Then, quantiles are forecasted for the same data using the proposed framework. Skill score, maximum deviation of the quantiles' coverage from perfect reliability and sum of the deviations for all 18 quantiles are calculated as 0.9%, 0.42% and 3.7% respectively and are much better than those calculated for the hours closer to midday (Table 4.2).

It should be noted that the proposed probabilistic forecast is based on an offline training algorithm and it uses the historical data to train the model and only the values for predictors are updated in real-time. In this thesis, the forecasting model is updated seasonally as it results the best performance for the three case studies, however, the applicability of the idea is not limited to the updating period used here. Depending on statistical behaviour of the PV generation and methodological variables in a specific region, the updating periods as well as the selection of the data used as predictors may change. Once the model is trained, it is ready to be used for online or real-time prediction. To predict m number of quantiles for each time instance in the future, only one simple matrix multiplication given in (3.2) is required. The multiplication can be executed in a small fraction of a second which makes the proposed method suitable for online and real-time applications.

4.7.4 Normalization with respect to clear sky PV power

In days with clear sky all day long, PV generations have a more predictable pattern. Therefore, one may expect that normalization of PV power with respect to the estimated clear sky power potentially can improve predictability PV generation.

In [15], 85% quantile forecast of PV power is considered as the estimated clear sky power. Measurements are then normalized with respect to the clear sky power. The approach has improved the forecast accuracy of the point forecasts in terms of *RMSE*. In order to check the performance of the normalization with respect to the clear sky power, here, three definitions of clear sky power are provided. Those are 85% and 95% quantile forecasts and physical estimation of the clear sky power. For the later one, firstly using geographic coordinate data and elevation from the sea level, the clear sky solar irradiance is calculated according to the model given in Appendix C [112]. Then, the clear sky solar irradiance is converted to the PV power according

Figure 4.12: PV power measurements (red colour curve), physical clear sky power (green colour curve), normalized PV power with respect to the clear sky values (black colour curve) and normalized values with respect to 85% quantile (cyan colour curve) for 6 randomly selected days from the dataset for *Case 2*. The diagram covers values for 5 am to 7 pm. Counting minutes of day is started from 5 am. Power values more than 1.5 pu are not shown.

to the formulation given in [115].

Physical estimation of the non-overcast power and 85% quantile are found closer to the measurements in days with clear sky all day. Therefore, investigations are carried out while

considering these two as the clear sky power. In Fig. 4.12, normalized PV power values with respect to the 85% quantile as well as physical clear sky power for 6 randomly selected days from the dataset for *Case 2* are shown. As can be seen in the figure, some of the measurements in partially cloudy days exceed both the clear sky power and 85% quantile. The same problem is reported in [15]. Reflections from the clouds and varying level of water vapor in atmosphere are the main reasons for this phenomenon. Looking at the figure, the normalization can help to improve stationarity in days with clear sky and hours closer to noon time.

To verify the extent to which normalization of PV power measurements can improve the performance of probabilistic forecasts, quantile forecasts for normalized time series are generated by deploying the proposed method for *Case 2*. Based on the investigations carried out, the normalization approach results in no significant improvement in probabilistic forecasting according to the specific framework proposed in this thesis. The following points can explain the results of the empirical investigations.

- In [15], the motivation behind the proposal of normalization of PV power with respect to the clear sky solar power is that in [15], classical time series methods are used for prediction. Classical time series methods require stationary time series as their inputs. Normalization can potentially create more stationary time series. However, there is no stationarity restriction for neural network based regression methods, such as ELM.
- Because normalization with respect to the low clear sky values for times close to sunrise and sunset leads to very high or infinity errors, according to [15], the daytimes with clear sky power lower than 20% of the maximum clear sky power for the same day should be excluded from the dataset. This results in exclusion of 15% to 30% of the daytime period considered in this study (5 am to 7 pm). Therefore, normalization is not a pragmatic approach to improve forecasts for times close to sunrise and sunset.
- In [15], 15-min and 1-hour averaged values are used. In this thesis, the resolution of data is one minute and both datasets are from locations where most of days of year are cloudy or partially cloudy. With 1-minute instantaneous values, cloud passages are the main reason of intermittency. As the movement of clouds is random, for partially cloudy days, the clear sky power values can no longer be representative of real measurements.

4.8 Conclusions

The focus in PV forecasting literature has been on point prediction only, while point prediction provides only conditional expectation of PV power output. In this chapter, a systematic framework is proposed for generating probabilistic forecasts for PV power generation. The empirical investigations show that PV power forecast errors do not follow common distributions like *Gaussian*, *Beta*, etc. Therefore, to avoid restrictive assumptions on the shape of the forecast densities, a nonparametric density forecasting method is proposed which works based on ELM as a regression tool, trained with an appropriate criterion.

The proposed nonparametric method is successfully applied on two PV power datasets with one-minute resolution and highly fluctuating patterns. The results show that the proposed method is able to efficiently provide reliable and sharp predictive densities for the very short-term (10-minute and one-hour lead times). Part of the future works should focus on demonstrating the functionality, practicability, and value of such probabilistic solar power forecasts in operational problems. Besides, The approach could be extended to account for Numerical Weather Prediction (NWP) as input when looking at further lead times, say, more than two hours ahead.

One of the contributions of this chapter is the investigations performed on the empirical distributions of the PV forecast errors. Proposing a new method for nonparametric density forecasting is another contribution. The proposed method is consisted of an optimization phase and a regression phase. An objective function is proposed for optimization-regression framework which accounts for feasibility, reliability and, sharpness of predictive densities. Another contribution of this work is to provide comparative discussion on the performance of probabilistic persistence and probabilistic climatology methods in generating predictive densities for the specific case of very short-term PV power.

Chapter 5

Space-Time Trajectories of Photovoltaic Power

5.1 Introduction

In the probabilistic energy forecasting literature, emphasis is mainly placed on deriving marginal predictive densities for which each random variable is dealt with individually, similar to the proposal and analyses provided in the previous chapter. Such marginals description is sufficient for power systems related operational problems if and only if optimal decisions are to be made for each lead time and each location independently of each other. However, many of these operational processes are temporally and spatially coupled, while uncertainty in photovoltaic (PV) generation is strongly dependent in time and in space. This issue is addressed in this chapter by analysing and capturing spatio-temporal dependencies in PV generation. Multivariate predictive distributions are modelled and space-time trajectories describing the potential evolution of forecast errors through successive lead times and locations are generated.

Discrimination ability of the relevant scoring rules on performance assessment of space-time trajectories of PV generation is also studied. Finally, the advantage of taking into account space-time correlations over probabilistic and point forecasts is investigated. The empirical investigation is based on the solar PV dataset of the Global Energy Forecasting Competition (GEFCom) 2014.

5.2 Experimental Data Description

As a basis for PV generation space-time dependency investigation, time series of PV generation for three contiguous zones are used. The installation specifications of these zones are given in Table 5.1. Technical specifications of the PV panels are given in [116–118].

Table 5.1: Description of the three PV zones

	Latitude	Longitude	Nominal Power	Altitude	Panel Type	Panel Orientation (from the north)	Panel Tilt
Zone 1	35°16'30" S	149°06'49" E	1,560 W	595 m	Solarfun SF160-24-1M195	38° Clockwise	36°
Zone 2	35°23'32" S	149°04'01" E	4,940 W	602 m	Suntech STP190S-24/Ad+	327° Clockwise	35°
Zone 3	35°32' S	149°09' E	4,000 W	951 m	Suntech STP200-18/ud	31° Clockwise	21°

To predict points and marginal densities of PV generations for three PV zones described above, 12 independent variables as the output of Numerical Weather Prediction (NWP) provided by European Centre for Medium-Range Weather Forecasts (ECMWF) are used as the explanatory variables. The period for which both NWP and PV measurements are available is from April 2012 until the end of June 2014. Therefore, in total around 800 days worth of data per zone with hourly resolution are used.

The NWP variables employed in this study are total column ice water, surface pressure, relative humidity at 1000 mbar, total cloud cover, 2 metre temperature, surface solar radiation down, surface thermal radiation down, net solar radiation at the top of the atmosphere, total precipitation, 10-metre U wind component and 10-metre V wind component.

Half of the available data is used as the training dataset to generate predictive densities and the second half is used as the evaluation dataset. The dataset has been prepared for the Global Energy Forecasting Competition (GEFCom) 2014 and is available online [105].

ECMWF generates NWPs not forecasts for PV generation which is the variable of interest in this study. However, as the forecast horizon in this practice is 1 to 24 hours ahead with hourly resolution, NWPs can be viewed as the most reliable explanatory variables for PV power forecasting [41]. It is to be noted that geostationary satellite imagery is not the best option for forecast horizons more than 5 hours [41].

5.3 Deriving and Evaluating Predictive Marginal Distribution

Given PV generation as a multivariate random variable denoted by $\mathbf{P}_t$ with univariate components represented by $P_{z,t+k}$ for lead time k and location z , and $F_{z,t+k}$ as univariate strictly increasing cumulative density function (CDF) and $f_{z,t+k}$ the corresponding density function, α -quantile is the value for which the probability of occurring $P_{z,t+k}$ below that is equal to $\alpha_i \in [0, 1]$. This function can be mathematically written by

$$\Pr(P_{z,t+k} < q_{z,t+k}^{\alpha_i}) = \alpha_i \quad \text{or} \quad q_{z,t+k}^{\alpha_i} = F_{z,t+k}^{-1}(\alpha_i) \quad (5.1)$$

Using explanatory variables up to time t , a quantile forecast with nominal proportion α_i for time $t+k$ and location z , denoted by $\hat{q}_{z,t+k|t}^{\alpha_i}$ is obtained. As a single quantile contains only a limited information about the likely behaviour of the respective random variable in the future, non-parametric probabilistic forecasts are informed as a set of m number of quantiles with increasing nominal levels to form a density forecast $\hat{f}_{z,t+k|t}$ as

$$\hat{f}_{z,t+k|t} = \{\hat{q}_{z,t+k|t}^{\alpha_i}; 0 \leq \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m \leq 1\} \quad (5.2)$$

$\hat{F}_{z,t+k|t}$ is the corresponding marginal predictive CDF and it is derived by looking at each look-ahead time k and each location z individually and independently.

Non-parametric probabilistic forecasts are referred to a group of forecasting methods with no restrictive assumption on the shape and features of the predictive distributions. Quantile regression is the most widely used type of probabilistic forecasting methods introduced in [119].

Continuous Rank Probability Score (CRPS) is used to compare the skill of predictive marginals for each component of $\hat{\mathbf{P}}_t$, i.e. each lead time in each contiguous location [108]. CRPS is a negatively-oriented proper score with lower values representing a higher skill of the respective marginal distribution. CRPS compares $\hat{F}_{z,t+k|t}$ with CDF of corresponding observation $F_{z,t+k}^0$. Because for each t , k and z , observation is a single value, the observed CDF is modeled as a single step function when the step from zero to one occurs at the observed value. CRPS for each k and z is given by

$$\text{CRPS}_{z,k} = \frac{1}{T} \sum_{t=1}^T \left(\int_0^1 \left(\hat{F}_{z,t+k|t}(p) - F_{z,t+k}^0(p) \right)^2 dp \right) \quad (5.3)$$

5.4 Space-time Dependency Modelling

In order to model aggregated uncertainties of a set of concurrent random variables in contiguous or geographically diverse locations, a joint or multivariate distribution is inevitable. Multivariate distributions are usually characterized by a set of trajectories or scenarios drawn from them where the generated trajectories can be treated in the same way as ensemble forecasts are treated in the realm of methodology. In the following, the derivation of the joint distribution and trajectory generation are outlined.

