



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a Doctor of Philosophy at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Data-Driven Modelling, Forecasting and Uncertainty Analysis of Disaggregated Demands and Wind Farm Power Outputs

Mingzhe Zou



Doctor of Philosophy

THE UNIVERSITY OF EDINBURGH

2022

Abstract

Correct analysis of modern power supply systems requires to evaluate much wider ranges of uncertainties introduced by the implementation of new technologies on both supply and demand sides. On the supply side, these uncertainties are due to the increased contributions of renewable generation sources (e.g., wind and PV), whose stochastic output variations are difficult to predict and control, as well as due to the significant changes in system operating conditions, coming from the implementation of various control and balancing actions, increased automation and switching functionalities, and frequent network reconfiguration. On the demand side, these uncertainties are due to the installation of new types of loads, featuring strong spatio-temporal variations of demands (e.g., EV charging), as well as due to the deployment of different demand-side management schemes.

Modern power supply systems are also characterised by much higher availability of measurements and recordings, coming from a number of recently deployed advanced monitoring, data acquisition and control systems, and providing valuable information on system operating and loading conditions, state and status of network components and details on various system events, transients and disturbances. Although the processing of large amounts of measured data brings its own challenges (e.g., data quality, performance, and incorporation of domain knowledge), these data open new opportunities for a more accurate and comprehensive evaluation of the overall system performance, which, however, require new data-driven analytical approaches and modelling tools.

This PhD research is aimed at developing and evaluating novel and improved data-driven methodologies for modelling renewable generation and demand, in general, and for assessing the corresponding uncertainties and forecasting, in particular. The research and methods developed in this thesis use actual field measurements of several onshore and offshore wind farms, as well as measured active and reactive power demands at several low voltage (LV) individual household levels, up to the demands at medium voltage (MV) substation level. The models are specifically built to be implemented for power system analysis and are actually used by a number of

researchers and PhD students in Edinburgh and elsewhere (e.g., collaborations with colleagues from Italy and Croatia), which is discussed and illustrated in the thesis through the selected study cases taken from this joint research efforts.

After literature review and discussion of basic concepts and definitions, the first part of the thesis presents data-driven analysis, modelling, uncertainty evaluation and forecasting of (predominantly residential) demands and load profiles at LV and MV levels. The analysis includes both aggregation and disaggregation of measured demands, where the latter is considered in the context of identifying demand-manageable loads (e.g., heating). For that purpose, periodical changes in demands, e.g., half-daily, daily, weekly, seasonal and annual, are represented with Fourier/frequency components and correlated with the corresponding exploratory meteorological variables (e.g., temperature, solar irradiance), allowing to select the combination of components maximising the positive or negative correlations as an additional predictor variable. Convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) are then used to represent dependencies among multiple dimensions and to output the estimated disaggregated time series of specific load types (with Bayesian optimisation applied to select appropriate CNN-BiLSTM hyperparameters). In terms of load forecasting, both tree-based and neural network-based models are analysed and compared for the day-ahead and week-ahead forecasting of demands at MV substation level, which are also correlated with meteorological data. Importantly, the presented load forecasting methodologies allow, for the first time, to forecast both total/aggregate demands and corresponding disaggregated demands of specific load types.

In terms of the supply side analysis, the thesis presents data-driven evaluation, modelling, uncertainty evaluation and forecasting of wind-based electricity generation systems. The available measurements from both the individual wind turbines (WTs) and the whole wind farms (WFs) are used to formulate simple yet accurate operational models of WTs and WFs. First, available measurements are preprocessed, to remove outliers, as otherwise obtained WT/WF models may be biased, or even inaccurate. A novel simulation-based approach that builds on a procedure recommended in a standard is presented for processing all outliers due to applied averaging window

(typically 10 minutes) and WT hysteresis effects (around the cut-in and cut-out wind speeds). Afterwards, the importance of distinguishing between WT-level and WF-level analysis is discussed and a new six-parameter power curve model is introduced for accurate modelling of both cut-in and cut-out regions and for taking into account operating regimes of a WF (WTs in normal/curtailed operation, or outage/fault).

The modelling framework in the thesis starts with deterministic models (e.g., CNN-BiLSTM and power curve models) and is then extended to include probabilistic models, building on the Bayesian inference and Copula theory. In that context, the thesis presents a set of innovative data-driven WT and WF probabilistic models, which can accurately model cross-correlations between the WT/WF power output (P_{out}), wind speed (WS), air density (AD) and wind direction (WD). Vine Copula and Gaussian mixture Copula model (GMCM) are combined, for the first time, to evaluate the uncertainty of P_{out} values, conditioning on other explanatory variables (which may be either deterministic, or also uncertain). In terms of probabilistic wind energy forecasting, Bayesian CNN-BiLSTM model is used to analyse and efficiently handle high dimensionality of both input meteorological variables (WS, AD and WD) and additional uncertainties due to WF operating regimes. The presented results demonstrate that the developed Vine-GMCM and operational WF model can accurately integrate and effectively correlate all propagated uncertainties, ultimately resulting in much higher confidence levels of the forecasted WF power outputs than in the existing literature.

Lay Summary

The past several years have seen rapid development in modern power systems. Due to the growth of the economy and people's concern about energy and the environment, on the demand side, there has been a fast penetration of new types of loads (e.g., electric vehicle charging) and development of different demand-side management schemes. On the supply side, there are increasing contributions from renewable generation sources (e.g., wind and photovoltaic), which also implement new control actions. For example, a wind turbine can adjust the blade angle and the rotation to control its power output. And a wind farm may reduce its generation by curtailing or shutting down some wind turbines. Those changes have greatly improved efficiency, asset utilisation, environmental impact and techno-economic system performance.

However, because the new types of load feature often much wider ranges of operating conditions, and renewable energy sources, such as wind and photovoltaic generation systems, have inherent variability thus are difficult to accurately predict and fully control the outputs, there is also a significant increase in uncertainty of the power system. In addition, as a result of the implementation of new technologies on both the demand and supply side, there are large amounts of measurement data collected from various widely installed sensors, and it is another challenge to efficiently process them and get insights into the data.

This thesis aims to develop novel data-driven models to more efficiently analyse the actual demand and supply measurements, to better understand the correlation of different variables and uncertainty associated with the process, and to reduce the ranges of uncertainty for more confident forecasting. The first part of the thesis researches on the demand side, where years of measurements of three aggregation levels are analysed: household, substation and city. To evaluate the variation of electricity load, a probability model is introduced, proving the existence of variation from two aspects. Then it also investigates several promising machine learning models and compares their performance on load forecasting, which helps the selection of a more proper model that has higher accuracy and is extensible and scalable. To identify demand-manageable loads, such as heating or cooling loads, and disaggregate them

from the total load, innovative frequency-domain modelling and correlation analysis method is proposed, which further improves the performance of the sophisticated load disaggregation model. And since there is always some uncertainty of model hyperparameter settings, and improper selection of values may lead to worse performance, an optimisation method is applied and guides the setting of model hyperparameters. The hyperparameters refer to the architecture/structure of the model and its training configurations. For example, to build and train a multiple-layer neural network model, the hyperparameters include the number of layers and how much the optimiser should update the model parameters in response to the estimated errors (i.e., the learning rate), etc.

The second part of the thesis focuses on supply side research, where years of measurements from multiple wind turbines and wind farms are investigated and analysed. Since the outlier and abnormal recordings in the measurements have strong negative impacts on applications related to data-driven models, which can cause lower accuracy and even errors, the thesis firstly presents a comprehensive and detailed review of approaches for the detection and treatment of outliers in processing wind turbine and wind farm measurements. Afterwards, it proposes a generic framework to detect and handle the outliers in the measurements, and a wind turbine simulation method is presented, which is based on and as an extension of the current standard that specifies the procedure of measuring the power performance of electricity producing wind turbines. Then, it also develops a novel power curve model that is capable of modelling the high wind speed region in which cut-out etc. of individual wind turbines may occur, whilst most other current models neglect. To correctly model the wind farm, the relationship between the wind turbine and wind farm is analysed, which leads to a simple but powerful operational/equivalent wind farm model. To capture the uncertainty of wind energy and improve decision making, this thesis also investigates and proposes several novel probability models from simple to complex. The proposed high-dimensional and multi-stage model can handle the uncertainty from input meteorological variables (wind speed, air density and wind direction), the power output, and the wind farm operating regime. And it can accurately integrate and effectively correlate all propagated uncertainties, ultimately resulting in much higher

confidence levels of the forecasted wind farm power outputs than in the existing literature

Acknowledgement

Firstly, I would like to thank my supervisor, Dr Sasa Djokic, for his constant guidance and support during the PhD study, especially for his encouragement and patience, which kept me motivated to make the most out of my PhD. I would like to acknowledge the internal grant from School of Engineering, the University of Edinburgh (grant name: School of Engineering PhD Scholarship), which funds tuition fees, stipend and research costs for this PhD research.

Thanks are also extended to my assistant supervisor, Prof. Gareth Harrison, for his instructions and help during hard times. Also, thanks to Prof. Roberto Langella and Prof. Igor Kuzle, for their constructive advice in wind energy research.

I am also grateful to Croatian WF Operators for providing a large part of the wind farm data in this thesis, and I am also indebted to China Electric Power Research Institute for setting up my productive internship and providing part of load data in this thesis.

Many thanks to my colleagues at the Institute for Energy Systems and other universities, who have been a real pleasure to work with. Thanks to Vincenzo Di Giorgio, Ninoslav Holjevac and Josip Đaković for their efforts in collaboration.

Thank you to all my friends, and special mention goes to Duo Fang and Zhihong Zeng for their continuous support during my difficult times from undergraduate to PhD, to Shuning Xu for her kind, helpful and inspiring suggestions when I was anxious and depressed seeking jobs, to both of my roommates (Jiacheng Gu and Yu Chen) for being like a “family” away from home and making my life in Edinburgh as easy and relaxed as possible.