5.4.1 Multivariate distributions

In space-time dependence modelling, PV generation for lead time k and location z is denoted by

$$P_{t,(z-1)K+k} \quad k = 1, 2, \dots, K, z = 1, 2, \dots, Z \quad (5.4)$$

Therefore, the underlying stochastic process is of dimension $D = K \times Z$ including random variables $P_{t,1}, P_{t,2}, \dots, P_{t,D}$. It should be noted that in this work, upper case letters symbolize random variables while their realizations are expressed by lower case letters. As an example, $[p_{t,1}, p_{t,2}, \dots, p_{t,D}]^\top$ denotes realizations of P_t written as a vector and the corresponding D -variate CDF, $\mathbf{F}_t$ is described by

$$\mathbf{F}_t(p_{t,1}, p_{t,2}, \dots, p_{t,D}) = \Pr(P_{t,1} \leq p_{t,1}, P_{t,2} \leq p_{t,2}, \dots, P_{t,D} \leq p_{t,D}) \quad (5.5)$$

$\hat{\mathbf{F}}_t$ as a predictive multivariate distribution conditional on information up to time t is an estimation of $\mathbf{F}_t$ generated using predictive marginal distribution for each random variable $F_{t,d}$ ($d = 1, \dots, D$).

Based on Sklar's theorem [120], to model the interdependence structure of a set of random variables, respective marginal distributions can be linked together using a copula function as

$$\hat{\mathbf{F}}_t(p_{t,1}, p_{t,2}, \dots, p_{t,D}) = C(\hat{F}(p_{t,1}), \hat{F}(p_{t,2}), \dots, \hat{F}(p_{t,D})) \quad (5.6)$$

A copula is defined as a multivariate distribution function for which the marginal distribution of each random variable is uniform on $[0, 1]$. Therefore, in order to use copulas, random variables $P_{t,d}$ ($d = 1, \dots, D$) should be transformed to a uniform domain, also called rank domain. Then,

dependency modelling is carried out in this common domain.

By definition of a CDF for a continuous variable, $\hat{F}_d(P_d)$ is distributed uniformly as $\hat{F}_d(P_d) \sim U[0, 1]$. Therefore, to transform random variables to a uniform domain, a simple CDF transformation can be used. The transformed variables preserve the dependency structure which is presented in the original domain [121]. It is noteworthy that continuous CDFs can be obtained by fitting a smooth curve through the set of quantiles available. As CDF is a strictly increasing function, the variables in the uniform domain can be transferred back to their original domain later using the inverse CDF transformation. This transformation can be mathematically written as

$$P_d = \hat{F}_d^{-1}(U) \Leftrightarrow U = \hat{F}_d(P_d) \quad (5.7)$$

Accordingly, a copula function can be represented by

$$\begin{aligned} C(U_1, U_2, \dots, U_D) &= \Pr(\hat{F}_1(P_1) \leq U_1, \hat{F}_2(P_2) \leq U_2, \dots, \hat{F}_D(P_D) \leq U_D) \\ &= \mathbf{F}(\hat{F}_1^{-1}(U_1), \hat{F}_2^{-1}(U_2), \dots, \hat{F}_D^{-1}(U_D)) \end{aligned} \quad (5.8)$$

The formulation for Gaussian, T and Gumbel copulas can be found in [121, 122]. Unfortunately, there is almost no guidance available on the proper copula family which can describe correlated variations in PV power. For the wind case though, some practices have investigated the performance of several families of copulas like Archimedean and Elliptical copulas [38]. To model wind power for systems with only few variates, the Gumbel copula is recommended as the right one to be used. However, as the dimension of the system increases, more probably the Gaussian copula can outperform the Gumbel one [123, 124]. Here, the focus is given to the Gaussian copula because based on the empirical tests discussed in section 5.6, it is able to provide adequate superiority over the benchmarks introduced. However, studying the right copula is essential piece of research that should be followed by future works in this area.

In the following subsections, with the focus on Normal Copula, derivation of the multivariate distribution and space-time trajectory generation are explained.

5.4.2 Multivariate Normal distribution with adaptive correlation matrix

Given $\hat{F}_{t,d}$, the random variable Y_d with $y_{t,d}$ as its realization at time t can be defined by

$$y_{t,d} = \hat{F}_{t,d}(p_{t,d}) \quad \forall t \quad (5.9)$$

where Y_d is uniformly distributed on interval $[0, 1]$. Given $Y_d \sim U[0, 1]$, a normally distributed X_d can be attained using the probit function Φ^{-1} as

$$x_{t,d} = \Phi^{-1}(y_{t,d}) \quad (5.10)$$

where the probit function is

$$x_{t,d} = \sqrt{2} \operatorname{erf}^{-1}(2y_{t,d} - 1) \quad (5.11)$$

If the forecasted CDFs are calibrated, the random variable $\{x_{t,d}\}_{t=1}^T$ for each d follows the standard Gaussian distribution, i.e. $X_d \sim \mathcal{N}(0, 1)$.

Denoting $\mathbf{X} = (X_1, X_2, \dots, X_D)^\top$, $\mathbf{X}$ follows a multivariate Gaussian distribution, $\mathbf{X} \sim \mathcal{MVN}(\mu_0, \Sigma)$, where μ_0 is a vector of zeros and size D and Σ_t is a variance-covariance matrix of size $D \times D$.

If historical data for a long period of time is available, the empirical correlation matrix can be calculated. Alternatively, an adaptive approach can be used to estimate Σ for each time t as

$$\Sigma_t = \lambda \Sigma_{t-1} + (1 - \lambda) \mathbf{X}_t \mathbf{X}_t^\top \quad \forall t > 1 \quad (5.12)$$

where λ is a forgetting factor ($\lambda \in [0, 1]$). Σ_t is initialized (for $t = 1$) by setting the diagonal elements equal to unity and all the other elements equal to zero.

With the assumption of perfect reliability of predictive quantiles, the diagonal elements of Σ_t should remain equal to 1 through updating steps. However, due to deviations from perfectly calibrated quantile forecasts, they may deviate from 1. If it is assumed that such deviations are caused by variance scaling, the following transformation can be used to make the covariance matrix a proper representative for a standard Gaussian variable.

$$\Sigma_t = \Sigma_t \oslash (\sigma_t \sigma_t^\top) \quad (5.13)$$

where σ_t is a standard deviation vector of size D containing the square root of the diagonal elements of Σ_t and $\oslash$ is an operator representing the element-by-element division.

The correlation matrix should be positive semi-definite. However, due to noisy estimation of the correlation matrix in either of the approaches (recursive or fully empirical), Σ often appears to be nonpositive semi-definite. Therefore, it is necessary to adjust the violation by approximating Σ to the most closely positive semi-definite matrix.

5.4.3 Trajectory generation

To generate several scenarios of the possible values that the random variable $P_{t,d}$ may take, the most up-to-date covariance matrix which can be deployed is Σ_{t-1} . It is to be noted that in this practice, t is the day index. Given the forecasted $\hat{f}_{t,d}$ and $\hat{F}_{t,d}$ for all involved random variables, and the multivariate distribution of $\mathbf{X}$, $\mathbf{X} \sim \mathcal{MVN}(\mu_0, \Sigma_{t-1})$, S scenarios are generated as follows:

- Draw S realizations of $\mathcal{MVN}(\mu_0, \Sigma_{t-1})$. Each of them is a vector of size D , denoting $\mathbf{x}_{s,(t-1)}$, $s = 1, 2, \dots, S$
- The elements of the vector of uniform variable $\mathbf{Y}_s = \{y_{s,d}\}_{d=1}^D$ for scenario s is achieved by employing the inverse of the probit function as

$$y_{s,d} = \Phi(x_{s,d}) \quad \forall s, d \quad (5.14)$$

- $[\hat{p}_{s,t,1}, \hat{p}_{s,t,2}, \dots, \hat{p}_{s,t,D}]$ represents the s^{th} space-time trajectory of PV generation for time t . $p_{s,t,d}$ can be obtained by

$$\hat{p}_{s,t,d} = \hat{F}_{t,d}^{-1}(y_{s,d}) \quad \forall s, d \quad (5.15)$$

5.5 Multivariate Scoring Rules

The benefits offered by probabilistic forecasting to decision making process are directly dependent on the quality of forecasts. While a range of univariate scoring rules can be found [93, 94], only a few quantitative assessment criteria for multivariate setting have been proposed.

Some of the most relevant scoring rules for evaluation of the multivariate predictive trajectories for the case of spatially-temporally correlated PV generations are used here.

5.5.1 Energy Score

The most commonly used scoring rule when distributions are represented by a finite number of trajectories is known as energy score (ES). ES as a multivariate generalization of the CRPS has been formulated and introduced in [108]. This score is proper and negatively oriented, i.e. a lower score represents a better forecast. Energy score is calculated as

$$ES_t = \frac{1}{S} \sum_{s=1}^S \left\| \mathbf{p}_t - \hat{\mathbf{p}}_t^{(s)} \right\|_2 - \frac{1}{2S^2} \sum_{s'=1}^S \sum_{s=1}^S \left\| \hat{\mathbf{p}}_t^{(s')} - \hat{\mathbf{p}}_t^{(s)} \right\|_2 \quad (5.16)$$

with $\hat{\mathbf{p}}_t^{(\cdot)}$ as trajectories distributed according to the predictive multivariate CDF $\hat{\mathbf{F}}_t$, and $\|\cdot\|$ is the $K \times Z$ dimensional Euclidean norm. ES is averaged over the T number of forecast time series.

ES has a good discriminating ability to evaluate forecasts relying on marginals with correct variances but biased means. However, it is insensitive to misspecification of dependence structures [99] and weak at discriminating multivariate forecasts when correlations between their components make the only distinction between them [100].

5.5.2 Variogram-based Score

Due to the limitations of ES on detecting misspecified dependence structures between elements of a multivariate quantity, Variogram-based Score (VS) as a proper score is introduced in [100]. VS is discriminating on misspecified means, variance and correlations. For a d -variate observation and $\hat{\mathbf{F}}_t$, VS of order γ can be written as

$$VS_t = \sum_{i,j=1}^d w_{ij} \left(|p_{t,i} - p_{t,j}|^\gamma - E_F |\hat{\mathbf{P}}_{t,i} - \hat{\mathbf{P}}_{t,j}|^\gamma \right)^2 \quad (5.17)$$

where $\hat{\mathbf{P}}_{t,i}$ and $\hat{\mathbf{P}}_{t,j}$ are the i^{th} and j^{th} element of random variable $\hat{\mathbf{P}}_t$ distributed according to $\hat{\mathbf{F}}_t$ for which the γ^{th} absolute moment exists. With a set of trajectories as a representative of

$\hat{\mathbf{F}}_t$, forecast variogram can be approximated by

$$\mathbb{E}_F |\hat{\mathbf{P}}_{t,i} - \hat{\mathbf{P}}_{t,j}|^\gamma \approx \frac{1}{S} \sum_{s=1}^S |\hat{\mathbf{p}}_{t,i}^{(s)} - \hat{\mathbf{p}}_{t,j}^{(s)}|^\gamma, \quad \forall i, j = 1, 2, \dots, d. \quad (5.18)$$

VS_t then is averaged over the time series available in the evaluation dataset as

$$VS = \frac{1}{T} \sum_{t=1}^T VS_t \quad (5.19)$$

It is discussed in [100] that down-weighting pairs of components of $\hat{\mathbf{P}}_t$ which are expected to have relatively low correlations would be helpful to improve signal to noise ratio and alleviate sampling error. Because it is expected that pairs of the elements with higher distance or longer intervals will have lower relative correlations, pairwise w_{ij} can be defined proportional to inverse distance between the i^{th} and j^{th} components of $\hat{\mathbf{P}}_t$.

5.5.3 Event-based Score

The scores mentioned above are helpful when the goal is to discriminate and compare the quality of different sets of trajectories. However, they do not identify and inform about the ability of a set of forecasts in mimicking the stochastic process and predicting the likelihood of the events occurring in the concerned process. An event-based score is proposed in [99] to fill this gap. This score measures the capability of a set of predicted trajectories for predicting long-lasting and gradient events. In [99], long lasting events are defined as the successive hours with wind generations continuously more than a predefined threshold. A gradient event also is described as the maximum absolute ramp which has realized or predicted to occur over a window of size h and centred on the k^{th} lead time.

There is a fundamental difference between wind and PV variations. Changes in PV generations can be divided into fairly predictable variability and uncertainty. Variability of PV power is the consequence of earth rotation around its own axis and can be described by clear sky solar irradiance. However, the intermittency attached to PV power is caused by various factors such as cloud passages. Even with clear sky and no sources of intermittency, it is not expected that PV generations stay in the same level for successive hours. This variability, for the solar irradiance is highly predictable and can be considered as deterministic variations, although when it comes to PV generation the accuracy of estimation would be lower. Therefore, different from

the case of wind power, for PV generation each time of day should be treated different from the others and dependent on the maximum expected power for that time.