Finally, a very special acknowledgement goes to my family, for their endless love and all of the encouragement, advice and support. I am grateful to my parents (Guoliang Zou and Aiwu Zhu), for their continued belief in me.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own, except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Mingzhe Zou

Contents

Abstract	I
Lay Summary	IV
Acknowledgement	VII
Declaration	VIII
Contents	IX
List of Figures	XV
List of Tables	XXII
Acronyms and Abbreviations	XXVI
List of Symbols	XXXI
Chapter 1 Introduction	1
1.1 Introduction	1
1.2 The Need for Improved Modelling, Forecasting and Uncertainty Analysis on Electricity Demand and Supply Sides	3
1.3 Research Objectives and Main Contributions of the Thesis	6
1.4 Thesis Structure	10
Chapter 2 Literature Review	12
2.1 Introduction	12
2.2 Overview of Basic Modelling and Forecasting Approaches	12
2.2.1 Machine Learning, Neural Network, and Deep Learning	13
Multilayer perceptron	13
Convolutional Neural Network	14
Recurrent Neural Network	15
Long Short-Term Memory	16
2.2.2 Copula Methodology	18
2.3 Commonly Applied Error Metrics	21
2.3.1 Deterministic Error Metrics	21
2.3.2 Probabilistic Error Metrics	22
2.4 Load Modelling and Forecasting Review	23
2.4.1 Load Forecasting	23
2.4.2 Load Disaggregation	24
2.5 Wind Turbine and Wind Farm Modelling and Forecasting Review	26

2.5.1	Overview of Wind Turbine Topologies and Control	27
2.5.2	Wind Turbine and Wind Farm Modelling.....	28
2.5.3	Wind Turbine and Wind Farm Forecasting.....	30
2.6	Review of Classification of Uncertainty	31
2.7	Discussions and Conclusions	34
Chapter 3 Weather Based Day-Ahead and Week-Ahead Load Modelling and Forecasting.....		35
3.1	Introduction	35
3.2	Description of Available Datasets.....	35
	Scotland Medium Voltage (MV) Substation Dataset.....	35
	China City Dataset	36
	The Almanac of Minutely Power Dataset (Version 2) [115].....	36
3.3	Time Series and Variation Analysis for Load Profiles	37
3.3.1	Sliding Window Approach.....	37
3.3.2	Second-Order Markov Chain Analysis	38
3.4	Day-Ahead Load Forecasting.....	40
3.4.1	Data Pre-processing: Dataset Splitting, One Hot Encoder, and Data Normalisation.....	41
3.4.2	Instantiation of MLP and SB-LSTM Model	41
3.4.3	Benchmarking NN-Based Model with Tree-Based Model: Gradient Boosting Regression Trees.....	43
3.4.4	Example Case Study 3.1: Comparison of Day-Ahead Load Forecasting Results by Three Considered Models	44
3.5	Week-Ahead Load Forecasting.....	49
3.5.1	Example Case Study 3.2: Comparison of Day-Ahead and Week-Ahead Load Forecasting Results by SB-LSTM	50
3.6	Forecasting of both the Total and Disaggregated Load	52
3.6.1	Example Case Study 3.3: Multi-dimensional Day-Ahead Load Forecasting Results by SB-LSTM	53
3.7	Discussions and Conclusions	55

Chapter 4 Heating and Lighting Load Disaggregation Using Frequency Components and Convolutional Bidirectional Long Short-Term Memory

Method 57

4.1 Introduction	57
4.2 Selection of Target Frequency Components for Fourier Series Decomposition and Correlational Analysis	59
4.2.1 Description of Available Datasets	60
4.2.2 Disaggregation through Targeted Single-Components Fourier Series Decomposition	61
4.2.3 Extraction of the Strongly Correlated Components	76
4.3 Neural Network-Based Load Disaggregation Model	78
4.3.1 Dataset Splitting and Model Input and Output Variables	79
4.3.2 Model Hyperparameter Tuning: Bayesian Optimisation	81
4.3.3 Load Disaggregation Results and The Benefits of Using Fourier Series Regression	84
4.4 Discussions and Conclusions	88

Chapter 5 A Review of Approaches for the Detection and Treatment of Outliers in Processing Wind Turbine and Wind Farm Measurements 90

5.1 Introduction	90
5.2 Wind Turbine and Wind Farm Measurements	93
5.3 Definitions, Characteristics and Causes of Outliers	97
5.3.1 Definitions of Outliers	98
5.3.2 Characteristics and Causes of Outliers in Available WT and WF Measurement Data	101
5.3.3 Data Rejection Requirements in IEC Standard 61400-12-1 [12]	108
5.3.4 Summary of Common Causes of Outliers in WT and WF Measurements in the Existing Literature	109
Outliers Related to Data Acquisition, Data Transfer and Data Management Errors	111
Outliers Determined by the Operational Logs and Event-Logging Systems	112
Outliers Due to Applied Period of Averaging (Averaging Window)	113
Outliers Due to Cut-Out Effects	114

5.4 Detection of Outliers in Processing Wind Turbine and Wind Farm Measurements	115
5.4.1 Statistical Methods for Outlier Detection.....	117
Wind Turbine Measurements	117
Wind Farm Measurements	122
5.4.2 Physical Constraint-Based Methods for Outlier Detection	123
5.4.3 Outlier Detection by Combinations of Statistical and Physical Constraints-Based Methods	123
5.4.4 Robust Statistical Models for Reducing Outlier Effects	125
5.5 Approaches for Testing Success of Outlier Detection and Removal/Treatment Procedures	125
5.6 Discussions and Conclusions	129
Chapter 6 Detection of Outliers in Wind Farm Measurements and Operational Wind Farm Model.....	132
6.1 Introduction	132
6.2 Description of Available Datasets.....	133
Scotland Wind Farm Dataset	133
Zelengrad Wind Farm Dataset	134
6.3 Detection of WT-Level Outliers	137
6.3.1 Identification of CAT-I and II Outliers	140
6.3.2 Identification of CAT-III and CAT-IV Outliers	140
6.3.3 WT-Level Outlier Detection Results.....	144
6.4 Six-Parameter Power Curve Model.....	146
6.5 Wind Farm Operating Regimes and Equivalent Wind Farm Model.....	147
6.5.1 WF Operating Regime	148
6.5.2 Equivalent Wind Farm Model.....	149
6.6 Single Sensor WF Measurements	151
6.7 Discussions and Conclusions	155
Chapter 7 Wind Farm 2D Probabilistic Modelling and Conditional Distribution Fittings 156	
7.1 Introduction	156
7.2 Probabilistic Models for Wind Speed	156

7.2.1	Two-Parameter Weibull Distribution.....	157
7.2.2	Three-Parameter Weibull Distribution.....	158
7.2.3	Generalised Normal Distribution	160
7.2.4	Wind Speed Fitting Results.....	161
7.3	Probabilistic Models for Wind Direction	163
7.3.1	Von Mises Distribution and Mixture of Distributions	164
7.3.2	Wind Direction Fitting Results	165
7.4	Probabilistic Models for Power Output.....	168
7.4.1	Power Output Fitting Results	168
7.5	Discussions and Conclusions	169
Chapter 8 3D Vine Copula Model for Evaluation of Wind Turbine Power		
Outputs with and without Uncertainties in Input Wind Speed and Wind		
Direction Data..... 171		
8.1	Introduction	171
8.2	WT Power Curve Models, Importance of Including WD and Empirical P_{out} Uncertainty	172
8.2.1	Manufacturer and Operational WT Power Curves.....	173
8.2.2	Empirical WT P_{out} Uncertainty	175
8.3	Copula-Based Model.....	175
8.3.1	P_{out} -WS Model: GMCM Framework	176
8.3.2	P_{out} -WS-WD Model: Vine-GMCM Framework	178
8.4	Markov Chain Model Forecasted Input WS and WD Data with Uncertainties	
	181	
8.4.1	Discretisation Process	181
8.4.2	Forecasted WS Data	181
8.4.3	Forecasted WS-WD Data	182
8.5	Case Study and Obtained Results.....	182
8.5.1	Metrics for Comparison of Uncertainty Ranges	183
8.5.2	Results of Copula-Based Models without Uncertainties in WS and WD	
	183	
8.5.3	Results of Copula-Based Models with Uncertainties in WS and WD ..	186
8.6	Discussions and Conclusions	187

Chapter 9 Bayesian CNN-BiLSTM and Vine-GMCM Based Probabilistic Forecasting of Hour-Ahead Wind Farm Power Outputs.....	190
9.1 Introduction	190
9.2 Description of Modelled WFs, Available Measurements and Used Evaluation Metrics.....	191
9.3 Bayesian CNN-BiLSTM for Forecasting Weather Variables and Wind Farm Operating Regimes.....	194
9.3.1 Deterministic CNN-BiLSTM Model	195
9.3.2 Handling of Uncertainties with Bayesian Neural Network.....	195
9.3.3 Bayesian CNN-BiLSTM Model.....	198
9.3.4 Example Case Study 9.1: Forecasting of Weather Variables.....	198
9.3.5 Example Case Study 9.2: Forecasting of WF Operating Regimes	201
9.4 Vine-GMCM for Correlation-Based Modelling of WF Power Outputs	203
9.4.1 Proposed 4-D Correlation Model	203
9.4.2 Example Case Study 9.3: Modelling of WF Power Output Correlations	204
9.5 Combining Bayesian CNN-BiLSTM and Vine GMCM for Wind Farm Output Forecasting	207
9.5.1 Example Case Study 9.4: Incorporating WF ORs in Forecasting.....	209
9.5.2 Model Comparison and Benchmarking.....	211
9.5.3 Benchmarking Results.....	214
9.6 Extended Example Case Study 9.5: Day-Ahead Forecasting with NWP	226
9.7 Discussions and Conclusions	229
Chapter 10 Conclusions and Future Work.....	231
10.1 Thesis Summary	231
10.2 Research Implications	233
10.3 Limitations of the Research and Future Work	235
Bibliography	240

List of Figures

Fig. 2.1. Typical MLP architecture.	14
Fig. 2.2. Typical RNN architecture.	15
Fig. 2.3. A general LSTM cell architecture (solid arrows represent weighted vectors, while dot arrows represent unweighted vectors).	16
Fig. 2.4. A bidirectional long short-term memory layer: (f) - “forward” and (b) - “backward”	17
Fig. 2.5. SB-LSTM architecture: (f) - “forward” and (b) - “backward”.	18
Fig. 2.6. D-Vine structure of three variables.	20
Fig. 2.7. C-Vine structure of three variables.	20
Fig. 3.1. Average daily load profiles of different types of days (Scotland MV substation dataset).	37
Fig. 3.2. Evaluation of demand variation with sliding window approach (Scotland MV substation dataset, Tuesday, 2 January 2007).	38
Fig. 3.3. Demand variation evaluation with the second-order MC model (Scotland MV substation dataset, Tuesday, 2 January 2007).	40
Fig. 3.4. The architecture of SB-LSTM model.	43
Fig. 3.5. Examples of day-ahead load forecasting in the China city dataset: (a) The minimum demand week. (b) The average demand week. (c) The maximum demand week.	46
Fig. 3.6. Examples of day-ahead load forecasting in the Scotland MV substation dataset: (a) The minimum demand week. (b) The average demand week. (c) The maximum demand week.	47
Fig. 3.7. Comparison of day-ahead and week-ahead load forecasting results for the week of minimum demand.	50
Fig. 3.8. Comparison of day-ahead and week-ahead load forecasting results for the week of average demand.	50
Fig. 3.9. Comparison of day-ahead and week-ahead load forecasting results for the week of maximum demand.	51
Fig. 3.10. The comparison of day-ahead forecasting of total demand and heating demand for the week of minimum total demand.	53