The maximum expected power can be assumed to be the expected non-overcast PV generation. Depending on the available information for the location of interest, various models can be found to estimate the clear sky solar irradiance. There is no perfect model for solar irradiance to PV power conversion though. Parameters such as shadow effect of building and trees, dust on the panels, maximum power point tracking system, power electronic circuits connected to the PV system cannot be modelled in a straightforward way. Therefore, inspired by [15], quantile regression as a statistical smoothing method is deployed here. The quantile with the coverage level of 99% is regarded as the maximum expected PV generations. Then, for event-based analyses, the intermittency with respect to the assumed maximum expected power is measured.

For the PV generation, three events are proposed in this study.

- **Event 1: Highly intermittent hours** This event describes the proportion of daytime when the difference between the measurement and maximum expected power exceeds a predefined threshold. Those periods are regarded as highly intermittent and less predictable hours. Comparing to less intermittent periods, during those hours more reserve or flexible generation powers are required to compensate fluctuations. This event can be mathematically written by

$$g_t(\mathbf{p}_t; \mathbf{p}\mathbf{c}_t, k, h, \xi) = \prod_{i=k-h/2}^{i=k+h/2} \mathbf{1}\{|p_{z,t+i} - p_{z,t+i}| \geq \xi\} \quad z = 1, \dots, Z. \quad (5.20)$$

where $p_{z,t+i}$ denotes maximum expected power for time t , zone z and step size i . ξ is the threshold value and h is the window size. k is the center of the window and so can be considered as a form of look-ahead time. All these three parameters are application dependent and are decided by the decision-maker. $\mathbf{1}\{\}$ is an indicator variable with a value of one if the condition realizes and zero value otherwise.

- **Event 2: Long-lasting intermittent hours** The difference between this event and event 1 is that, event 2 monitors the periods with intermittency continuously more than a predefined thresholds. The formulation is given by

$$g_t(\mathbf{p}_t; \mathbf{p}\mathbf{c}_t, k, h, \xi) = \prod_{i=k-h/2}^{i=k+h/2} \mathbf{1}\{|pc_{z,t+i} - p_{z,t+i}| \geq \xi\} \quad z = 1, \dots, Z. \quad (5.21)$$

• **Event 3: Gradient predictability**

If the difference between maximum and minimum deviations of measurements from the maximum expected power over a window of size h and centered on the k^{th} exceeds a predefined threshold, this status can be regarded as a gradient event. A gradient event can be given by

$$g_t(\mathbf{p}_t; k, h, \xi) = \mathbf{1}\left\{\left(\max_{i \in \{k-h/2, \dots, k+h/2\}} |\Delta_{z,t+i}| - \min_{i \in \{k-h/2, \dots, k+h/2\}} |\Delta_{z,t+i}|\right) \geq \xi\right\} \quad (5.22)$$

with

$$\Delta_{z,t+i} = pc_{z,t+i} - p_{z,t+i} \quad (5.23)$$

Other related events can be formulated depending on requirements of the decision-maker. For example, in gradient events, one might be more concerned about the times during the noon time when PV generation drops below a predefined threshold.

Eq. (5.20) to (5.22) represent events in an observed trajectory. To calculate the proportion of the trajectories predicting the concerned event, (5.24) can be used.

$$\hat{g}_t(\hat{\mathbf{p}}_t, \mathbf{p}\mathbf{c}_t; \theta) = \frac{1}{S} \sum_{s=1}^S g_t(\hat{\mathbf{p}}_t^{(s)}, \mathbf{p}\mathbf{c}_t; \theta) \quad (5.24)$$

where h , k , and ξ are substituted by parameter set θ .

Brier Score (BS) as a proper scoring rule then can be employed to measure the quadratic difference between probability forecast of a set of trajectories and observed one for a defined event by

$$\text{BS} = \frac{1}{T} \sum_{t=1}^T \left(\hat{g}(\hat{\mathbf{p}}_t^{(s)}; \theta) - g_t(\mathbf{p}_t; \theta) \right)^2 \quad (5.25)$$

5.5.4 Probability Integral Transformation (PIT)

To evaluate the probabilistic calibration of generated trajectories with respect to marginal predictive distributions, PIT histograms can be employed. In PIT histograms, the percentage of generated scenarios lying between two successive quantiles of marginal distributions are compared with the difference between nominal probability of those quantiles (known as ideal case).

5.6 Results

In order to get an insight about the aggregated uncertainties in the system, it is crucial to model the related dependencies. The database described in section 5.2 is used to investigate dependency structures existing in PV generations. Time-dependency is studied here by looking at zone 1 only, hence neglecting correlations in space. Space-time correlations are also investigated by treating all three zones and forecast horizons simultaneously. The forecast horizons in this practice are 6 to 19 hours ahead. It is to be noted that NWP for the dataset used in this thesis are issued at midnight. For instance when predicting PV generations for 6 am and 12 pm, the forecast horizons are 6 and 12 hours respectively.

To get an idea of how PV generations for the three zones used in this practice are correlated in space, a pairwise scatter plot of historical data is provided in Fig. 5.1. From Fig. 5.1, it can be observed that pairwise zones are highly and positively correlated. PV generations for diverse geographical regions can also be negatively correlated. This means that for instance if the western part of a region is rainy, there is a certain probability that the eastern part is experiencing a sunny time.

5.6.1 Benchmark methods

5.6.1.1 Naive method

A natural benchmark here would be an analogy-like approach. In this method, trajectories are randomly selected from the past observations. For instance, if the number of required trajectories is S , the observed time series for S randomly selected days from the past act like the set of predicted time-dependent trajectories for the current day. It is like assuming that one possibility is that what happens today is exactly the same as one of the previous days. If

Figure 5.1: All pairwise scatter plots of PV measurements in three zones

the number of trajectories was tending towards infinity, one would end up with a generalization of climatology forecast in the multivariate context.

5.6.1.2 Independent method

In this benchmark, for each t , k and z , trajectories are generated by uniformly sampling from the corresponding predictive marginal distribution. Therefore, this benchmark takes advantage of probabilistic forecasts though neglects space-time correlations.

5.6.1.3 Independent Gaussian

In probabilistic decision-making approaches such as Monte Carlo based methods, usually it is assumed that forecast errors follow a normal distribution centered on point forecasts. This benchmark is considered here to investigate the benefits of probabilistic forecasts over the common practice of normality assumption for uncertainty attached with intermittent generations. In this benchmark, trajectories for each t , k and z are generated by sampling from a normal distribution centred on corresponding point predictions.

In order to find standard deviation values which can reasonably describe uncertainties around point predictions, two options are tried. In the first option, the standard deviation for each t , k and z is calculated using historical data. For the second option, it is assumed that standard deviation for each t , k and z is $\beta\%$ of the corresponding point prediction. As the calculated scores for the second option found to be much better than the first one, here only results for the second option are given. Two common assumptions for the value of β are 5 and 10. The same values are chosen here for independent Gaussian-5 % and independent Gaussian-10 % respectively. It is to be noted that point forecasts are generated by deploying the method proposed in [125].

5.6.1.4 Gaussian copula with full empirical covariance

This benchmark is considered to make a comparison between multivariate normal distributions with the recursive covariance matrix and empirical one. Hence, in benchmark 5, the covariance matrix is calculated using historical data available up to time t .

5.6.2 Time grid

High dependency of PV generations on seasonally sunshine variations enforces a major difference between time dependent wind and PV generations. Given data with hourly resolution, sunrise and sunset vary seasonally (or monthly). In order to keep the size of the covariance matrix fixed and at the same time take into account sunrise and sunset changes, one approach is to project the information originally available (of varying dimension, since the beginning and end of day may change) onto a grid called time grid whose dimension stays fixed. For each day, the first element of the corresponding grid is the first PV measurement after sunrise and the last one is the last measurement before the sunset. The rest of the elements are found

Figure 5.2: PV observations (yellow colour curves) along with 49 prediction intervals with coverage ranging from 0.02 to 0.98 by 0.02 increments (from the darkest to the lightest)

by interpolation among values for daytime hours. Then, for all the covariance modelling, only information on this grid is considered. Eventually, when generating trajectories, they have to be transformed back to the original space and skill verifications are performed in the original space. The size of the grid is chosen to be 15 while there is no restriction on selecting a higher grid size.

5.6.3 Marginal distributions

Fig. 5.2 illustrates probabilistic forecasts in the form of non-parametric predictive distributions obtained using quantile regression [119] for the three randomly selected days from the evaluation dataset. In this figure, column 1 represents data for zone 1, column 2 data for zone 2 and column 3 data for zone 3. Univariate predictive distributions are described by 99 quantile forecasts with nominal coverage ranging from 0.01 to 0.99 with 0.01 increments. Then, as by

products of quantile forecasts, 49 prediction intervals with nominal proportion from 0.02 to 0.98 with 0.02 increments are generated. Each prediction interval is central on median and is generated by using two quantile forecasts with symmetrical nominal probabilities with respect to the median. Twelve independent NWP variables described in section 5.2 are used as the predictors and PV power is treated as the output of the quantile regression.

In order to benefit from quantile forecasts, they should be calibrated. Calibration of the quantile forecasts can be verified via a reliability diagram in which the proportion of the measurements with values less than each quantile of the marginal distribution is compared with the nominal probability of the corresponding quantile (ideal case). Fig. 5.3 illustrates the reliability diagram for zone 1 when all forecast horizons are covered. As can be seen, the deviations from the ideal reliability are quite low and that lets us use the generated marginal distributions for further investigations.

Figure 5.3: Reliability evaluation of 99 quantiles with 0.01 increasing nominal proportion generated by quantile regression for zone 1 and all forecast horizons.

Skill verification of predictive marginal distributions based on CRPS is visualized in Fig. 5.4 for all three zones available. From the figure, it can be concluded that for zones 1 and 3, marginal densities are less skilful around 9 am in the morning while zone 2 has its peak in the afternoon. However, for all zones it can be seen that CRPS mimics the typical daily sunshine

Figure 5.4: Skill evaluation of the probabilistic forecasts of PV generation as a function of forecast horizon

curve with lowest values close to sunrise and sunset.

5.6.4 The goodness of fit of Gaussian Copula

In order to verify the goodness of fit of the Gaussian copula, here four approaches are given. Firstly, the goodness of fit of Gaussian copulas used to model PV generation at pairs of time intervals can be visualized by scatterplots. We can compare the observed ranks versus the simulated bivariate Gaussian process having the same rank correlation. Also, the dependency patterns of real measurements can be compared with random draws of the fitted bivariate Gaussian distribution. Fig. 5.5 provides the scatterplot of the observed PV power measurements at 15:00 versus those at 16:00. In the same figure, simulated bivariate Gaussian data is plotted. The bivariate Gaussian copula is fitted to data using maximum likelihood estimator. As seen

in this figure both patterns are very similar. This indicates that the Gaussian copula could be considered a suitable option for describing the spatio-temporal dependency structure. The same conclusion can be drawn from Fig. 5.6 where the observed ranks of PV power at 15:00 ($(y_{t,15:00} = \hat{F}_{t,15:00}(p_{t,15:00}) \forall t)$) versus those at 16:00 ($(y_{t,16:00} = \hat{F}_{t,16:00}(p_{t,16:00}) \forall t)$) are plotted. The results obtained while considering other pairs of space-time variables are all deemed qualitatively similar.

Figure 5.5: Comparison of real measurement and the simulated Gaussian process for the dependency between PV power at 15:00 and 16:00 over the all data available for Zone 1.

Although the visual comparison provide basic idea on the goodness of fit of Gaussian copula for the application in this thesis, there are various fit evaluation criteria which could be used if one aims at finding an optimal copula. The corrected Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are among the mostly used criteria for evaluation of best fitting model [126]. In theory, likelihood is a function of the parameters of a statisti-

Figure 5.6: Comparison of empirical and Gaussian copulas for the dependency between PV power at 15:00 and 16:00 over all data available for zone 1.

cal model given observations. It can be calculated as the product of the probability density functions evaluated at the observed data values. Often it is preferred to work with the natural logarithm of the likelihood function, the so-called log-likelihood function which is defined by

$$\log \mathcal{L}(\theta|P) = \sum_{i=1}^n \log f_i(\theta|P_i) \quad (5.26)$$

with $\mathcal{L}(\theta|P)$ as the set of model parameter values θ given observations P and n is the number of observations.