Fig. 3.11. The comparison of day-ahead forecasting of total demand and heating demand for the week of average total demand.....	54
Fig. 3.12. The comparison of day-ahead forecasting of total demand and heating demand for the week of maximum total demand.....	54
Fig. 4.1. A block diagram illustration of the proposed load disaggregation method and overall analysis presented in Chapter 4.....	59
Fig. 4.2. Window lengths for Fourier decomposition into seven predefined considered frequency components.	64
Fig. 4.3. Components of the proposed Fourier series regression on AMPds2 (Monday, 24 March 2014) for load and temperature recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.....	67
Fig. 4.4. Results of the proposed Fourier series regression on AMPds2 (Monday, 24 March 2014) for load and temperature recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of temperature frequency components, (d) phase angles of temperature frequency components.	68
Fig. 4.5. Reconstructions using the proposed Fourier series regression on AMPds2 (Monday, 24 March 2014) for load and temperature recordings: (a) reconstruction of non-disaggregated demand and temperature daily profiles using considered frequency components; (b) reconstruction of non-disaggregated demand and temperature daily profiles using conventional Fourier transform.....	68
Fig. 4.6. Components of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and temperature recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.	69
Fig. 4.7. Results of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and temperature recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of temperature frequency components, (d) phase angles of temperature frequency components.	70

Fig. 4.8. Reconstructions using the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and temperature recordings: (a) reconstruction of non-disaggregated demand and temperature daily profiles using considered frequency components; (b) reconstruction of non-disaggregated demand and temperature daily profiles using conventional Fourier transform.	70
Fig. 4.9. Components of the proposed Fourier series regression on UK-DALE (Monday, 11 January 2016) for load and solar irradiance recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.	71
Fig. 4.10. Results of the proposed Fourier series regression on UK-DALE (Monday, 11 January 2016) for load and solar irradiance recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of solar irradiance frequency components, (d) phase angles of solar irradiance frequency components.	72
Fig. 4.11. Reconstructions using the proposed Fourier series regression on UK-DALE (Monday, 11 January 2016) for load and solar irradiance recordings: (a) reconstruction of non-disaggregated demand and solar irradiance daily profiles using considered frequency components; (b) reconstruction of non-disaggregated demand and solar irradiance daily profiles using conventional Fourier transform.	72
Fig. 4.12. Components of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and solar irradiance recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.	73
Fig. 4.13. Results of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and solar irradiance recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of solar irradiance frequency components, (d) phase angles of solar irradiance frequency components.	74
Fig. 4.14. Reconstructions using the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and solar irradiance recordings: (a) reconstruction of non-disaggregated demand and solar irradiance daily profiles using	

considered frequency components; (b) reconstruction of non-disaggregated demand and solar irradiance daily profiles using conventional Fourier transform.	74
Fig. 4.15. Combinations of more negatively correlated frequency components for demand and temperature: (a) AMPds2 (Monday, 24 March 2014). (b) MV level dataset (Monday, 6 February 2012).	78
Fig. 4.16. Combinations of more negatively correlated frequency components for demand and solar irradiance: (a) UK-DALE (Monday, 11 January 2016). (b) MV level dataset (Monday, 6 February 2012).	78
Fig. 4.17. Proposed CNN-BiLSTM architecture and hyperparameters.	80
Fig. 4.18. Example load disaggregation results. (a) AMPds2, heating load disaggregation, Monday, 24 March 2014. (b) UK-DALE, lighting load disaggregation, Monday, 11 January 2016.	85
Fig. 4.19. Example load disaggregation results. (a) AMPds2, heating load disaggregation, 24 - 30 March 2014. (b) UK-DALE, lighting load disaggregation, 11 - 17 January 2016.	85
Fig. 5.1. Actual 3-year measurements with indicated different types of outliers for an individual WT.	102
Fig. 5.2. Actual 3-year measurements with indicated different types of outliers for: (a) an individual WT, and (b) a whole WF (derived from the individual WTs in the same WF). (The A-E marks indicate characteristics of various types of WT outliers.) ...	102
Fig. 5.3. Actual 1-year measurements with indicated constant WS-variable P_{out} outliers for a directly measured whole WF.	102
Fig. 6.1. Dalry wind farm layout.	134
Fig. 6.2. Actual measurements with indicated different types of outliers. $\rho_{0.97}$, $\rho_{1.12}$ and $\rho_{1.27}$ refer to air densities equal to 0.97 kg/m ³ , 1.12 kg/m ³ and 1.27 kg/m ³ , respectively: (a) 3-year recordings for an individual WT in Scotland Wind Farm Dataset (WT1); (b) 3-year recordings for the whole WF in Scotland Wind Farm Dataset (derived from the individual WTs in the same WF [142]); (c) 4-year recordings for a directly measured whole WF in Zelengrad Wind Farm Dataset.	136
Fig. 6.3. WT-level outlier detection flow chart (Mfr PC ρ_x refers to air density specified manufacturer power curve).	139
Fig. 6.4. WT1 in Scotland Wind Farm Dataset outlier identification results.	145

Fig. 6.5. The averages of percentages of outlier detection results for all six WTs in Scotland Wind Farm Dataset.	145
Fig. 6.6. 6P-PC fitting results on normal data of WT1, Scotland Wind Farm Dataset.	147
Fig. 6.7. WF operating regime classification results for Scotland Wind Farm Dataset, indicated by different colours and (A, B, C) combinations at the top.	149
Fig. 6.8. Equivalent wind farm model for Scotland Wind Farm Dataset.....	151
Fig. 6.9. Error evaluation for “linear scaling (plus bias) relationship”.	151
Fig. 6.10. Results of equivalent wind farm model fitting algorithm for single sensor WF measurements (Zelengrad Wind Farm Dataset). (a) WF operating regime classification results indicated by different colours. (b) Equivalent wind farm model.	154
Fig. 7.1. Comparison of 2-pW and 3-pW fittings on a synthetic dataset.....	160
Fig. 7.2. Illustration of GND with different shape parameters.	160
Fig. 7.3. Comparison of PDFs of fitted distributions for wind speed.	163
Fig. 7.4. Comparison of CDFs of fitted distributions for wind speed.....	163
Fig. 7.5. Wind rose plots (N = 0 degree, E = 90 degrees, S = 180 degrees, and W = 270 degrees) for wind direction in different wind speed bins: a) 6 – 7 m/s; b) 11 – 12 m/s; c) 17 – 18 m/s; d) 24 – 25 m/s.....	166
Fig. 7.6. Fitting wind direction in different wind speed bins: a) 6 – 7 m/s; b) 11 – 12 m/s; c) 17 – 18 m/s; d) 24 – 25 m/s.....	167
Fig. 7.7. Wind direction fitting error metrics: a) RMSE; b) MAEAAB; c) MAE; d) WAE.....	167
Fig. 7.8. Fitting power output in different wind speed bins: a) 6 – 7 m/s; b) 11 – 12 m/s; c) 17 – 18 m/s; d) 24 – 25 m/s.....	168
Fig. 7.9. Power output fitting error metrics: a) RMSE; b) MAEAAB; c) MAE; d) WAE.....	169
Fig. 8.1. An illustration of the impact of WD on WT power outputs: a) comparison of Mfr-PC and AO-PC models for the four main WD quadrants: b) and c) WF layout, with positions of WTs within the WF.	173
Fig. 8.2. An example of empirical/actual WT Pout uncertainty	175

Fig. 8.3. Comparison of results for one whole day in winter for input WS and WD without uncertainties: a) mean values; b) 2D Copula P_{out} -WS GMCM model with P_{out} uncertainties; and c) 3D Copula P_{out} -WS-WD with P_{out} uncertainties.	184
Fig. 8.4. Comparison of results for one whole day in winter (same as in Fig. 8.3) for input WS and WD with uncertainties: a) mean values; b) 2D Copula P_{out} -WS GMCM model with P_{out} uncertainties; and c) 3D Copula P_{out} -WS-WD with P_{out} uncertainties.	187
Fig. 9.1. An example of clustered P_{out} values for different ORs of Bruska WF: different colours and numbers at the top of the Fig. represent different numbers of WTs in normal operation.	194
Fig. 9.2. A general CNN-BiLSTM architecture.....	195
Fig. 9.3. The results of Bruska WF forecasting weather variables in “test week”: (a) WS, (b) AD, (c) WD (zero degree corresponds to a true North direction).	200
Fig. 9.4. The results of forecasting WF Ors in “test week”: (a) Bruska WF, (b) Jelinak WF (Y-axis represents the number of WTs in normal operation).	202
Fig. 9.5. Implemented 4-D Vine Copula structure.....	204
Fig. 9.6. Conditional distribution of WF P_{out} given actual measurements of WS, AD, WD and known ORs in “test week”: (a) Bruska WF, (b) Jelinak WF.....	205
Fig. 9.7. Flowchart of Bayesian CNN-BiLSTM and Vine-GMCM framework....	208
Fig. 9.8. The results of forecasting of WF P_{out} by combined Bayesian CNN-BiLSTM and Vine-GMCM model in “test week”: (a) Bruska WF, (b) Jelinak WF.....	208
Fig. 9.9. Reliability diagram with and without considering wind farm operating regimes in “test week”.	210
Fig. 9.10. The results of two benchmarking models for forecasting of WF P_{out} in “test week”: (a) Bruska WF by Seq2Seq, (b) Jelinak WF by Seq2Seq, (c) Bruska WF by MC, (d) Jelinak WF by MC, (e) Bruska WF by persistence, (f) Jelinak WF by persistence, (g) Bruska WF by single-stage Bayesian CNN-BiLSTM, (h) Jelinak WF by single-stage Bayesian CNN-BiLSTM.....	216
Fig. 9.11. Visualisation of Winkler score (α is the significance level, see Equation (2.16))......	224
Fig. 9.12. Example of prediction distributions.....	225
Fig. 9.13. Forecasting results for the first 100 hours in Jan 2013 for WF1.	229

List of Tables

Table 3.1 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.5(a))	46
Table 3.2 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.5(b))	46
Table 3.3 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.5(c))	47
Table 3.4 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.6(a))	47
Table 3.5 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.6(b))	48
Table 3.6 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.6(c))	48
Table 3.7 Evaluation of SB-LSTM Performance (For 7 Days in Fig. 3.7).....	50
Table 3.8 Evaluation of SB-LSTM Performance (For 7 Days in Fig. 3.8).....	51
Table 3.9 Evaluation of SB-LSTM Performance (For 7 Days in Fig. 3.9).....	51
Table 3.10 Performance of Day-ahead Forecasting of Total Demand and Heating Demand (For 7 Days in Fig. 3.10)	53
Table 3.11 Performance of Day-ahead Forecasting of Total Demand and Heating Demand (For 7 Days in Fig. 3.11)	54
Table 3.12 Performance of Day-ahead Forecasting of Total Demand and Heating Demand (For 7 Days in Fig. 3.12)	54
Table 4.1 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.3 – 4.5	75
Table 4.2 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.6 – 4.7	75
Table 4.3 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.8 – 4.11	75

Table 4.4 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.12 – 4.14	75
Table 4.5 Spearman Correlation Coefficients between the Frequency Components of Total Load and Temperature/Solar Irradiance in Example Results (D_i only).....	77
Table 4.6 Selection of Optimal Settings of Hyperparameters by BO for All Load Disaggregation Models	84
Table 4.7 Comparison of Performance of Load Disaggregation Models on Two Example Days	85
Table 4.8 Comparison of Performance of Load Disaggregation Models on Two Example Weeks.....	86
Table 4.9 Comparison of Performance of Load Disaggregation Models on the Whole Test Sets	86
Table 4.10 Comparison of Training Time for All Models.....	88
Table 5.1 Summary of Wind Turbine and Wind Farm Measurements in Available Outlier-Related Literature	97
Table 5.2 Summary of Characteristics and Causes of Outliers in WT and WF Measurements	107
Table 5.3 Summary of Common Causes of Outliers in WT and WF Measurements	115
Table 5.4 Summary of Possible Problems or Limitations of Reviewed Outlier Detection Methods in Practical Applications.....	124
Table 6.1 Operating Characteristics of Vestas V90-3MW [87].....	133
Table 6.2 Dalry Wind Farm Parameters	134
Table 6.3 Zelengrad Wind Farm Parameters	135
Table 6.4 Summary of Characteristics and Causes of Outliers in WT and WF Measurements	137
Table 6.5 Suggestions for Treatments of WT-Level Outliers for both WT-Level and WF-Level Studies.....	145
Table 6.6 Percentages of WF Operating Regimes	149
Table 6.7 Comparisons of Power Output Estimations for Scotland Wind Farm Dataset (Derived from WTs).....	151

Table 6.8 Comparisons of Power Output Estimations for Zelengrad Wind Farm Dataset (Single Sensor Measurements).....	155
Table 7.1 Distribution Fitting Results for Wind Speed Measurements	162
Table 8.1 Discretisation Classes for WD (Columns 1-3), and WS (Columns 4-7) .	181
Table 8.2 Comparison of P_{out} -WS Model, P_{out} -WS-WD Model, Manufacturer Power Curve and Actual Measurements for the Same Day in Fig. 8.3.....	184
Table 8.3 Comparison of P_{out} -WS Model, P_{out} -WS-WD Model, Manufacturer Power Curve and Actual Measurements for Whole Seasons	185
Table 8.4 Comparison of P_{out} -WS model and P_{out} -WS-WD model for the same day in Fig. 8.4	187
Table 9.1 Bruska Wind Farm Parameters	192
Table 9.2 Jelinak Wind Farm Parameters	192
Table 9.3 Error of Forecasting Weather Variables in “Test Week”	201
Table 9.4 Errors of WF P_{out} Estimations of WS, AD, WD and Operating Regimes by Vine-GMCM in “Test Week”	206
Table 9.5 Comparison of Results with and without Considering Wind Farm Operating Regimes in “Test Week”	206
Table 9.6 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Week”	217
Table 9.7 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Year”	218
Table 9.8 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Spring Season”	219
Table 9.9 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Summer Season”	220
Table 9.10 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Autumn Season”	221
Table 9.11 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Winter Season”	222
Table 9.12 Comparison and Benchmarked Numerical Values of the Reduced Model with the Results from [256, 258].....	228

Table 9.13 Comparison and Benchmarked Numerical Values of the Reduced Model with the Results from [259].....	228
---	-----

Acronyms and Abbreviations

2-pW	Two-parameter Weibull distribution
3-pW	Three-parameter Weibull distribution
4PLF	Four-parameter logistic function
5PLF	Five-parameter logistic function
6PLF	Six-parameter logistic function
6P-PC	Six-parameter power curve model
ACE	Average coverage error
AD	Air density
AI	Artificial intelligence
AMPds2	The Almanac of Minutely Power Dataset (Version 2)
ANFIS	Adaptive neuro-fuzzy inference
AO-PC	Average operational power curve
AP	Active power
BIC	Bayesian information criterion
BiLSTM	Bidirectional long short-term memory
BNN	Bayesian neural network
BO	Bayesian optimisation
BPTT	Backpropagation through time
CDF	Cumulative density function
CMF	Cumulative mass function
CNN	Convolutional neural network
CO	Combinatorial optimisation
Conv. Fourier	Conventional Fourier transform
CRPS	Continuous ranked probability score
CV	Computer vision
DBSCAN	Density-based spatial clustering of applications with noise
DD	Day of the month
DEM	Demand
DFIG	Doubly-fed induction generator
DFT	Discrete Fourier transform
DL	Deep learning
DST	Daylight saving time

DW	Day of the week
EM	Expectation–maximisation
EV	Electric vehicle
FCM	Fuzzy c-means
FHMM	Factorial hidden Markov model
FN	False negative
FOMC	First order Markov chain
FP	False positive
FR	Fourier series regression and correlation analysis results
Freq.	Frequency
FSR	Fourier series regression
GA	Genetic algorithm
GARCH	Generalised autoregressive conditional heteroskedasticity
GBRT	Gradient boosting regression trees
GEFCom2014	Global Energy Forecasting Competition 2014
GMCM	Gaussian mixture Copula model
GMM	Gaussian mixture model
GND	Generalised normal distribution
GP	Gaussian process
GRU	Gated recurrent units
GW	Gigawatt
HH	Hour
HLD	Holiday
HMM	Hidden Markov model
HP	Hyperparameter
i.i.d.	Independent and identically distributed
IALM	Intrusive appliance load monitoring
ILM	Intrusive load monitoring
IQR	Interquartile range
k -CV	k -cross validation
KL	Kullback-Leibler
KLOF	Kernel-based local outlier factor
KNN	k -nearest neighbour

k -NN	k -nearest neighbour
kW	Kilowatt
LASSO	Least absolute shrinkage and selection operator
LMedS	Least median of squares
LOF	Local outlier factor
LReLU	Leaky rectified linear units
LSBoost	Least-square boosting
LS-SVR	Least squares support vector regression
LSTM	Long short-term memory
LV	Low voltage
MAD	Median absolute deviation
MAE	Mean absolute error
MAEAAB	Maximum absolute error among all Bins
MAPE	Mean absolute percentage error
MAR	Missing at random
MC	Markov chain
MCAR	Missing completely at random
MCMC	Markov chain Monte Carlo
MD	Meteorological data
MERRA-2	The Modern-Era Retrospective Analysis for Research and Applications, Version 2
Mfr-PC	Manufacturer's power curve
MGND	Mixture of generalised Normal distributions
ML	Machine learning
MLE	Maximum likelihood estimation
MLP	Multilayer perceptron
MM	Month
MMO	Mathematical morphology operation
MNAR	Missing not at random
MOM	Method of moments
MP	Moon phase
MV	Medium voltage
MVM	Mixture of Von-Mises distributions
MWB	Mixture of three-parameter Weibull distributions

NI	Nonignorable
NIALM	Nonintrusive appliance load monitoring
NILM	Nonintrusive load monitoring
NLP	Natural language processing
NN	Neural network
NWP	Numerical weather prediction
Ofgem	The Office of Gas and Electricity Markets
OLS	Ordinary least squares
OR	Operating regime
OSTAR	Outlier smooth transition autoregressive
PC	Power curve
PCA	Principal component analysis
PDF	Probability density function
PICP	Prediction interval coverage probability
PINC	Prediction interval nominal confidence
PLF	Parameter logistic function
PMF	Probability mass function
P_{out}	Power output
$P_{\text{out_max}}$	Maximum power output
PPT	Precipitation
Pred.	Prediction
PSO	Particle swarm optimisation
PV	Photovoltaic
ReLu	Rectified linear unit
RMSE	Root-mean-square error
RNN	Recurrent neural network
SA	Sun azimuth angle
SB-LSTM	Stacked bidirectional long short-term memory
SCV	Standard calendar variable
SDAE	Stacked denoising autoencoder
SE	Sun elevation angle
Seq2Seq	Sequence-to-sequence
SI	Solar irradiation

std.	Standard deviation
STW	Sliding time window
SVM	Support vector machine
SVR	Support vector regression
TEMP	Temperature
TN	True negative
TP	True positive
UK	United Kingdom
UK-DALE	UK domestic appliance-level electricity
VI	Variational inference
VMD	Von-Mises distribution
w/	With
w/o	Without
WD	Wind direction
WF	Wind farm
WMAE	Weighted mean absolute error
WRF	Weather research and forecasting model
WS	Wind speed
WT	Wind turbine
YY	Year

List of Symbols

(A, B, C)	(Number of WTs operating normally, Number of WTs in curtailment, Number of WTs in outage)
$[\cdot]$	Floor function
Δu	Normalised uncertainty range estimation error
A	Wind turbine rotor swept area
a	The first asymptote of five-parameter logistic function
a_0	DC component in Fourier series decomposition
a_k	Fourier coefficients ($k = 1, 2, \dots, \infty$) in Fourier series decomposition
a_x	A Fourier coefficient in single frequency component extraction
b	The hill slope of five-parameter logistic function
b_f	LSTM cell forget gate bias parameter
b_i	LSTM cell remember gate bias parameter
b_k	Fourier coefficients ($k = 1, 2, \dots, \infty$) in Fourier series decomposition
b_o	LSTM cell output gate weights parameter
b_x	A Fourier coefficient in single frequency component extraction
b_z	LSTM cell candidate vector bias parameter
c	The inflection point of five-parameter logistic function
$\mathbf{C}$	Markov chain cumulative transition probability matrix
$C(\cdot)$	Copula distribution function
$c(\cdot)$	Copula density function
c_{ijk}	Markov chain cumulative probability of transferring from state I and following state j to the states $1, 2, \dots, k$
$\text{cov}(\cdot)$	Covariance function
C_p	Wind turbine power coefficient
c_t	LSTM cell state at time t
c_x	DC component in single frequency component extraction
d	The second asymptote of five-parameter logistic function
$\mathbf{D}$	Wind speed standard deviation matrix
D_i	i -th day
E_o	Overestimated energy consumption

E_T	Total energy consumption
E_U	Underestimated energy consumption
$F(\cdot)$	Marginal distribution function
$f(\cdot)$	Marginal density function
$f(Y; X)$	The deterministic relationship between predictor X and response Y
fm	Filtered matrix after convolution
f_t	LSTM cell forget gate at time t
f_{TN}	PDF of truncated normal distribution
g	Asymmetry factor of five-parameter logistic function
gm	Target matrix in convolution
g_N	PDF of normal distribution
$H(\cdot)$	Cross-entropy function
h_t	RNN (including LSTM) cell hidden state at time t
i	Row index for a matrix; or index for the data
$I_0(\cdot)$	The modified Bessel function of the first kind and order zero
i_t	LSTM cell remember gate at time t
j	Column index for a matrix; or index for the data
K	Kernel matrix
k	Index for the data
km	Kernel matrix in convolution
n	Index for the data (usually represent the length)
N	Number of wind turbines in a wind farm
N_0	Number of normally operating wind turbines in a wind farm
o_t	LSTM cell output gate at time t
P	Markov chain transition probability matrix
$P(X)$	Probability of X
$P(X, Y)$	The joint probability of X and Y
$P(X Y)$	The conditional probability of X for given Y
P_{CLT}	Wind turbine curtailed maximum power output
p_{ijk}	Markov chain probability of transferring from state i and following state j to the state k
P_{out}	Power output matrix
$P_{out,5\%}$	5 th percentile of power output