AIC is defined as

$$AIC = -2 \log \mathcal{L}(\theta|P) + 2\kappa \quad (5.27)$$

Table 5.2: Comparison of BIC and AIC values for various copulas

	BIC	AIC
Gumbel	-329.8	-333.8
Clayton	-445.4	-449.3
Frank	-310.1	-314.1
Joe	-189.2	-193.2
Gaussian	-4351.9	-4571.4

where κ is the number of estimated parameters in the set θ . The model with the lowest AIC is expected to have a higher goodness of fit.

BIC is formulated as

$$BIC = -2 \log \mathcal{L}(\theta|P) + \kappa \ln(n) \quad (5.28)$$

Similarly, the model with the lowest BIC is more preferred. In this thesis, an empirical experiment is performed to compare Gaussian copula versus Joe, Clayton, Gumbel and Frank copulas in terms of their BIC and AIC for the data used in this chapter. Table 5.2 gives the results of this experiment. As can be seen in the table, Gaussian copula provides much better goodness of fit than the other copulas.

5.6.5 Multivariate analysis

A set of 20 space-time trajectories for three randomly selected days (the same days as those in Fig. 5.2) for all three zones are depicted in Fig. 5.7. In Fig. 5.7, column 1 represents data for zone 1, column 2 data for zone 2 and column 3 data for zone 3. Trajectories are drawn from the joint Normal distribution as explained in Subsection 5.4.3. As can be seen in the figure, generated scenarios are both temporally and spatially correlated. With temporal correlations, generated scenarios would not evolve through lead times as a random walk whose magnitudes are formed and limited by marginal densities. Likewise, correlations in space ensures that by looking at a zone the level of uncertainties in the other zones can be estimated.

Figs. 5.8 and 5.9 illustrate the reliability diagram and PIT histogram over whole evaluation dataset containing all look-ahead times for temporal and space-time trajectories respectively. It should be emphasized that these two evaluation criteria measure reliability of trajectories

Figure 5.7: PV observations (dark black colour curves) along with 20 generated space-time trajectories (grey colour curves)

with respect to the marginal distributions. Therefore, in case the later ones as the inputs of the dependency study are not reliable, unreliability would be expanded to the generated trajectories. In Figs. 5.8 and 5.9, the perfect situation is the case where each bin contains 5 percent of the generated trajectories. The ideal case of perfect probabilistic correctness is represented by the dash line. Figs. 5.8 gathers the results over the whole evaluation dataset for all look-ahead times in zone 1. The PIT histogram in Fig. 5.9, however, gathers the results for all look-ahead times and zones. As can be seen deviations from the ideal line are very low which indicate that trajectories are distributed identical to the predictive distributions.

The skill of all the methods described in this thesis are verified based on the skill scores

Figure 5.8: Time trajectories: evaluation of the probabilistic correctness of generated scenarios with a PIT histogram.

explained in section 5.5 namely ES, VS and BS. To investigate the impact of the number of generated trajectories on discriminating ability of a method, sets of trajectories in 5 different sizes $S = 20, 100, 200, 400$ and 3000 are generated. Hereafter, dependency modelling based on multivariate normal distribution with recursive covariance is denoted by MVN. The scores for MVN and five other benchmarks for time and space-time dependencies are visualized in Figs. 5.10 and 5.11 respectively. For the Variogram-based score, two different settings are assessed. For VS1, the weights for all pairs of the components of $\hat{\mathbf{P}}_t$ are set to 1. Moreover, in order to degrade the impact of pairs of components with relatively lower correlations, in VS2 $w_{i,j}$ is set to the historical correlation between components i and j . As reported in [100], γ equal to 0.5 for most cases provides good discriminating ability. Therefore, for both VS1 and VS2, γ is set to 0.5. BS1, BS2 and BS3 represent the BS for events 1, 2 and 3 as explained in subsection 5.5.3 respectively. For BS1 and BS3 $k = 11$ and $h = 4$ and for BS2 $k = 11$ and $h = 2$. ξ for all three events is equal to 0.2. It is to be noted that in all tests for the MVN, Independent and Gaussian copula with full empirical covariance approaches, the same marginal distributions are

used. The score for each experiment is calculated by averaging over 30 independent runs and the values are rounded to 3 digits.

Figure 5.9: Space-time trajectories: evaluation of the probabilistic correctness of generated scenarios with a PIT histogram.

By looking at Figs. 5.10 and 5.11, it can be concluded that the impact of number of trajectories varies from one score to another. Increasing the number of trajectories has almost no distinguishable influence on ES when evaluating PV power trajectories. This feature, on one hand can be viewed as an advantage when only limited trajectories are available such as the case with small size ensemble. On the other hand, it can be interpreted as a weakness of this score at discriminating between a very good representation of an predictive density of PV power and a very sparse one. By increasing the number of trajectories, though, VS scores improve continuously for all the four settings. A higher number of trajectories as a better representative for predictive density can also get a better BS, although the improvements are not as noticeable as the case with VS.

The skill of Independent benchmark has its lowest difference with that of MVN when the evaluation metric is ES. This supports the conclusion in [99] that ES is weak at discriminating

Figure 5.10: Temporal trajectories: the scores for multivariate normal distribution with recursive covariance and five other benchmarks. 5 different sizes for the trajectories set are considered.

multivariate forecasts when correlations between their components make the only distinction between them.

Analysing Figs. 5.10 and 5.11 reveals that for both time and space-time dependency, MVN has the best or close to best discriminating ability to detect highly intermittent time of day, long lasting and gradient intermittency. The Gaussian independent method with standard deviation 10% of the point forecasts presents better skills in comparison with the one with 5% standard deviation.

Overall, the Independent method shows better performance than independent Gaussian. This informs about the benefits of deploying probability forecasting to model predictive distributions of PV generations for each time in the future because normality assumption is too naive to model variations of PV power.

The skill of the MVN and Gaussian copula with full empirical covariance are quite comparative which means that the recursive approach is able to mimic the dependence structure of the multivariate quantity. Therefore, in case the historical data is not available to calculate the empirical covariance matrix, the recursive approach can be a proper alternative.

Analysing the skill of the Independent benchmark and MVN can reveal the importance of modelling dependency structures. While the independent benchmark takes advantage of prob-

Figure 5.11: Space-time trajectories: the scores for multivariate normal distribution with recursive covariance and five other benchmarks. 5 different sizes for the trajectories set are considered.

abilistic forecasts, it treats each look-ahead time k and location z independently. However, intermittent generations are dependent both in time and space. To get an idea about the significance of dependency of PV generations in time and space, the empirical covariance matrix for the transformed observations on the time grid is visualized in Fig. 5.12. According to this figure, time dependency decreases as the difference between lead times increases. Furthermore, there is high dependency in space specially for adjacent lead times. As the generated trajectories are to be used in operational problems, providing miscalibrated forecasts based on wrong assumption of independency leads to suboptimal decisions.

Although Independent benchmark is insufficiently skillful with respect to MVN, it provides much better ES in comparison with Gaussian independent methods. In terms of VS, again the Independent approach surpasses Gaussian independent benchmarks. This can validate that probabilistic forecasts can offer more accurate information about the uncertainties involved than the naive assumption of normality distribution for forecast errors.

According to Fig. 5.11, the naive method shows the best BS2 and BS3 scores. However, further investigations showed that as the ξ increases (toward the rare events), MVN gets better BS2 and BS3 in comparison with the naive method. Fig. 5.13 illustrates BS2 and BS3 for the MVN, naive and Independent methods for five different values of ξ . As can be seen for ξ of more than 0.2, BS2 scores given by the MVN and Independent methods are better than those

Figure 5.12: Visualization of correlation matrix for the transformed measurements on the time grid, the matrix is of size of 45×45 . Red to white colours represent the lowest to the highest correlations

given by the naive method. The same results obtained for BS3 when ξ is of more than 0.3. It can be concluded that the scenarios generated by the naive method are able to describe the frequent events, however, they are weak at informing about the less frequent or rare events.

It is to be emphasized that in generating space-time trajectories, the aim is to capture the development of forecast errors through successive lead times and interdependent generation in contiguous locations. This is in contrast with point forecasts where the target is to minimize the square prediction errors.

Considering solar spectral distribution among the other relevant explanatory variables can potentially improve point forecasts and even lead to generation of marginal distributions with

Figure 5.13: Space-time trajectories: BS2 and BS3 scores for MVN, naive and Independent methods for five different values of ξ . For BS2 $k = 11$ and $h = 2$ and for BS3 $k = 11$ and $h = 4$.

higher quality. However, it does not impose any changes in the space-time correlation framework used here. The initial aim of the thesis is to demonstrate that in case the calibrated predicted marginal distributions are available (no matter which explanatory variables are used to generate them), to what extent the space-time dependency modeling can add value to forecasting. This is the reason that in the thesis, the “Independent” benchmark is introduced which deploys the same marginal distributions as those of the proposed method but neglects the space-time dependencies. Then, when the methods are compared based on the similarity of their predicted trajectories with the measured ones, we can see that space-time dependency modeling can make a noticeable difference in the quality of forecasts and leads to better skill scores. If more relevant explanatory variables are available to predict marginal distributions, it is expected when those marginal distributions are plugged into the proposed framework for the space-time correlation modeling, the superiority of the predicted trajectories over the introduced benchmarks can be even more noticeable.

As a final remark, it is to be noted that generating these trajectories is not straightforward, and to the best of our knowledge, there is not any established method today for application to

solar PV generation that we can make a comparison with it. Hence, in this thesis, the importance and the value of such space-time dependency modeling are investigated by introducing a few relevant benchmarks. In designing the benchmarks, it is tried to compare this proposal with the common practices in uncertainty modeling.

5.7 Conclusions

As there is always a level of uncertainty attached to the PV generation forecasting, uncertainty quantification is an integral factor to maintain acceptable levels of reliability, security and profitability in power systems. Due to the inertia of meteorological systems, prediction errors are correlated in time and space.

The main contributions of this chapter are to develop a framework for generation of multivariate trajectories of PV generation and to propose scoring rules to evaluate their performances. In order to track time dependency of PV generations for successive lead-times while taking into account seasonally variations of sunrise and sunset time, it is proposed to transfer the observations and marginal distributions to a time grid with a predefined fixed size. The dependency study is carried out on this grid. Eventually, the generating trajectories are transformed back to the original space. In this study, investigations are performed to evaluate the goodness of fit of Gaussian copula in describing the joint characterization of PV power generation in time and space. The goodness of fit of Gaussian copula is compared with a number of Archimedean copulas namely Frank, Joe, Clayton and Gumbel in terms of AIC and BIC. The joint predictive distributions based on marginal densities are modelled and space-time trajectories describing evolution of forecast errors through successive lead times and locations are generated. Several scoring rules are used to measure the similarity of different set of forecasts with respect to the PV generation observations. Three events are proposed which cater for the specific case of PV generations. These events measure the level of intermittency in PV generations with respect to the maximum expected PV power for each time of day. A number of benchmark methods are introduced to allow evaluation of the spatio-temporal modelling leading to emphasize on the potential values that multivariate analysis can add to decision making.

Analysing the results validates the superiority of the probabilistic forecasts in describing the uncertainties over the common practice of normality assumption for forecast errors. Moreover, trajectories drawn from multivariate distributions present better skill in almost all the exper-

iments. Therefore, in operational problems which are sensitive to the dependency structures, one may conclude that using probabilistic forecasts when space-time correlations are neglected may provide only a suboptimal solution. The investigation verified that in case that historical data is not available to find dependency structures of PV generations, modelling covariance matrix recursively can be viewed as an alternative approach.

While in this thesis, multivariate distribution is assumed to be Gaussian, further investigations are required to find the best copula function which can describe the behaviour of the PV generations in time and space. Therefore, in the future work, it will be tried to perform various tests to evaluate Archimedean, empirical and T copulas and find the right one for the case of PV generations.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

PV power generation has experienced an extensive upward trend in power systems over the last decade. However, the unpredictability and intermittency in PV generation are still the most severe challenges in operation and control of power systems with high penetration of PV energy. Improvement and development in the PV power forecast sector are crucial requirements for managing the related variability and intermittency.

This thesis proposed a number of complementary solutions for prediction and uncertainty quantification of PV power generation. The proposed solutions presented in three different forms of point forecasts, predictive marginal densities and space-time trajectories.

As far as point forecasts are of interest, two statistical prediction models are developed, one based on batch learning and the other one based on sequential learning. The former one suits for longer prediction horizons and lower temporal resolutions e.g. day-ahead prediction while the former one is designed as an adaptive and ensemble technique for the case of online forecast with very short-term lead times and high forecast periodicity. The adaptive and ensemble model called nOSELM can address the need for forecasting in newly established PV sites with no long archived data available for training and fitting the model. The proposed model can be trained with very few training samples and continuously get updated when the real-time measurements are available. This is an important feature when dealing with non-stationery time series because the model can be updated and adapted based on specific kinds of dynamic and fluctuations which exist at the time of forecasting. Each of the ensemble models in nOSELM can be trained

for a specific regime in the dynamic of PV generation time series. The results showed that for the specific datasets used in this thesis the predictive performance of the adaptive models is better than that of the offline training models in terms of the commonly used accuracy metric in point forecasting.

The emphasize in this thesis is to give credit to probabilistic forecasts rather than deterministic forecasts, i.e. point or single-valued forecasts. Forecast users may still prefer to get point forecasts, because they are more transparent, easier to use and understand. However, there is always a non-negligible level of uncertainty attached to forecasts, meaning that forecasts are always precisely wrong. Therefore, the decisions and policies made based on point forecasts can only be suboptimal. Decision-makers need to have an estimation of the uncertainty while this cannot be provided by point forecasts. The accuracy metrics like *RMSE* or *MAE*, only describe the historical performance of a method not the uncertainty associated to each forecast.

When this research was initiated, the area of probabilistic forecast of solar energy was barely touched. In this project, the characteristics of forecast errors of PV power generation is investigated to gain an insight on the stochastic behavior of forecast errors. One-sample Kolmogorov-Smirnov (KS) test is conducted to hypothesize relevant parametric distribution to describe the stochastic behavior of PV generation forecast errors. The empirical investigations revealed that PV power forecast errors do not shape any kinds of parametric distributions. Without using additional explanatory variables, e.g., from sky imaging, it is unlikely that the empirical distributions of forecast errors could look much different. Therefore, to avoid restrictive assumptions on the shape of the forecast densities, a nonparametric density forecast method is proposed which consists of regression and optimization phases. For quantification of uncertainty associated with forecasts, uncertainties are modeled in the forms of quantile forecasts and prediction intervals.

The proposed method to generate density forecasts is performance oriented and estimates the quantiles based on the evaluation criteria of the probabilistic forecasts. The proposed nonparametric method was successfully applied on two PV power datasets with one-minute resolution and highly fluctuating patterns.

In this thesis, it is discussed that despite the substantial information provided by probabilistic forecasts they are only suboptimal inputs to decision-making when problems are temporally and/or spatially coupled. Marginal predictive densities described above are estimated on per location, per lead time basis. Hence, they do not provide information about the interdepen-

dence structure of uncertainty such as the overall forecast errors of PV generations spread over the control zone or the evolution of forecast errors over a time period. Due to the inertia in meteorological systems and complex meteorological mechanisms like cloud passages, PV generations show residual cross-correlations in time and in space. Therefore, it is highly plausible that leveraging spatio-temporal dependencies can lead to an improved forecast accuracy. As very first studies in the area of space-time correlation modeling of PV generation, a framework was developed for generation and evaluation of spatio-temporal trajectories of PV generation.

In order to appraise the predictive performance of our proposed frameworks for estimation of space-time dependencies, rival benchmarks based on the most common practices for representation of uncertainty in power systems are designed. The skill of the forecasts given by space-time modeling is compared with the case of trajectories generated directly from predictive marginal distributions, uncovering the facts that space-time trajectories are able to describe the underlying stochastic behavior of interest much better. In addition, a benchmark is devised to compare the predictive performance of our spatio-temporal model with the case of commonly used normality assumption for uncertainty estimation.

Rival approaches are compared based on a number of relevant skill evaluation criteria, revealing a higher skill of the proposed dependency modeling. Three event-based skill scores are proposed for the specific case of PV generation where PV generations are compared with clear sky solar irradiance.

The empirical results show that for the case of space-time trajectories of PV generation, Variogram-based score has a much better discrimination ability than the energy score. When evaluating PV power trajectories, It is observed that increasing the number of trajectories has almost no distinguishable influence on energy score while it continuously improves Variogram-based score. This can be interpreted as a weakness of energy score in discriminating between a very good representation of an predictive density of PV power and a very sparse one.

6.2 The Applications of the Developed Forecasting Frameworks

In practice, the applications of forecasting methods can be analyzed based on two specifications. The first one is the forecast horizon and the second one is the type of forecasts, namely point, probabilistic or full space-time characterization. In the following, a number of potential applications for the various methods proposed in this thesis will be reviewed in terms of the

above-mentioned specifications.

6.2.1 Forecast horizon

In this thesis, 10-minute, 1-hour, 1 to 24-hour ahead (day-ahead) forecast horizons are considered. In California which accommodates the highest penetration of photovoltaic energy in the US, the independent system operator (CAISO) uses day ahead and hour ahead forecast horizons. CAISO also is planning to deploy intra-hour forecasts on 5-minute to 10-minute forecast horizons. In Europe also, day-ahead forecasts are required throughout the different power exchange markets. 10 minutes forecasts have applications in balancing/real-time/regulation markets, load following and baseline ramps. Among the applications of the hours (short-term) forecasts, we can refer to unit-commitment, scheduling conventional plant, dynamic ED, and designing trading strategies of wind/solar plant owners. Day(s) ahead forecasts have applications in day-ahead scheduling and large CHP operation.

6.2.2 Forecast type

The most commonly used type of forecast is point forecast. Although point forecasts are easily understandable to the end-users, they are always subject to errors. If the consequences of relying on inaccurate forecasts are significant, it is required to provide a full picture of the estimated error using probabilistic forecasts. Therefore, relying on point forecasts depends mainly on the consequences of making forecast errors in a particular application and the extent to which the decision maker can tolerate loss. The desired type of forecasts is also grounded on the expertise of decision-makers. The reason is that both implantation and deployment of probabilistic forecasts require much higher expertise than those of point forecasts. Probabilistic forecasts can take over the role of point forecasts in many of the well-known power systems operation and planning applications such as unit-commitment, optimal power flow and economic dispatch [127–129], operating combined solar-wind-storage, reserve quantification [130], assessment of operating costs [131]. There is information provided by probabilistic forecasts, which cannot be given by point forecasts: such as what is the probability that PV generation does not exceed a certain level at a particular time. In [132], it is mentioned that when the penetration of intermittent generation in the system is more than 10%, basic point forecasts are not sufficient and reliable probabilistic forecasts are needed. Generally, the applications of the probabilistic forecasts in well-known power system operational problems such as UC and OPF can be justi-

fied by the fact that for a large class of decision making problems the optimal solution is linked to a specific quantile rather than the expectation of the future outcome [21, 133].

It is to be noted that deployment of the probabilistic forecasts requires reformulating the deterministic problems as stochastic ones. Stochastic, interval, chance constrained, robust optimization approaches can be used to solve those problems. Stochastic, chance constrained and robust unit-commitment and optimal power flow are examples of applications among others.

In the probabilistic energy forecasting literature, emphasis is mainly placed on generating marginal predictive densities for which each random variable is dealt with individually, on per look-ahead time and per location basis. However, following complex meteorological mechanisms like cloud passages, PV generations act like a random variable distributed over time and space. Therefore, it is highly plausible that by leveraging spatio-temporal correlations, improved forecast accuracy can be gained.

Although probabilistic forecasts provide valuable information about the uncertainty associated to forecasts, they fail to inform about interdependence structure of uncertainty at different locations and/or different lead times. Therefore, if operational problems are spatially and/or temporally coupled, both point forecasts and probabilistic forecasts would be only suboptimal inputs to decision-making. Stochastic unit-commitment, optimal power flow, operation of combined renewable-storage are examples of operational problems which requires information about uncontrollable generation spread over a control area as well as the development of forecast errors through successive lead times. Space-time dependency modeling also can answer to the questions such as

- What is the probability that a given storage is large enough to balance production/consumption for the next four hours?
- What is the probability that up-regulation is going to exceed a certain level over the next five hours due to a decrease in PV generation?
- What is the probability that PV generation will exceed a certain level over the next three hours?
- What is the probability that PV generation at site A be less than X_1 and PV generation at site B be less than X_2 and PV generation at site C be less than X_3 at a same time.

As one of the simple implementations of space-time trajectories in power systems, scenario-

based unit-commitment can be considered. In this case, instead of generating scenarios independently from marginal distributions, they are generated from the predictive multivariate distribution and are deployed in the same optimization framework as before. By considering space-time dependencies, the decision maker can be prepared for the aggregated forecast errors to be observed in different generation zones in his/her control area as well as the evolution of forecasts errors through successive time intervals. Space-time trajectories should be considered as one the most important components of the operation management of combined solar-storage units because the performance of storage units is highly time dependent.

6.3 Future Works

In this thesis, it is tried to provide the forecast users with high skilled forecasts in view of the extend to which they describe the underling characteristics of the stochastic process of interest, i.e., PV power generation. The reader should note that the skill of forecasts does not directly justify their value. Here, the value of forecasts represents the economic benefit obtained from deploying forecasts as inputs to decision-making. To appraise the value of forecasts, it is required to reformulate those operational problems which involve intermittent variables such that they can incorporate more advanced forms of forecasts, e.g. space-time trajectories. This requires more research in the realm of stochastic decision-making to find the relation between the skill of forecasts and the operational/economic benefits provided by them.

While in this thesis multivariate predictive densities are estimated through predictive marginal distributions, one interesting research direction would be to generate multivariate densities directly without the need to have marginals first.

In the following subsections, a few relevant research perspectives are discussed.

6.3.1 Finding the right multivariate copula for modelling spatio-temporal dependency in PV generation

In this thesis, It is assumed that temporal/spatio-temporal dependences in PV generation follow Gaussian distribution. While this is the most straightforward assumption, it might not be true in general. For example, a Gaussian copula model has zero tail dependence. This implies that there is asymptotically zero probability that random variables are in their extremes unless their Pearson correlation coefficient is unity [134]. More investigations are required to explore

different approaches for multivariate density construction to find the most proper one for the case of PV generation. The rival approaches (copulas) should be examined on their goodness-of-fit as well as skill scores. The potential copulas which can describe the behaviour of the PV generations in time and space are the Archimedean family of copulas (Frank, Clayton, Gumbel, Joe) [135], and T and vine (R-vine, D-vine and C-vine, truncated vine) copulas [136].

In theory and practice, Archimedean copulas are very popular and treated as the building block for more complicated constructions such as nested/hierarchical Archimedean copulas. In Archimedean copulas, dependency in all kinds of dimensions is modelled with only one parameter, governing the strength of dependence. This can be considered either an advantage or a drawback, depending on the type and strength of dependency, required accuracy and affordable computation burden.

Vine copula modeling is a pair-copula construction where a multivariate density function of dimension d is constructed using $d(d-1)/2$ bivariate copulas [137] and a nested set of trees. Because in vine copula modeling, multidimensional dependencies are factorized into pair-copulas, it is expected that this approach adds more flexibility into multidimensional dependency modeling. It is claimed that the vine copula can be more efficient in cases where different pairs of the dependent variables have different dependency structures [138]. Vine copulas leverage from bivariate copulas and enable extensions to arbitrary dimensions. The number of available parametric bivariate copulas are much more than the available parametric multidimensional copulas. This in turn extends the number of copula options for parametric modeling of dependencies [139].