$P_{out,95\%}$	95 th percentile of power output
$\mathbf{P}_{out; w}$	Windowed power output matrix
P_{rated}	Wind turbine rated power output; v_i , v_r , and v_o are cut-in, rated and cut-out
P_{WF}	Wind farm power output
$P_{WF FULL}$	Wind farm power output when all WTs are normally operating
P_{WT}	Wind turbine power output
Q_x	Quantile value
R_e	Empirical uncertainty range of P_{out} (difference between 95 th percentile of power output and 5 th percentile of power output)
R_M	Uncertainty range of P_{out} by the model (difference between estimated 95 th percentile of power output and estimated 5 th percentile of power output)
r_x	Rank of random variable x
r_y	Rank of random variable y
s	Markov chain number of states
s_t	LSTM cell input at time t
t	Time
$\mathbf{T}$	Time stamp matrix
$\mathbf{T}_w$	Windowed time stamp matrix
$\mathbf{V}$	Wind speed matrix
v_i	Cut-in wind speed
v_o	Cut-out wind speed
v_r	Rated wind speed
$\mathbf{V}_w$	Windowed wind speed matrix
W_f	LSTM cell forget gate weights parameter
W_i	LSTM cell remember gate weights parameter
W_o	LSTM cell output gate bias parameter
W_z	LSTM cell candidate vector weights parameter
$\mathbf{X}$	Matrix of predictor variable
x_i	i -th predictor variable
x_t	Predictor variable at time t
$\mathbf{y}$	Matrix of response variable
y_i	i -th response variable
y_t	Response variable at time t

z_t	LSTM cell candidate vector at time t
α	Significance level
α	Generalised normal distribution scale parameter
β	Matrix of regression coefficients
β	Generalised normal distribution shape parameter
$\Gamma(\cdot)$	Gamma function
$\delta(\cdot)$	Digamma function
Δi	Shifting along the row axis for a matrix
Δj	Shifting along the column axis for a matrix
ζ	Generalised normal distribution location parameter
θ	Weibull distribution (3-pW only) location parameter
κ	Weibull distribution (2-pW and 3-pW) shape parameter
λ	Weibull distribution (2-pW and 3-pW) scale parameter
μ	Mean value
$\mathbf{v}$	Vector of parameters for individual distributions of a mixture distribution
ρ	Von-Mises distribution location parameter
ρ_{xy}	Spearman correlation coefficient between x and y
σ	Standard deviation
τ	Von-Mises distribution concentration parameter
$\Phi(\cdot)$	Standard normal CDF
$\phi(\cdot)$	Standard normal PDF
$\psi(\cdot)$	Gaussian mixture model probability density function
$\Psi(\cdot)$	Gaussian mixture model probability distribution function
ω	Weight of a member distribution of a mixture distribution
Ω	Vector of weights for individual distributions of a mixture distribution
δu	The arithmetic mean value of the error of 5 th percentile estimation and error of 95 th percentile estimation
$\delta u_{5\%}$	Error of 5 th percentile estimation
$\delta u_{95\%}$	Error of 95 th percentile estimation

1.1 Introduction

Power systems have seen continuous development and evolution since the late 19th century. In recent years, decarbonisation and concerns about the climate change bring new challenges to power systems and set requirements for higher efficiency and lower emission [1]. Many countries have set grand targets for carbon reduction and clean energy transition [2]. In the United Kingdom (UK), the government has set in law the most ambitious climate change target all over the world, which is to cut emissions by 78 % by 2035 compared to 1990 levels. The UK also determines to achieve net zero emissions by 2050 [3].

To achieve those targets, there will be various smart technologies introduced to modernise the power systems, and also rapid penetration of renewable energy-based generations. For example, the Office of Gas and Electricity Markets (Ofgem) in the UK has set plans to advocate smart technologies and enable smart homes and business [4]. For example, the electricity generation from wind power in the UK has increased by 715 % from 2009 to 2020 [5].

Correct assessment of the performance of modern power systems requires to evaluate much wider ranges and higher levels of uncertainties introduced by the implementation of new technologies on both demand and supply sides. On the demand side, the deployment of new types of load, emergence of demand with strong spatio-temporal variation patterns (e.g., electric vehicle, EV, charging) and various demand-side management schemes all contribute to the increased variability and uncertainty of the power system. On the supply side, the significant increase in numbers and sizes of inherently stochastic renewable electricity generation systems (e.g., wind and photovoltaic, PV) also introduces much higher uncertainties in the operation, control, dispatching and forecasting, than in the traditional power supply systems.

Modern power supply systems also feature a much larger amount of available measurement data, due to the strong development and much wider implementation of

advanced metering and monitoring technologies. Extracting useful information from the large available datasets by suitable processing and then understanding of the monitored phenomena, processes and systems is often a challenging task, but there are opportunities to utilise data-driven methods and gain insights into, and understanding of the related operational, functional and “behavioural” characteristics on both demand and supply sides, which then can be implemented for a more realistic modelling and more confident forecasting, i.e., addressing and resolving the problems.

The improved modelling and forecasting of electricity demand and supply are becoming increasingly important for the operation and planning of existing electricity networks and future smart grids. The understanding and correct modelling of underlying uncertainties can improve system reliability, security, resilience and supply quality, and it can also enhance the ability to accurately predict variable outputs of renewable energy sources, such as wind and PV generation systems, and then apply suitable control and balancing actions, particularly in case of highly dispersed and non-dispatchable distributed generation units connected in distribution networks. The research in modelling and forecasting techniques and uncertainty analysis is crucial for more optimal balancing of variable generation with controllable loads (and vice versa), including individual and multi-objective evaluation of different and operational requirements, such as improved efficiency, higher asset utilisation, minimised environmental impact and enhanced techno-economic system performance.

This thesis addresses the above problems by processing and analysing years of actual measurement from wind farms, households, and substations. It has developed and evaluated novel and improved data-driven methodologies to capture the correlation and uncertainty. State-of-the-art models have been investigated during the PhD, including neural network and deep learning for forecasting, optimisation for curve and distribution fitting, and Copula model for correlation estimation, and the Bayesian method for uncertainty analysis. And the novel utilisation and improvement of the models also lead to satisfactory performance when solving the problems described above, which means the calculated error metrics of proposed models indicate better results than the benchmarking models in other publications.

1.2 The Need for Improved Modelling, Forecasting and Uncertainty Analysis on Electricity Demand and Supply Sides

Electricity demand side analysis is a pillar of this PhD research. And it starts from load forecasting, which is generally aimed at providing an accurate estimation of future energy consumption, typically based on available historical recordings and other relevant data. Different forecasting horizons are of interest to different studies. Long-term forecasting is important for power system planning, while short-term forecasting, especially day-ahead forecasting, is essential for operational studies [6]. Typical load profiles exhibit periodicity that follows day-night cycle, but are also influenced by meteorological, economic and socio-behavioural factors, which introduce additional uncertainties. To improve the model performance, additional predictor variables may be added to the model so that all related information can be analysed, which however increases the complexity and could bring a challenge known as “the curse of dimensionality”. In addition, among the advanced and accurate models, the proper model selection for a specific practical forecasting problem is also non-trivial, requiring an understanding of both the modelling task and the property of the model.

In addition to load forecasting, load disaggregation research is also important to the electricity demand side research. Load disaggregation denotes a process of identifying specific load types in the total demands, e.g., proportion of demand-manageable loads, such as heating or cooling loads. Demand side management is becoming increasingly important for modern power systems, because it can balance renewable generation and provide various system support services [7, 8]. Different load types have different energy consumption patterns, and they are also impacted by appliance-specific characteristics, weather conditions, user settings and habits, as well as wider socio-economic factors [9]. This introduces considerable difficulties for load disaggregation, and deep learning models are being increasingly used for load disaggregation due to improvement in accuracy and performance. However, these models are often criticised as “black box” models, due to a lack of interpretability [10]. Furthermore, the optimisation of model hyperparameters and integrating domain knowledge (feature engineering) into the models are also important for decision making and improving

model performance. Frequency domain analysis and demand-weather correlation can lead to additional predictor variables for the sophisticated deep learning models, which can even further improve accuracy and, importantly, provide an increased level of interpretability.

On the electricity demand and supply side, over the last few decades, there was a significant increase in numbers and sizes of installed renewable energy generation systems, particularly wind and solar energy, which has effectively resulted in the much wider variations of network power flows and has also introduced higher ranges of uncertainties in network operating conditions. In that context, correct representation of the power conversion process and accurate characterisation of the wind turbine (WT) and whole wind farm (WF) operation can benefit both the wind farm operator and also the network operator.

In wind energy research, power curves (PCs), e.g., the power curves provided by the wind turbine manufacturers, are commonly used for modelling wind-based electricity generation systems, giving the deterministic relationship between the input wind speed (WS) and power output (P_{out}). However, since manufacturer power curves (Mfr-PCs) are obtained under predefined standard conditions [11, 12], they cannot correctly reflect all practical factors on different sites, which introduces additional uncertainties. On one hand, wind energy features strong variability, which should be also modelled in terms of wind direction (WD) and air density (AD) variations. On the other hand, WTs are complex energy conversion devices and dynamics of individual WTs, as well as wake effects of WTs operating in a WF [13], also have impact on the wind farm power output. A simple wind speed-power output PC model neglects these (and other) explanatory variables and influential factors. Additionally, deterministic power curves cannot be used to assess the uncertainty of power outputs, which significantly limits their application in operating studies.

Due to a general need to assess the variations and uncertainties in renewable generation, probabilistic modelling of, e.g., wind-based electricity generation systems using field measurements and data-driven methods, is becoming increasingly popular in operating research. Many wind turbine and wind farm models are typically obtained by processing large amounts of measurements from the individual wind turbines and

the whole wind farms. These measurements usually contain corrupted or abnormal data and outliers, and it is therefore important to understand their causes and origins, in order to identify suitable methods for identification and handling, as otherwise the developed models will be biased, or even erroneous.

To model the variation of wind power outputs, some of the existing probabilistic modelling methods (e.g., [14, 15]) make simple assumptions that they should follow normal distribution between cut-in and rated wind speeds, because power outputs are impacted by many factors, which usually leads to normal distribution. And they also model the power outputs outside this wind speed interval to be deterministic values. More specifically, the power outputs are rated values between rated and cut-out wind speeds, and they are zeros for either below cut-in or above cut-out wind speeds. However, according to the investigation performed on several years of actual measurements of the wind farm in this thesis (see Chapter 7), in many wind speed bins, the distributions of power outputs within the bins have clearly different skewness and kurtosis from normal distributions, and there are variations of power outputs when wind speeds are between cut-in and rated, and also between rated and cut-out wind speeds. Therefore, it is essential to develop improved probabilistic methods for a more accurate wind farm power output modelling, which can more precisely describe the operational relationships in terms of all factors and variables of influence.

Several explanatory variables have impacts on the variation of power outputs. Many current wind turbine and wind farm models take only wind speed as the predictor variable, but neglect other meteorological variables (e.g., wind direction and air density), although Chapter 8 and [16] clearly show that besides uncertainties in wind speed, variations in air density and wind direction may also have strong impact on the wind farm power outputs. This is easy to understand, as the air density is included in the formulae/equation for calculating wind energy (see Equation (9.9) in Chapter 9). In terms of wind direction, although wind turbine with the automatic yawing system will always try to orient the WT so its blades' plane of rotation is perpendicular on the wind direction, the yawing mechanism is relatively slow and any change of wind direction will affect the power output. Therefore, it is beneficial to have an improved

multi-dimensional model that can correctly analyse the impacts of wind direction and air density.