Akaike and Bayesian Information criteria [126, 140] and Clarke and Vuong Tests [141, 142] as well as the energy, event-based and variogram-based scores are among the criteria which can be used to compare the fitness and skill of the rival copulas.

6.3.2 Estimation of multivariate prediction regions for random generation

The prediction intervals/quantiles proposed in Chapter 4, describe regions from the predictive marginal distribution of PV generation for each time and location individually. Where space-time dependency and modelling multivariate densities are of interest, the multivariate prediction regions would be preferred. A multivariate uncertainty region informs about the region where the next realization of correlated random variables for the stochastic process is expected to be observed with a certain probability.

The multivariate prediction regions can be used as the critical inputs for stochastic decision-making in operational processes which are temporally and spatially coupled. The regions should be used as the constraints in the operational problems such as optimal power flow and unit commitment. Chance constrained [143], robust [144, 145], interval [146] and scenario-based optimization can be used to solve those problems with multivariate prediction regions as constraints.

In [147–149], two methods are used to find multivariate (simultaneous) intervals. The methods are called Adjusted Intervals (AI) and Chebyshev-based Intervals. These approaches are originally introduced in [148]. The idea is to find the boundaries for intervals based on the proportion of the generated trajectories which are enveloped by them.

It is desired that if an interval envelops $(1 - \alpha)\%$ of the predicted trajectories, it also covers the same proportion of the observed trajectories in the long run. A trajectory of size of T is covered by an interval if for $t = 1, \dots, T$, the trajectory is inside or equal to the limits of that interval. These multivariate intervals are obtained for temporal trajectories in [147]. One can refer to [148] for the detailed description of the methodology.

The AI method for GEFCom2014 dataset, Zone 1 (refer to Section 5.2) [105] is simulated here where time-dependency is modelled only. PV generations from 6:00 am to 7:00 pm are studied. In Fig. 6.1, one can see the calibration of the simultaneous prediction intervals generated based on this method. Maximum deviation from ideal calibration for this method is 0.152 and average deviation is 0.0457.

In Fig. 6.2, PV observations along with 19 simultaneous prediction intervals with nominal coverage ranging from 0.05 to 0.95 by 0.05 increments, for 3 randomly selected days from the evaluation set are depicted. Looking at Fig. 6.2 reveals that the simultaneous prediction bands with lower nominal coverage are very wide. Furthermore, the differences between widths of the intervals with nominal coverage lower and higher than 50% are very low.

The simultaneous prediction intervals shown in Fig. 6.2 does not provide appropriate calibration, they are too wide and the approach for their construction is time-consuming. The target prediction region is different with the simultaneous prediction intervals described above.

I aim at developing multivariate ellipsoidal prediction regions (PEs). A PE represents a contour of constant density in the joint distribution of random variables (for elliptical densities). The predictive ellipsoids can be used as generation constraints as they are or they can be approximated by maximum-volume inner polyhedron and minimum-volume outer polyhedron.

Figure 6.1: Calibration of the simultaneous prediction intervals.

The goal is to develop an approach to construct PEs with lower volume and better calibration than simultaneous prediction intervals described above.

The PE has the following form,

$$PE = \{X \in R^p | (X - \mu)^\top B^{-1} (X - \mu) \leq \gamma\} \quad (6.1)$$

where μ is the center of the ellipsoid, B is the ellipsoid matrix which is positive definite, R^p is the Euclidean space of p and γ is called robust parameter.

Recently, ellipsoidal uncertainty sets have been used in robust optimization. However, the parameters of the ellipsoids, μ , γ and B are chosen based on assumptions [150,151]. A thorough reserach on prediction of γ for the specific nominal coverage is definitely a promising research direction and the outcome can be helpful in a variety of applications of stochastic programming.

I will also work on introducing new scoring rules for evaluation of PEs. The scoring rules should examine the prediction regions based on their collaborations and their volumes.

Figure 6.2: PV power observations (green colour curves) along with 19 simultaneous prediction intervals with nominal coverage ranging from 0.05 to 0.95 by 0.05 increments (from the darkest to the lightest), for 3 randomly selected days from the evaluation set. t refers to day number in the year.

6.3.3 Regime switching models for PV generation

PV generation time series with high temporal resolutions of few minutes or seconds exhibits fluctuations of different dynamic nature with dramatic breaks in their behaviour and magnitude. Therefore, a potentially useful approach to model the uncertainty in PV generation is studying

the regime switching aspects of the PV generation time series. This approach mostly relies on the Markov-switching AutoRegressive models [152].

6.3.4 Sky imaging

The most effective method for solar irradiance forecasts varies depending on the required spatial resolution and lead times. For lead times more than 6 hours and the spatial resolutions more than 2 km, NWP is the most accurate method. Satellite-based prediction provides the best accuracy for predictions with 1 km resolutions and the lead times 1 to 6 hours. For the spatial resolutions of less than 1 km and the intra hour prediction, though, ground sky imaging is the most optimal approach [41].

Ground-based sky imaging and cloud image processing techniques along with other weather data can be used to predict cloud location and movement. Clouds movement forecasts can be used as additional information in solar irradiance and PV energy generation prediction. This requires installing adequate sensors and sky imagers spread over the site of interest to measure weather variables such as temperature, solar irradiance, humidity and wind speed as well as data for estimation of clouds movement. The measured weather data along with clouds forecast can be used in our developed forecast approaches to provide valuable information on uncertainty quantification of the PV power.

In our models for very short-term predictions, cloud forecasts based on ground sky imaging are not used. A number of researchers have worked on taking advantage of ground sky imaging to improve PV generation point forecasts [153–155]. Investigating the benefits that very short-term cloud forecasts can bring about in improving predictive performance of probabilistic forecasts or space-time dependency modelling is a propitious research direction.

6.3.5 The impacts of PV power fluctuations on power market and demand response management

It could be of interest to study the impacts of PV power fluctuations on the energy or regulation market prices and then investigate how price-responsive loads can react to these fluctuations. In other words, the uncertainty forecasts of PV power can be employed in characterizing the uncertainty forecasts of energy price. The forecasts then can be provided to the flexible loads as a decision-aiding tool to manage their operation and risks.

List of Publications

- [1] Faranak Golestaneh, Pierre Pinson and Hoay Beng Gooi, “Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation-With Application to Solar Energy,” *IEEE Transactions on Power Systems*, Volume 31, pp. 3850-3863, January 2016.
- [2] Faranak Golestaneh, Hoay Beng Gooi and Pierre Pinson, “Generation and evaluation of spacetime trajectories of photovoltaic power,” *Applied Energy*, Volume 176, pp. 80-91, August 2016.
- [3] Faranak Golestaneh, Hoay Beng Gooi, “Batch and sequential forecast models for photovoltaic generation,” *2015 IEEE Power & Energy Society General Meeting*, July 2015, Denver, USA.
- [4] R. A-Abarghoee, Faranak Golestaneh, Hoay Beng Gooi, Jeremy Lin, Farhad Bavafad and Vladimir Terzija, “Chance Constrained Probabilistic Unit Commitment with Incentive Demand Response Management and High Wind Power Integration,” *Applied Energy*, Volume 182, pp. 634-651, November 2016.

References

- [1] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.
- [2] "Unlocking the sunbelt potential of photovoltaics, european photovoltaic industry association." [Online]. Available: <http://www.epia.org>
- [3] E. Ela, V. Diakov, E. Ibanez, and M. Heaney, "Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales," *Contract*, vol. 303, pp. 275–3000, 2013.
- [4] E. T. FLAGSHIP, "Solar intermittency: Australias clean energy challenge," 2012.
- [5] P. Pinson, "Estimation of the uncertainty in wind power forecasting," 2006.
- [6] R. H. Inman, H. T. Pedro, and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535–576, 2013.
- [7] R. Marquez, V. G. Gueorguiev, and C. F. Coimbra, "Forecasting of global horizontal irradiance using sky cover indices," *Journal of Solar Energy Engineering*, vol. 135, no. 1, p. 011017, 2013.
- [8] Y. Chu, H. T. Pedro, and C. F. Coimbra, "Hybrid intra-hour dni forecasts with sky image processing enhanced by stochastic learning," *Solar Energy*, vol. 98, pp. 592–603, 2013.
- [9] A. Mellit, M. Benghanem, and S. Kalogirou, "An adaptive wavelet-network model for forecasting daily total solar-radiation," *Applied Energy*, vol. 83, no. 7, pp. 705–722, 2006.
- [10] S. Al-Awadhi and N. El-Nashar, "Stochastic modelling of global solar radiation measured in the state of kuwait," *Environmetrics*, vol. 13, no. 7, pp. 751–758, 2002.
- [11] A. Mellit, M. Benghanem, A. H. Arab, and A. Guessoum, "A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of markov transition matrices approach," *Solar Energy*, vol. 79, no. 5, pp. 469–482, 2005.
- [12] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet, "Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation," *Energy*, vol. 36, no. 1, pp. 348–359, 2011.

- [13] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of arma and tdnn," *Solar Energy*, vol. 85, no. 5, pp. 808–817, 2011.
- [14] O. Kärner, "Arima representation for daily solar irradiance and surface air temperature time series," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 71, no. 8, pp. 841–847, 2009.
- [15] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.
- [16] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.
- [17] D. Sarewitz, R. A. Pielke, and R. Byerly, *Prediction: science, decision making, and the future of nature*. Island Press, 2000.
- [18] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets: Operational Problems*. Springer, 2014.
- [19] P. Pinson and H. Madsen, "Ensemble-based probabilistic forecasting at horns rev," *Wind Energy*, vol. 12, no. 2, pp. 137–155, 2009.
- [20] A. Timmermann, "Density forecasting in economics and finance," *Journal of Forecasting*, vol. 19, no. 4, pp. 231–234, 2000.
- [21] T. Gneiting, "Quantiles as optimal point forecasts," *International Journal of Forecasting*, vol. 27, no. 2, pp. 197–207, 2011.
- [22] L. Wu, M. Shahidehpour, and T. Li, "Stochastic security-constrained unit commitment," *Power Systems, IEEE Transactions on*, vol. 22, no. 2, pp. 800–811, 2007.
- [23] J. M. Morales and J. Perez-Ruiz, "Point estimate schemes to solve the probabilistic power flow," *Power Systems, IEEE Transactions on*, vol. 22, no. 4, pp. 1594–1601, 2007.
- [24] T. Niknam, A. Kavousi Fard, and A. Baziar, "Multi-objective stochastic distribution feeder reconfiguration problem considering hydrogen and thermal energy production by fuel cell power plants," *Energy*, vol. 42, no. 1, pp. 563–573, 2012.
- [25] 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.
- [26] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004.
- [27] G. Anastasiades and P. McSharry, "Quantile forecasting of wind power using variability indices," *Energies*, vol. 6, no. 2, pp. 662–695, 2013.
- [28] 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.

- [29] J. Tastu, “Short-term wind power forecasting: probabilistic and space-time aspects,” 2013.
- [30] P. Pinson, H. A. Nielsen, H. Madsen, and G. Kariniotakis, “Skill forecasting from ensemble predictions of wind power,” *Applied Energy*, vol. 86, no. 7-8, pp. 1326–1334, 2009.
- [31] 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.
- [32] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating renewables in electricity markets: Operational problems*. Springer Science & Business Media, 2013, vol. 205.
- [33] C. Yang and L. Xie, “A novel arx-based multi-scale spatio-temporal solar power forecast model,” in *2012 North American Power Symposium (NAPS)*, 2012, pp. 1–6.
- [34] J. Widén, “Correlations between large-scale solar and wind power in a future scenario for sweden,” *Sustainable Energy, IEEE Transactions on*, vol. 2, no. 2, pp. 177–184, 2011.
- [35] C. Yang, A. Thatte, L. Xie *et al.*, “Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation,” *Sustainable Energy, IEEE Transactions on*, vol. 6, no. 1, pp. 104–112, 2015.
- [36] R. J. Bessa, A. Trindade, and V. Miranda, “Spatial-temporal solar power forecasting for smart grids,” *Industrial Informatics, IEEE Transactions on*, vol. 11, no. 1, pp. 232–241, 2015.
- [37] V. Berdugo, C. Chaussin, L. Dubus, G. Hebrail, and V. Leboucher, “Analog method for collaborative very-short-term forecasting of power generation from photovoltaic systems,” *Next Generation Data Mining Summit: Ubiquitous Knowledge Discovery for Energy Management in Smart Grids and Intelligent Machine-to-Machine (M2M) Telematics*, 2011.
- [38] H. Louie, “Evaluation of bivariate archimedean and elliptical copulas to model wind power dependency structures,” *Wind Energy*, vol. 17, no. 2, pp. 225–240, 2014.
- [39] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A fast and accurate online sequential learning algorithm for feedforward networks,” *Neural Networks, IEEE Transactions on*, vol. 17, no. 6, pp. 1411–1423, 2006.
- [40] J. Han and M. Kamber, *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann, 2006.
- [41] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, “Photovoltaic and solar forecasting: state of the art,” *IEA PVPS, Task*, vol. 14, 2013.
- [42] M. MAKhyoun, “Predicting solar power production,” *Solar Electric Power Association*, 2014.
- [43] A. Yona, T. Senjyu, T. Funabashi, and C.-H. Kim, “Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead pv power output correction,” *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 2, pp. 527–533, 2013.