It is also crucial to understand that a wind farm is an aggregation of wind turbines. The existing wind energy research often neglects the relationship between wind turbine level modelling and wind farm level modelling and forecasting. A wind farm typically consists of multiple individual wind turbines and at any moment, each wind turbine in a wind farm may be in one of three general states: a) normally operating, b) curtailed, or c) disconnected due to the planned or unplanned outages. The combinations of the possible wind turbines states form the wind farm operating regime. In the existing literature, only some of the references make distinctions between different operating regimes (e.g., [17, 18]), although the wind farm in various operating regimes can have significantly different power output characteristics. An improved multiple state wind farm model is needed, because it can more accurately model the whole wind farm and provide a better understanding of operating regime uncertainty. This is further discussed in Chapter 6 and Chapter 9 in this thesis.

1.3 Research Objectives and Main Contributions of the Thesis

The main objective of the research presented in this thesis is development of improved probabilistic, correlational and forecasting methodologies for a more accurate capturing, analysis and quantification of uncertainties in the operation of modern power systems. Processing, cleaning and outlier analysis of the measured data are essential for accurate modelling and are therefore an important additional target of the presented PhD research. The final objective is to incorporate developed operational and forecasting models into a general framework for the analysis, and to validate and test them using the actual measurements of load demands and wind farm outputs, in order to demonstrate higher accuracy and improved performance when compared to and benchmarked against the existing state-of-the-art methods and models.

The main results of the thesis have been accepted and published in 5 journal papers ([19-23]). In addition, the methods and results of the thesis have also been presented in 13 conference papers ([24-36]). More specifically, the related work in Chapter 3 has

been published in [19, 29, 30, 35]. The research in Chapter 4 has been published in [22]. The review in Chapter 5 has been published in [21]. The related work in Chapter 6 has been published in [25] and [28], and it is currently being improved and drafted to be submitted to IEEE Transactions on Sustainable Energy. The methods and results in Chapter 7 has been published in [24]. The investigation and research outcomes of Chapter 8 have been published in [20, 27]. And the findings and output of Chapter 9 has been published in [23].

The main contributions are summarised as the followings:

- ***Comparison of three state-of-the-art load forecasting models and inclusion of weather variables in the input data:*** Three different day-ahead load forecasting models, multilayer perceptron (MLP), gradient boosting regression trees (GBRT) and stacked bidirectional long short-term memory (SB-LSTM), are built in Chapter 3, based on historical load recordings and meteorological data. Their performance is compared and analysed, showing that SB-LSTM model produces the best results and smallest errors for the two considered datasets. Therefore, SB-LSTM is investigated further and extended to produce a week-ahead load forecasting (also see Chapter 3), where meteorological factors (temperature, solar irradiation, precipitation, etc.) are also included in the model. The presented results and analysis allow to evaluate how accurately SB-LSTM approach can provide predictions for both day-ahead and week-ahead load forecasting, taking into account meteorological information.
- ***Proposal of a novel frequency domain decomposition and correlation method for improved load disaggregation results of LSTM-based models, and application of Bayesian optimisation (BO) to reduce the effects of uncertain hyperparameters:*** A novel approach in which periodical changes in the total demands (e.g., daily, weekly, and seasonal variations) are disaggregated into corresponding frequency components and correlated with the same frequency components in the meteorological variables (e.g., temperature and solar irradiance) is introduced in Chapter 4, allowing to select combinations of frequency components with the strongest correlations as the additional explanatory variables. These results are used as an additional feature of the

convolutional neural network and bidirectional long short-term memory (CNN-BiLSTM) load disaggregation model. And the hyperparameters of CNN-BiLSTM are selected using Bayesian optimisation to reduce the side effects of uncertainty in the hyperparameters. The obtained results and comparisons demonstrate that the CNN-BiLSTM model in which correlated frequency components are used as the additional explanatory variables is more accurate for load disaggregation than the conventional/naïve CNN-BiLSTM model without frequency components (“black box model”).

- ***A detailed review of approaches for the detection and treatment of outliers in processing WT and WF measurements:*** A comprehensive and systematic review of approaches for the detection and treatment of outliers in processing WT and WF measurements is presented in Chapter 5. It starts from the discussion of the commonly measured parameters, variables and resolutions, as well as the corresponding requirements and recommendations in related standards. Afterwards, characteristics and causes of outliers reported in existing literature are discussed and illustrated, as well as the requirements for the data rejection in related standard. Next, outlier identification methods are reviewed, which is followed by a review of approaches for testing the success of outlier removal procedures, with a discussion of their potential negative effects and impact on the WT and WF models. Finally, some issues and concerns that could be of interests for further research on the detection and treatment of outliers in processing WT and WF measurements are discussed.
- ***Proposal of a generic framework for WT-level outlier detection and a simulation-based approach that can explain outliers due to WT cut-out effects and averaging window, including proposal of a new six-parameter power curve model and analysis of the relationship between WT and WF level data:*** Chapter 6 firstly identifies outliers in WT recordings, using a framework consisting of sliding window and a novel simulation-based approach with random walk Metropolis and manufacturer power curve with hysteresis control. A six-parameter power curve model is proposed and fitted on the filtered WT measurements, which can simultaneously model the cut-in and cut-out regions. Then, WF-level analysis is performed for the measured data with known

individual WT recordings. An equivalent WF model is presented, which can represent the relationships among various WF operating regimes. Afterwards, another WF-level case study without individual WT measurements is investigated, where an iterative algorithm is proposed to fit the proposed equivalent WF model. Both case studies show the importance of considering actual operating conditions in identifying outliers and improving the estimations of WF power outputs.

- ***Proposal of a multi-dimensional probabilistic WT and WF model, which can handle both the uncertain inputs (e.g., ranges of wind speed, wind direction and air density values) and model parameter uncertainties:*** A general model for the analysis of the P_{out} -WS-WD correlational structure is presented in Chapter 8, based on Gaussian mixture Copula model (GMCM) and Vine Copula, where GMCM accurately represents the pair dependencies in Vine Copula (i.e., Vine-GMCM framework). This model shows that wind direction is essential for the accurate representation of wind energy resources. The Vine-GMCM model is modified so that it can provide conditional probability estimation for $P(P_{out}|WS, WD)$ when WS and WD are uncertain. The obtained results show that the inclusion of WD will provide noticeable improvement for models with no uncertainties in input WS and WD data, while in the case of the forecasted WS and WD data with uncertainties, WS is a much stronger contributor to the total WT P_{out} uncertainty than WD.

In Chapter 9, this Vine-GMCM model is extended to capture P_{out} -WS-WD-AD correlations, as air density is also an important predictor variable in wind energy analysis. Another sophisticated probabilistic model, Bayesian convolutional neural network-bidirectional long short-term memory (Bayesian CNN-BiLSTM), is also investigated and applied, which can capture both epistemic and aleatoric uncertainties [37] and provide accurate forecasting for the weather variables and operating regimes. The proposed combined model is applied and validated using the actual field measurements from two WF clusters in close locations in Croatia and is also benchmarked against several other models. Additionally, the reduced P_{out} -WS-WD model is also benchmarked against some of the models participating in the Global Energy Forecasting Competition 2014 (GEFCom2014) competition [38]. The results in Chapter 9 show the satisfactory performance of all the

proposed models, confirming that the most complex P_{out} -WS-WD-AD multidimensional model achieves the highest accuracy, and the two-stage WF P_{out} prediction framework provides the most confident probabilistic forecasting.

1.4 Thesis Structure

This PhD thesis has ten chapters. In addition to this first chapter, which provides a general introduction and background of the thesis and also discusses the motivation, research objectives, and main contributions of the thesis, the summary of the contents of the other chapters is given as follows:

Chapter 2: This chapter presents a literature review of the theoretical methods, error metrics, domain knowledge and relevant discussions about uncertainty in this thesis.

Chapter 3: This chapter analyses the variations of electricity demand time series data at three different aggregation levels: a household, a substation and a city. It compares three models for a day-ahead load forecasting, where the one with the highest accuracy (SB-LSTM) is then extended to a week-ahead load forecasting and both the total load and heating load forecasting. All models have considered the impacts of weather and calendar variables.

Chapter 4: In this chapter, a novel frequency domain decomposition and correlation method is proposed. The results are utilised by the CNN-BiLSTM load disaggregation model. Comparisons are also made to analyse the impact of including the frequency domain and correlation features on forecasting. In addition, to help select an appropriate set of hyperparameters and thus further the uncertainty, Bayesian optimisation is executed before training and tuning, to guide the hyperparameters setting.

Chapter 5: This chapter presents a detailed and comprehensive literature review about outlier detection and treatment methods in processing WT and WF measurements.

Chapter 6: In this chapter, a novel outlier detection framework is presented, which consists of both time series analysis and simulation-based methods. It proposes a six-parameter power curve model for all wind speed regions and it also demonstrates the

relationship between WT-level and WF-level analysis, which led to the introduction of an important concept of WF operating regimes.

Chapter 7: This chapter applies several mixture distributions with explicit expressions to fit the historical measurements of wind speed, wind direction and power output data. These models are specifically built for comparison with other models developed in this thesis and can be used for the classical probabilistic representation of a WF power output performance.

Chapter 8: This chapter combines Vine Copula and GMCM to build a correlational 3D model for P_{out} -WS-WD, which can take both deterministic and uncertain predictor variables.

Chapter 9: This chapter uses Bayesian CNN-BiLSTM to produce probabilistic forecasting for the weather variables. An extended 4D Vine-GMCM model is developed and applied to estimate the conditional probability of power output, $P(P_{\text{out}}|\text{WS, WD, AD})$. The wind farm operating regime is also integrated with the forecasting framework.

Chapter 10: In this chapter, the main conclusions of the work are presented. It also discusses the limitations of the research and proposes recommendations for future research.

2.1 Introduction

This chapter starts with a review of artificial intelligence (AI) tools and mathematical probabilistic models, which are gaining increasing attention in power system studies due to high accuracy and flexibility in modelling and forecasting tasks. Then, it presents an overview of commonly used metrics for assessing both deterministic and probabilistic models. Afterwards, a review of load disaggregation and load forecasting literature is presented, followed by another literature review on wind turbine and wind farm modelling and forecasting. A detailed review of wind turbine and wind farm measurement outliers is presented separately in Chapter 5. Finally, this chapter reviews the definitions and sources of various types of uncertainties and also discusses potential approaches and solutions to some of the considered issues.

2.2 Overview of Basic Modelling and Forecasting Approaches

In the most general sense, a model is a representation of a physical or abstract system, process, procedure or phenomenon, whose main purpose is to ascertain information and understanding/knowledge, which is in practice typically manifested as an ability to accurately or confidently establish relationship between the model inputs and model outputs. Particularly in terms of measurement-based models, the process of building or establishing the model is equivalent to formulating a mapping function that may either be deterministic or probabilistic and may either be single-dimensional or multi-dimensional. Literature in different research fields (e.g., physics, statistics, computer science, etc.) may use various terminologies to represent ‘inputs’ and ‘outputs’. This thesis uses ‘input variables’, ‘independent variables’, ‘explanatory variables’ and ‘predictor variables’ interchangeably for ‘inputs’, while ‘output variables’, ‘dependent variables’, ‘outcome variables’ and ‘response variables’ are used interchangeably as ‘outputs’. This thesis considers both modelling and forecasting studies, where the main difference is assumed to lie in the synchronisation between predictor/input and response/output variables. If the model inputs and outputs are synchronised, then the

model is built or used for modelling or hindcasting purposes. However, if the model takes the observations in the past and/or the current time or state and provides the estimations for the future time or state, then it is a forecasting model.

With the above definition, the model itself may not make any difference (may not be aware of) whether it is handling modelling or forecasting problems, as that simply depends on the settings of synchronisation of the predictor and response variables by the model user. Therefore, it is possible to utilise exactly the same model for both modelling and forecasting. Accordingly, this section gives a general overview of the approaches for building, designing and specifying models that could be used for modelling or forecasting in the current literature.

2.2.1 Machine Learning, Neural Network, and Deep Learning

Machine Learning (ML) uses computer algorithms to explore and extract the useful information and relationships among the provided data. Generally, the ML contains a set of advanced models, including support vector regression (SVR) [39], tree-based model [40], Gaussian process (GP) [41], neural network (NN) [6, 42], etc. In recent years, neural networks are playing increasingly important roles in many areas, particularly when the dimensionality of the problem is too high for traditional methods. When there is a sufficient amount of labelled data (e.g., in computer vision and natural language processing tasks), neural networks are also widely applied, whose training results depend on the size and quality of the available data. . The conventional NN model in machine learning usually has only one hidden layer, and deep learning (DL) usually refers to NN model that has a cascade of multiple layers, through which the analysed data are transformed into various hidden domains, allowing the extraction of deeper features. In relatively simple problems, or if there is not enough data, deep learning may even be less accurate than a classic machine learning, but its learning ability is much stronger in complex problems with huge amounts of available data.

Multilayer perceptron

Multilayer perceptron (MLP) is a classical feedforward artificial neural network architecture consisting of at least three layers: an input layer, an output layer and a hidden layer, Fig. 2.1. Hidden layer typically uses a nonlinear activation function to

map the weighted inputs to its outputs. For extremely complex problems, the depth of the network usually increases by adding more hidden layers, but a deeper MLP is more likely to have overfitting issue. The overfitting means the model is capable of fitting exactly against the training data (even the noise inside the dataset) but cannot provide good estimations on the test set or unseen data. To resolve that, random dropout of hidden and visible neurons in a certain layer has been presented as an efficient regularisation technique [43]. Furthermore, L2 regularisation can improve accuracy by adding the squared values of weight coefficients to the loss function as penalty factors [44].

In classical MLP, layers are full-connected (also known as “dense layers”), which means the outputs of all neurons in the previous layer are connected to every neuron in its next layer. In the training process, MLP learns through changing weights of all neurons in the steepest direction that reduces the error between the network outputs and the actual outputs, which is known as gradient descent.

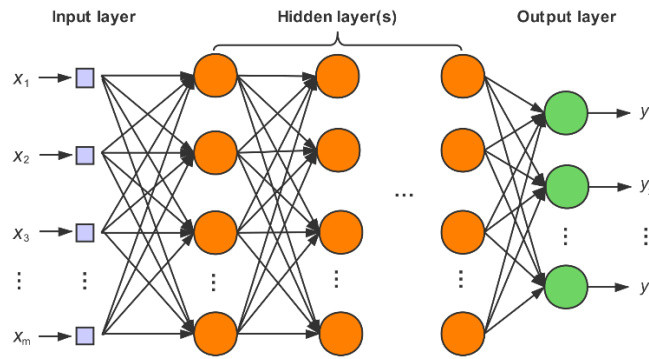


Fig. 2.1. Typical MLP architecture.

Convolutional Neural Network

Convolutional neural network (CNN) is a type of deep learning model commonly used in image processing and computer vision (CV) applications. Compared to a more general neural network model architecture, e.g., MLP, CNN is less prone to overfitting issues, as it does not use simple full-connections, but it embeds the complex data patterns into smaller and simpler filters through convolution [45]. The convolution operation is performed by a convolution matrix, i.e., kernel, which dot-multiplies the target matrix from which the features are extracted as:

$$fm(i, j) = km \cdot tm(i, j) = \sum_{\Delta i = \Delta i_{\min}}^{\Delta i_{\max}} \sum_{\Delta j = \Delta j_{\min}}^{\Delta j_{\max}} km(\Delta i, \Delta j) gm(i + \Delta i, j + \Delta j) \quad (2.1)$$

where: gm is the target matrix, km is the kernel matrix, and fm is the filtered matrix after convolution; i and j are the row and column indices of the matrix, and Δi and Δj indicate the shifting along the row and column directions. In terms of time series, the convolution operation is performed along the temporal dimension (a single dimension), which is effectively a 1D convolution, and ReLu (rectified linear unit) is commonly used as the activation function.

Recurrent Neural Network

In regression problems, another type of deep learning model specialised for data sequences is a recurrent neural network (RNN), which is widely researched and reported to achieve satisfactory accuracy in natural language processing (NLP), speech recognition, machine translation and many other applications. The RNN is designed to have an internal state, where both the previous internal state and the current predictor variables are inputted into the RNN cell to produce the current output. The recursively updated internal state makes the RNN capable of memorisation. Fig. 2.2 shows a typical architecture of a traditional RNN. It can be seen that each RNN cell uses both the previous internal state h_{t-1} and its current input x_t to predict y_t , where internal state h_t is updated and passed to the next step.

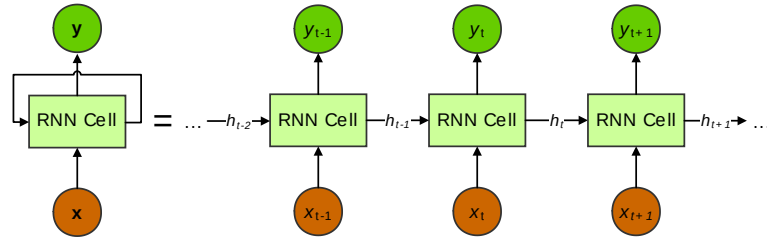


Fig. 2.2. Typical RNN architecture.

When handling long data time series, traditional RNN tends to have “gradient vanishing” and “gradient exploding” issues [46]. The training process of traditional RNN uses the Backpropagation Through Time (BPTT) algorithms to calculate the gradients and propagation of error will need to be multiplied with weights in all steps [47]. In learning long-length dependencies, if the weights are smaller than 1, then the error will be very close to zero after the backpropagation (i.e., gradient vanishing), and

if the weights are larger than 1, then the error will be too large at the end (i.e., gradient exploding).

Long Short-Term Memory

To overcome “gradient vanishing” and “gradient exploding” problems, a special type of RNN, long short-term memory (LSTM) has been developed, as it is relatively insensitive to gap lengths between the important sequence events [48]. Accordingly, every LSTM cell has three gates to control whether to remember, or to forget certain information, and whether to output it in each recursion. These three gates are trained, so that LSTM could remember important information and forget the irrelevant data, even in case of a long sequence. LSTM is also a very flexible model, as its inputs and outputs can be either a single value, or a sequence, and it is, therefore, suitable for “one-to-one”, “one-to-many”, “many-to-one” and “many-to-many” problems. A general architecture of the LSTM cell is illustrated in Fig. 2.3.

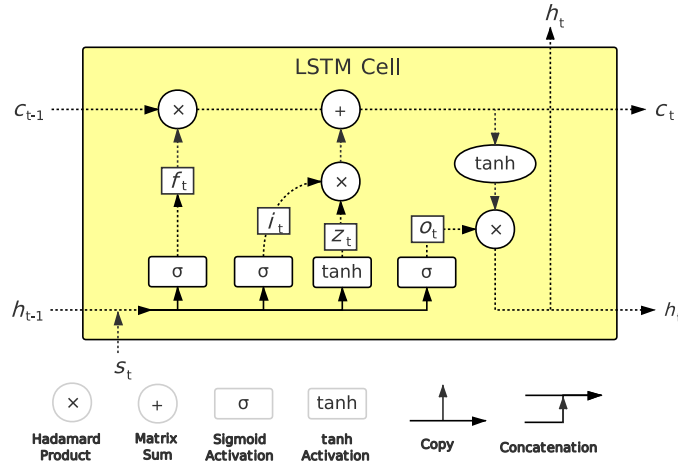


Fig. 2.3. A general LSTM cell architecture (solid arrows represent weighted vectors, while dot arrows represent unweighted vectors).

In Fig. 2.3, s_t is the current input and h_t is the hidden state (also the current output) of the LSTM cell. State c is the cell state that changes extremely slow and, in most cases, it is very similar to the previous cell state. State h stands for the hidden state, which varies a lot in the recursion and usually depends on the current input, as well as on the previous hidden state. The computation can be summarised by the equations for the three gates:

$$f_t = \text{sigmoid}(W_f \cdot [h_{t-1}, s_t] + b_f) \quad (2.2)$$

$$\begin{cases} i_t = \text{sigmoid}(W_i \cdot [h_{t-1}, s_t] + b_i) \\ z_t = \tanh(W_z \cdot [h_{t-1}, s_t] + b_z) \\ c_t = f_t \times c_{t-1} + i_t \times z_t \end{cases} \quad (2.3)$$

$$\begin{cases} o_t = \text{sigmoid}(W_o \cdot [h_{t-1}, s_t] + b_o) \\ h_t = o_t \times \tanh(c_t) \end{cases} \quad (2.4)$$

where: W_f , W_i , W_z , W_o are weights parameters and b_f , b_i , b_z and b_o are bias parameters of the LSTM cell. f_t is the forget gate with h_{t-1} and s_t as inputs and it is activated by a sigmoid activation function; the remember gate signal i is also activated by a sigmoid activation function and another candidate vector z is activated by tanh function; the output gate signal o again has a sigmoid activation and uses h_{t-1} and s_t as the inputs. More detailed discussion of LSTM cell can be found in the original paper [48]. The original LSTM architecture has only one direction, and a bidirectional structure is firstly introduced in [49], to allow the signal to travel in both forward and backward directions. This is denoted as a bidirectional LSTM (BiLSTM) method, which could further reduce the errors. The schematic of a BiLSTM layer is plotted in Fig. 2.4.

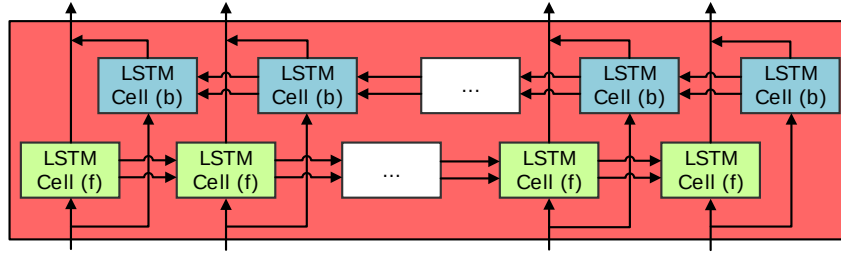


Fig. 2.4. A bidirectional long short-term memory layer: (f) - “forward” and (b) - “backward”.