- [44] U. EPA, “Exposure factors handbook 2011 edition (final),” *Washington, DC*, 2011.
- [45] A. Mills, M. Ahlstrom, M. Brower, A. Ellis, R. George, T. Hoff, B. Kroposki, C. Lenox, N. Miller, J. Stein *et al.*, “Understanding variability and uncertainty of photovoltaics for integration with the electric power system,” Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), Tech. Rep., 2009.
- [46] W. Traunmüller and G. Steinmaurer, “Solar irradiance forecasting, benchmarking of different techniques and applications of energy meteorology,” in *Proceedings of the EuroSun 2010 conference*. Graz, Austria, 2010.
- [47] E. Lorenz, D. Heinemann, H. Wickramaratne, H. Beyer, and S. Bofinger, “Forecast of ensemble power production by grid-connected pv systems,” in *20th European PV Conference*. Milano, 2007, pp. 3–9.
- [48] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, “Irradiance forecasting for the power prediction of grid-connected photovoltaic systems,” *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 2, no. 1, pp. 2–10, 2009.
- [49] B. Ernst, F. Reyer, and J. Vanzetta, “Wind power and photovoltaic prediction tools for balancing and grid operation,” in *Integration of Wide-Scale Renewable Resources Into the Power Delivery System, 2009 CIGRE/IEEE PES Joint Symposium*. IEEE, 2009, pp. 1–9.
- [50] R. Perez, M. Beauharnois, K. Hemker, S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, and G. Van Knowe, “Evaluation of numerical weather prediction solar irradiance forecasts in the us,” in *preparation (to be presented at the 2011 ASES Annual Conference)*, 2011.
- [51] B. Kirby and M. Milligan, “Facilitating wind development: The importance of electric industry structure,” *The Electricity Journal*, vol. 21, no. 3, pp. 40–54, 2008.
- [52] T. E. Hoff and R. Perez, “Quantifying pv power output variability,” *Solar Energy*, vol. 84, no. 10, pp. 1782–1793, 2010.
- [53] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom, “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.
- [54] H. G. Beyer, J. Polo Martinez, M. Suri, J. L. Torres, E. Lorenz, S. C. Müller, C. Hoyer-Klick, and P. Ineichen, “D 1.1. 3 report on benchmarking of radiation products,” *MESOR Report*, 2009.
- [55] E. Lorenz, J. Remund, S. C. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J. A. Ruiz-Arias, V. L. Fanego, L. Ramirez, M. G. Romeo *et al.*, “Benchmarking of different approaches to forecast solar irradiance,” in *24th European photovoltaic solar energy conference*. Hamburg, Germany, 2009, pp. 1–10.
- [56] S. C. Müller and J. Remund, “Advances in radiation forecast based on regional weather models mm5 and wrf,” in *Proceedings of the 25th EUPVSEC Conference*, 2010, pp. 6–9.

- [57] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, and X. Zhou, "Comparative study of power forecasting methods for pv stations," in *Power System Technology (POWERCON), 2010 International Conference on*. IEEE, 2010, pp. 1–6.
- [58] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo, and J. Sumita, "Forecasting electric power generation in a photovoltaic power system for an energy network," *Electrical Engineering in Japan*, vol. 167, no. 4, pp. 16–23, 2009.
- [59] S. Pelland, G. Galanis, and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model," *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 3, pp. 284–296, 2013.
- [60] E. Lorenz, D. Heinemann, and C. Kurz, "Local and regional photovoltaic power prediction for large scale grid integration: Assessment of a new algorithm for snow detection," *Progress in Photovoltaics: Research and Applications*, vol. 20, no. 6, pp. 760–769, 2012.
- [61] H. T. Pedro and C. F. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, 2012.
- [62] A. Mellit, "Artificial intelligence technique for modelling and forecasting of solar radiation data: a review," *International Journal of Artificial intelligence and soft computing*, vol. 1, no. 1, pp. 52–76, 2008.
- [63] S. Jafarzadeh, M. S. Fadali, and C. Y. Evrenosoglu, "Solar power prediction using interval type-2 tsk modeling," *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 2, pp. 333–339, 2013.
- [64] L. Ciabattoni, G. Ippoliti, S. Longhi, M. Pirro, and M. Cavalletti, "Solar irradiation forecasting for pv systems by fully tuned minimal rbf neural networks," in *Neural Nets and Surroundings*. Springer, 2013, pp. 289–300.
- [65] A. Sözen, E. Arcaklioğlu, and M. Özalp, "Estimation of solar potential in turkey by artificial neural networks using meteorological and geographical data," *Energy Conversion and Management*, vol. 45, no. 18, pp. 3033–3052, 2004.
- [66] 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.
- [67] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
- [68] X. Liu, C. Gao, and P. Li, "A comparative analysis of support vector machines and extreme learning machines," *Neural Networks*, vol. 33, pp. 58–66, 2012.
- [69] Y. Chu, B. Urquhart, S. M. Gohari, H. T. Pedro, J. Kleissl, and C. F. Coimbra, "Short-term reforecasting of power output from a 48 mwe solar pv plant," *Solar Energy*, vol. 112, pp. 68–77, 2015.
- [70] E. Y. Bitar, R. Rajagopal, P. P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1225–1235, 2012.

- [71] A. Botterud, Z. Zhou, J. Wang, R. J. Bessa, H. Keko, J. Sumaili, and V. Miranda, "Wind power trading under uncertainty in Imp markets," *IEEE Transactions on power systems*, vol. 27, no. 2, pp. 894–903, 2012.
- [72] U. Linnet, "Tools supporting wind energy trade in deregulated markets," Ph.D. dissertation, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark, 2005.
- [73] S. Bofinger, A. Luig, and H. Beyer, "Qualification of wind power forecasts," in *2002 Global Windpower Conference, Paris*, vol. 2, no. 5, 2002.
- [74] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [75] M. Lange, "On the uncertainty of wind power predictions-analysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, vol. 127, no. 2, pp. 177–184, 2005.
- [76] H. Bludszuweit, J. A. Domínguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *Power Systems, IEEE Transactions on*, vol. 23, no. 3, pp. 983–991, 2008.
- [77] H. Bludszuweit, "Reduction of the uncertainty of wind power predictions using energy storage," Ph.D. dissertation, PhD Thesis, Universidad de Zaragoza, Spain, 2009.
- [78] 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.
- [79] 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.
- [80] 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.
- [81] 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.
- [82] 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.
- [83] 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.
- [84] 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.
- [85] J. K. Møller, H. A. Nielsen, and H. Madsen, "Time-adaptive quantile regression," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292–1303, 2008.

- [86] 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.
- [87] 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*. IEEE, 2011, pp. 1–8.
- [88] 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.
- [89] R. Bessa, J. Sumaili, V. Miranda, A. Botterud, J. Wang, and E. Constantinescu, "Time-adaptive kernel density forecast: a new method for wind power uncertainty modeling," in *Proceedings of the 17th power systems computation conference, Stockholm, Sweden*, 2010.
- [90] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Optimal prediction intervals of wind power generation," 2014.
- [91] C. Wan, Z. Xu, P. Pierre, 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.
- [92] T. Heskes, W. Wiegierinck, and H. Kappen, "Practical confidence and prediction intervals for prediction tasks," 1997.
- [93] 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.
- [94] 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.
- [95] 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.
- [96] P. Mathiesen, J. M. Brown, and J. Kleissl, "Geostrophic wind dependent probabilistic irradiance forecasts for coastal california," *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 2, pp. 510–518, 2013.
- [97] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *Power Systems, IEEE Transactions on*, vol. 24, no. 1, pp. 40–49, 2009.
- [98] J. Tastu, P. Pinson, and H. Madsen, "Space-time trajectories of wind power generation: Parameterized precision matrices under a gaussian copula approach," *Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension*, 2014.
- [99] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12–20, 2012.

-
- [100] 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.
- [101] C. A. Gueymard and S. M. Wilcox, "Assessment of spatial and temporal variability in the us solar resource from radiometric measurements and predictions from models using ground-based or satellite data," *Solar Energy*, vol. 85, no. 5, pp. 1068–1084, 2011.
- [102] D. Yang, C. Gu, Z. Dong, P. Jirutitijaroen, N. Chen, and W. M. Walsh, "Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging," *Renewable Energy*, vol. 60, pp. 235–245, 2013.
- [103] X. G. Agoua, R. Girard, and G. Kariniotakis, "Spatio-temporal models for photovoltaic power short-term forecasting," in *Solar Integration workshop 2015*, 2015.
- [104] Uq solar, the university of queensland, australia.
- [105] Global energy forecasting competition 2014 probabilistic solar power forecasting. [Online]. Available: <https://crowdanalytix.com>
- [106] J. Ma, Y. V. Makarov, C. Loutan, and Z. Xie, "Impact of wind and solar generation on the california iso's intra-hour balancing needs," in *Power and Energy Society General Meeting, 2011 IEEE*. IEEE, 2011, pp. 1–6.
- [107] P. Pinson, P. McSharry, and H. Madsen, "Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation," *Quarterly Journal of the Royal Meteorological Society*, vol. 136, no. 646, pp. 77–90, 2010.
- [108] 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.
- [109] 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.
- [110] P. Pinson and R. Hagedorn, "Verification of the ecmwf ensemble forecasts of wind speed against analyses and observations," *Meteorological Applications*, vol. 19, no. 4, pp. 484–500, 2012.
- [111] M. J. Reno, C. W. Hansen, and J. S. Stein, "Global horizontal irradiance clear sky models: Implementation and analysis," *SANDIA report SAND2012-2389*, 2012.
- [112] I. A. Walter, R. G. Allen, R. Elliott, M. Jensen, D. Itenfisu, B. Mecham, T. Howell, R. Snyder, P. Brown, S. Echings *et al.*, "Asces standardized reference evapotranspiration equation," in *Task Committee on Standardization of Reference Evapotranspiration of the Environmental and Water Resources Institute*, 2005.
- [113] F. Golestaneh and H. B. Gooi, "Batch and sequential forecast models for photovoltaic generation," in *Power & Energy Society General Meeting, 2015 IEEE*. IEEE, 2015, pp. 1–5.

- [114] V. Akhmatov, "Influence of wind direction on intense power fluctuations in large offshore windfarms in the north sea," *Wind Engineering*, vol. 31, no. 1, pp. 59–64, 2007.
- [115] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine, and C.-H. Kim, "Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system," in *Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on.* IEEE, 2007, pp. 1–6.
- [116] Solarfun, "SF 160-24-M datasheet," [http://www.energymatters.com.au/images/solarfun/SF160-24-M\(190W\).pdf](http://www.energymatters.com.au/images/solarfun/SF160-24-M(190W).pdf), [Online].
- [117] SUNTECH, "Suntech STP190S-24/Ad+ datasheet," <http://www.energymatters.com.au/images/suntech/STP190S24Adplus.pdf>, [Online].
- [118] SUNTECH, "Suntech STP200-18/ud datasheet," [https://www.ibc-solar.pt/fileadmin/content/homepage/produtos/Download/Datasheet_Suntech_Power_STP_200_205_210_18Ud_\(101029\)_EN.pdf](https://www.ibc-solar.pt/fileadmin/content/homepage/produtos/Download/Datasheet_Suntech_Power_STP_200_205_210_18Ud_(101029)_EN.pdf), [Online].
- [119] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.
- [120] M. Sklar, *Fonctions de répartition à n dimensions et leurs marges.* Université Paris 8, 1959.
- [121] R. B. Nelsen, *An introduction to copulas.* Springer Science & Business Media, 2013, vol. 139.
- [122] S. Demarta and A. J. McNeil, "The t copula and related copulas," *International statistical review*, vol. 73, no. 1, pp. 111–129, 2005.
- [123] G. Díaz, P. G. Casielles, and J. Coto, "Simulation of spatially correlated wind power in small geographic areassampling methods and evaluation," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 513–522, 2014.
- [124] G. Díaz, "A note on the multivariate archimedean dependence structure in small wind generation sites," *Wind Energy*, vol. 17, no. 8, pp. 1287–1295, 2014.
- [125] F. Golestaneh and H. Gooi, "Batch and sequential forecast models for photovoltaic generationl," in *IEEE PES 2015 General Meeting, Denver, Colorado, USA*, 2015.
- [126] Y. Sakamoto, M. Ishiguro, and G. Kitagawa, "Akaike information criterion statistics," *Dordrecht, The Netherlands: D. Reidel*, 1986.
- [127] X. Liu, "Economic load dispatch constrained by wind power availability: A wait-and-see approach," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 347–355, 2010.
- [128] C. Liu, J. Wang, A. Botterud, Y. Zhou, and A. Vyas, "Assessment of impacts of phev charging patterns on wind-thermal scheduling by stochastic unit commitment," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 675–683, 2012.

- [129] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda, "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of illinois," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 250–261, 2013.
- [130] R. J. Bessa, M. A. Matos, I. C. Costa, L. Bremermann, I. G. Franchin, R. Pestana, N. Machado, H.-P. Waldl, and C. Wichmann, "Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 827–836, 2012.
- [131] M. A. Ortega-Vazquez and D. S. Kirschen, "Assessing the impact of wind power generation on operating costs," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 295–301, 2010.
- [132] J. Dowell, "Probabilistic forecasting of wind power generation."
- [133] J. Tastu, P. Pinson, P.-J. Trombe, and H. Madsen, "Probabilistic forecasts of wind power generation accounting for geographically dispersed information," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 480–489, 2014.
- [134] L. Hu, "Dependence patterns across financial markets: a mixed copula approach," *Applied financial economics*, vol. 16, no. 10, pp. 717–729, 2006.
- [135] A. J. McNeil, "Sampling nested archimedean copulas," *Journal of Statistical Computation and Simulation*, vol. 78, no. 6, pp. 567–581, 2008.
- [136] D. Kurowicka, *Dependence modeling: vine copula handbook*. World Scientific, 2011.
- [137] U. Schepsmeier, "A goodness-of-fit test for regular vine copula models," *arXiv preprint arXiv:1306.0818*, 2013.
- [138] W. Wu, K. Wang, B. Han, G. Li, X. Jiang, and M. L. Crow, "A versatile probability model of photovoltaic generation using pair copula construction," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1337–1345, 2015.
- [139] C. Czado, "Pair-copula constructions of multivariate copulas," in *Copula theory and its applications*. Springer, 2010, pp. 93–109.
- [140] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [141] K. A. Clarke, "A simple distribution-free test for nonnested model selection," *Political Analysis*, vol. 15, no. 3, pp. 347–363, 2007.
- [142] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: Journal of the Econometric Society*, pp. 307–333, 1989.
- [143] H. Zhang and P. Li, "Chance constrained programming for optimal power flow under uncertainty," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2417–2424, 2011.
- [144] A. A. Sousa, G. L. Torres, and C. A. Canizares, "Robust optimal power flow solution using trust region and interior-point methods," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 487–499, 2011.

-
- [145] R. A. Jabr, "Adjustable robust opf with renewable energy sources," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4742–4751, 2013.
- [146] L. Wu, M. Shahidehpour, and Z. Li, "Comparison of scenario-based and interval optimization approaches to stochastic scuc," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 913–921, 2012.
- [147] R. J. Bessa, "From marginal to simultaneous prediction intervals of wind power," in *Intelligent System Application to Power Systems (ISAP), 2015 18th International Conference on*. IEEE, 2015, pp. 1–6.
- [148] D. Kolsrud, "Time-simultaneous prediction band for a time series," *Journal of Forecasting*, vol. 26, no. 3, pp. 171–188, 2007.
- [149] J. S.-H. Li and W.-S. Chan, "Simultaneous prediction intervals: An application to forecasting us and canadian mortality," 2011.
- [150] P. Li, X. Guan, J. Wu, and X. Zhou, "Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1594–1605, 2015.
- [151] Y. Guan and J. Wang, "Uncertainty sets for robust unit commitment," *IEEE Transactions on Power Systems*, vol. 3, no. 29, pp. 1439–1440, 2014.
- [152] P. Pinson and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with markov-switching autoregressive models," *Journal of Forecasting*, vol. 31, no. 4, pp. 281–313, 2012.
- [153] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima, and J. Kleissl, "Solar irradiance forecasting using a ground-based sky imager developed at uc san diego," *Solar Energy*, vol. 103, pp. 502–524, 2014.
- [154] B. Urquhart, B. Kurtz, E. Dahlin, M. Ghonima, J. Shields, and J. Kleissl, "Development of a sky imaging system for short-term solar power forecasting." *Atmospheric Measurement Techniques*, vol. 8, no. 2, 2015.
- [155] C. W. Chow, S. Belongie, and J. Kleissl, "Cloud motion and stability estimation for intra-hour solar forecasting," *Solar Energy*, vol. 115, pp. 645–655, 2015.
- [156] W. Härdle, *Nonparametric and semiparametric models*. Springer, 2004.
- [157] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*. Springer, 2009, vol. 2, no. 1.

Appendix A

Kernel Density Estimate (KDE)

KDE is a nonparametric approach to estimate the PDF of a random variable by using a finite data samples. Let us denote $y_1, y_2, \dots, y_N$ as N independently and identical samples extracted from the unknown PDF $f(Y)$ of the random variable Y . Using KDE, a kernel estimation of $f(Y)$, denoted by $\hat{f}_h(y)$, can be found by [156]

$$\hat{f}_h(y) = \sum_{i=1}^N \frac{1}{Nh} K\left(\frac{y - y_i}{h}\right) = \frac{1}{N} \sum_{i=1}^N K_h(y - y_i) \quad (\text{A.1})$$

where $K_h(\cdot)$ is the scaled kernel; h is called smoothing parameter or kernel bandwidth; and $K_h(y) = \frac{1}{h} K\left(\frac{y}{h}\right)$. The area under KDE is equal to one and the area under each of the rescaled kernels can be calculated by [156]

$$\int \frac{1}{Nh} K\left(\frac{y - y_i}{h}\right) dx = \frac{1}{Nh} \int K(u) h du = \frac{1}{Nh} h \int K(u) du = \frac{1}{n} \quad (\text{A.2})$$

The KDE and rescaled kernel functions for an artificial dataset containing 10 samples are depicted in Fig. A.1. Rescaled kernels are usually called “bumps” and each of them is centered on one of the observations. By vertically summing over these little bumps, KDE can be obtained. It is self-evident that different values of h result in different $\hat{f}_h(y)$. It can be mathematically proven that increasing h over-smooths the KDE and increases the bias in estimation. In contrast, decreasing it makes the estimated density spiky and increases the variance. Therefore, bias-variance trade-off is crucial in KDE to choose the optimal h . In this study, cross validation is employed to find the optimal h for the datasets in hand.

Figure A.1: Kernel density estimate for an artificial dataset

Appendix B

Cross Validation

Cross validation (CV) is one of the widely used approaches for tuning and validating statistical learning methods [157]. It is also an efficient method to estimate the performance of a given machine learning model when applied to an independent dataset. Using the CV the expected extra sample error made by $\hat{f}_{t+k|t}$ when applied to an independent data sample can be estimated.

B.1 K -Fold Cross Validation

In the K -fold CV, a subset of the available dataset is used to fit a model and the remaining data is set aside for testing the model. To do this, the dataset is randomly split into K folds with almost the same size. Fig. B.1 shows a typical 5-fold cross validation partitioning.

Figure B.1: A typical diagram for 5-fold cross validation

Firstly, the initial validation fold (e.g j^{th} fold) is determined and excluded from the dataset. The remaining samples are used to train the model. It is followed by testing the model on the validation fold. The procedure is repeated for $j = 1, 2, \dots, K$ and each time a different fold

is treated as the validation fold. Considering a dataset of N samples and using K -fold CV, the estimated error (ε_{cv}) in predication of the expected value of a random variable Y_t can be achieved by [157]

$$\varepsilon_{cv} = \sum_{j=1}^K \sum_{i=1}^{N_{\varsigma_j}} \frac{1}{N_{\varsigma_j}} L(y_{t_i}, \hat{f}_{t_i+k|t_i}^{-\varsigma_j}) \quad (\text{B.1})$$

where N_{ς_j} is the number of data items in the fold ς_j ; $\hat{f}_{t_i+k|t_i}^{-\varsigma_j}$ are the forecasts using the model trained by the dataset when fold ς_j is excluded; y_{t_i} are the true values for the samples in the validation set and L is the error function.

Considering $rRMSE$ as the error function, the estimated $rRMSE$ when the trained model is tested on an independent dataset can be obtained by averaging the given $rRMSE_j$ ($j = 1, \dots, K$) as

$$rRMSE_{cv} = \frac{1}{K} \sum_{j=1}^K rRMSE_j \quad (\text{B.2})$$

Apart from K -fold CV, the validation set and leave-one-out CV are two other CV approaches and both of them can be considered as a special case of K -fold CV. For the former one, $K = 2$ while the later one takes $K = N$. Among different CV methods, K -fold CV provides the best trade-off between bias and variance therefore, it is chosen in this project for tuning and validation of the forecast models.

Appendix C

Clear Sky Solar Irradiance

Clear sky solar radiation is defined as the amount of terrestrial solar radiation that would be received at the solar irradiance sensor under a cloudless sky. The available clear sky models rely on solar elevation angle, site altitude and various atmospheric variables. As the accuracy of models increases, the number of required variables increases. A review on different clear sky models are presented in [111]. In this study, by considering the limited atmospheric variables available for the datasets in hand, the model formulated in [112] is used.

$$R_{so} = (0.75 + 2 \times (10^{(-5)}) \times z) \times R_a \quad (C.1)$$

with R_a as extraterrestrial radiation [$MJm^{-2}h^{-1}$], z as station elevation above sea level [m] and R_{so} as the clear-sky solar radiation [$MJm^{-2}h^{-1}$].

Extraterrestrial radiation is the short-wave radiation outside of the earth's atmosphere and can be estimated with a good accuracy based on solar declination, the day of the year, time of day and sites' latitude as

$$R_a = \left(\frac{24}{\pi}\right) G_s d_r ((\omega_2 - \omega_1) \sin(\phi) \sin(\delta) + \cos(\phi) \cos(\delta) (\sin(\omega_2) - \sin(\omega_1))) \quad (C.2)$$

with ϕ as latitude of the site, ω_2 and ω_1 as solar time angles at the beginning and end of the period, respectively and δ as solar declination. G_s is the solar constant $4.92 [m^{-2}h^{-1}]$ and d_r

is defined as inverse relative distance factor for the earth-sun and is formulated by

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi J}{365}\right) \quad (\text{C.3})$$

with J as the day of year [1-365]. Solar declination can be calculated by

$$\delta = 0.409 \sin\left(\frac{2\pi J}{365} - 1.39\right) \quad (\text{C.4})$$

Note that all the angles in the above formulation are in radians. ω_2 and ω_1 are defined in (C.5) and (C.6).

$$\omega_1 = \omega - \left(\frac{\pi t_1}{24}\right) \quad (\text{C.5})$$

$$\omega_2 = \omega + \left(\frac{\pi t_1}{24}\right) \quad (\text{C.6})$$

with ω as the solar time angle at the midpoint of the period and t_1 as the length of the period [112].