The stacked bidirectional LSTM (SB-LSTM) has multiple BiLSTM layers and forms a deeper network model, which, therefore, could learn more efficiently and accurately in case of more complex and more challenging problems [50]. However, a deeper SB-LSTM is more likely to have overfitting issues or difficulties in converging during the training. Dropout or L2 regularisation may be applied to increase the ability of generalisation, like in MLP [43, 44]. The SB-LSTM architecture is depicted in Fig. 2.5.

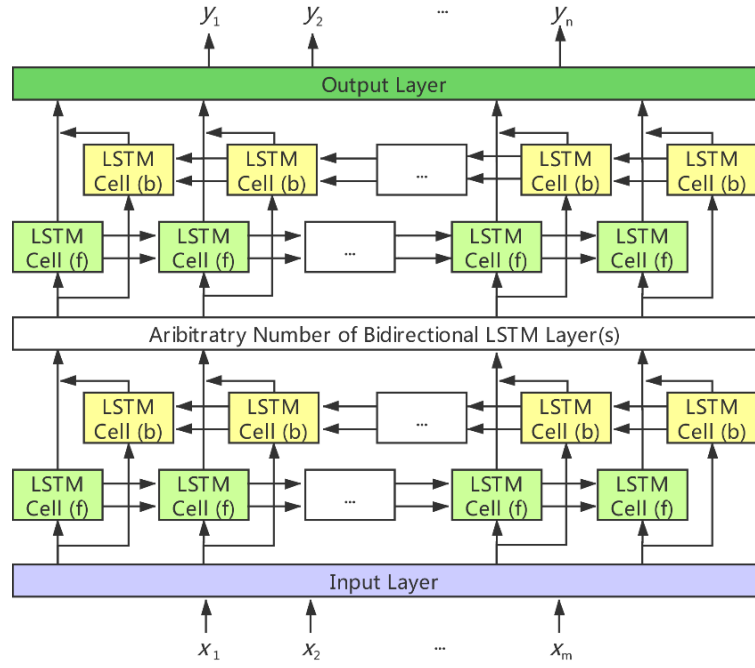


Fig. 2.5. SB-LSTM architecture: (f) - “forward” and (b) - “backward”.

2.2.2 Copula Methodology

Copula is a multivariate probability distribution function with uniformly distributed marginal probability for each variable [51]. Applications of Copula are attracting increasing attention in power systems and renewable energy research, such as load profile clustering [52], forecasting of wind-based generation systems [53-55] and, especially, modelling of uncertainties and correlations [56-59]. Copula theory separates the analysis of univariate marginal distributions and their correlations. According to Sklar's theory [51], the joint distribution of multiple variables can be expressed by marginal distributions and a Copula function that describes the dependency structure among the variables. For known marginal distributions $F_1, F_2, \dots, F_d$, their joint distribution $F_{1,2,\dots,d}$ can be written as $F_{1,2,\dots,d} = C_{1,2,\dots,d}(F_1, F_2, \dots, F_d)$, where $C(\cdot)$ is the Copula distribution function. If $F_1, F_2, \dots, F_d$ are continuous, then $C(\cdot)$ is unique and for a d -dimensional variable structure:

$$C_{1,2,\dots,d}(F_1, F_2, \dots, F_d) = F_{1,2,\dots,d}(x_1, x_2, \dots, x_d) \quad (2.5)$$

$$\begin{aligned} & c_{1,2,\dots,d}(F_1, F_2, \dots, F_d) \\ &= \frac{\partial^d C_{1,2,\dots,d}(F_1, F_2, \dots, F_d)}{\partial F_1 \partial F_2 \dots \partial F_d} \end{aligned}$$

$$\begin{aligned}
&= \frac{\partial^d F_{1,2,\dots,d}(x_1, x_2, \dots, x_d)}{\partial x_1 \partial x_2 \dots \partial x_d} \times \prod_{i=1}^d \frac{\partial x_i}{\partial F_i} \\
&= \frac{f_{1,2,\dots,d}(x_1, x_2, \dots, x_d)}{\prod_{i=1}^d f_i(x_i)} \tag{2.6}
\end{aligned}$$

where: $C_{1,2,\dots,d}(\cdot)$ and $c_{1,2,\dots,d}(\cdot)$ are Copula distribution and density functions; $F_i(\cdot)$ and $f_i(\cdot)$ are marginal distribution and density functions of x_i and $F_i(\cdot) \sim U(0, 1)$; $F_{1,2,\dots,d}(\cdot)$ and $f_{1,2,\dots,d}(\cdot)$ are joint distribution and density functions; f_i and F_i stand for $f_i(x_i)$ and $F_i(x_i)$, respectively.

The main idea of Copula is to firstly transform x_i into $U(0, 1)$ domain and then capture their correlations using a Copula function. Different choices of Copula functions have various desired dependence patterns and properties. Widely-used Copula families are elliptical Copulas (e.g., Gaussian Copula, and the t -Copula) and Archimedean Copulas (e.g., Clayton Copula, Frank Copula, and Gumbel Copula).

Although a single simple Copula may model arbitrary high-dimensional correlations, it will be difficult to implement/converge in practice given the highly nonlinear or even multi-modal correlations. Vine Copula [52, 57, 60] is particularly effective in modelling high-dimensional dependencies, as it divides the full correlation structure into several pair dependencies, for which the joint density function can be expressed recursively as:

$$\begin{aligned}
&f_{1,2,\dots,d}(x_1, x_2, \dots, x_d) \\
&= f_1(x_1) f_{2|1}(x_2|x_1) \dots f_{d|1,2,\dots,d-1}(x_d|x_1, x_2, \dots, x_{d-1}) \\
&= \prod_{t=2}^d f_{t|1,2,\dots,t-1}(x_t|x_1, x_2, \dots, x_{t-1}) f_1(x_1) \tag{2.7}
\end{aligned}$$

$$\begin{aligned}
&f_{t|1,2,\dots,t-1}(x_t|x_1, x_2, \dots, x_{t-1}) \\
&= c_{1,t|2,\dots,t-1}(F_{1|2,\dots,t-1}, F_{t|2,\dots,t-1}) \times f(x_t|x_2, x_3, \dots, x_{t-1}) \\
&= \dots \\
&= \left[\prod_{s=1}^{t-2} c_{s,t|s+1,s+2,\dots,t-1}(F_{s|s+1,s+2,\dots,t-1}, F_{t|s+1,s+2,\dots,t-1}) \right] \times \\
&c_{t-1,t}(F_{t-1}, F_t) \times f_t(x_t) \tag{2.8}
\end{aligned}$$

Therefore, according to Equation (2.7) and Equation (2.8), one possible rearrangement for $f_{1,2,\dots,d}$ calculation is:

$$\begin{aligned}
& f_{1,2,\dots,d}(x_1, x_2, \dots, x_d) \\
&= \left[\prod_{t=2}^d \prod_{s=1}^{t-2} c_{s,t|s+1,s+2,\dots,t-1}(F_{s|s+1,s+2,\dots,t-1}, F_{t|s+1,s+2,\dots,t-1}) \right] \times \\
& \left[\prod_{t=2}^d c_{t-1,t} \right] \times \left[\prod_{k=1}^d f_k(x_k) \right] \\
&= \left[\prod_{i=1}^{n-1} \prod_{j=1}^{n-i} c_{j,j+i|j+1,j+2,\dots,j+i-1}(F_{j|j+1,j+2,\dots,j+i-1}, F_{j+i|j+1,j+2,\dots,j+i-1}) \right] \times \\
& \left[\prod_{k=1}^n f_k(x_k) \right] \tag{2.9}
\end{aligned}$$

where: $c_{j,j+i|j+1,j+2,\dots,j+i-1}(F_{j|j+1,j+2,\dots,j+i-1}, F_{j+i|j+1,j+2,\dots,j+i-1})$ is conditional Copula describing correlation between $F_{j|j+1,j+2,\dots,j+i-1}$ and $F_{j+i|j+1,j+2,\dots,j+i-1}$. Generally:

$$F_{j|D}(x_j|\mathbf{x}_D) = \frac{\partial C_{j,v}(F_{j|D-v}, F_{v|D-v})}{\partial F_{v|D-v}} \tag{2.10}$$

Note that $x_v \cap x_{D-v} = \emptyset$ and $x_v \cup x_{D-v} = x_D$. More specifically, the correlation structure by Equation (2.9) is known as D-Vine Copula. For example, when $n = 3$, Equation (2.9) will yield pair dependence $c_{1,2}$, $c_{2,3}$, and $c_{1,3|2}$, which can be represented by the tree structure in Fig. 2.6.

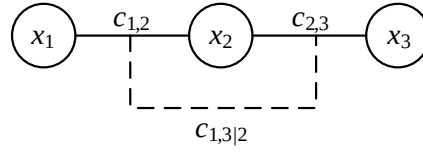


Fig. 2.6. D-Vine structure of three variables.

Alternatively, $f_{1,2,\dots,d}$ can be derived using C-Vine structure [52], whose analytical expression is:

$$\begin{aligned}
& f_{1,2,\dots,d}(x_1, x_2, \dots, x_d) \\
&= \left[\prod_{i=1}^{n-1} \prod_{j=1}^{n-i} c_{i,i+j|1,2,\dots,i-1}(F_{i|1,2,\dots,i-1}, F_{i+j|1,2,\dots,i-1}) \right] \times \left[\prod_{k=1}^n f_k(x_k) \right] \tag{2.11}
\end{aligned}$$

When $n = 3$, its tree structure is shown in Fig. 2.7.

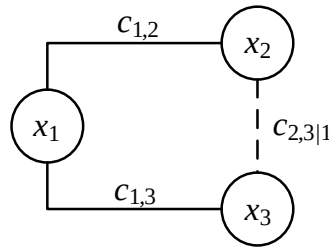


Fig. 2.7. C-Vine structure of three variables.

Compared to D-Vine, C-Vine is usually more suitable for the cases where there exists a centre variable of high importance. In D-Vine, all variables share similar attentions from the model. It also should be noted that there are also more complex Vine structures in the current literature, e.g., R-Vine [53].

2.3 Commonly Applied Error Metrics

This section reviews some commonly applied deterministic and probabilistic error metrics for assessing model accuracy, which are also used in evaluating performance of different models developed in this thesis. In addition, some other metrics proposed in this thesis are illustrated in later chapters (e.g., energy-based error metrics in Section 3.4.4 and bin-related error metrics in Section 7.2.4).

2.3.1 Deterministic Error Metrics

The three most widely used error metrics are: root-mean-square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

The RMSE is expressed as the square root of the average squared errors of estimation, calculated as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2.12)$$

where: y_i is the actual observation at i , and $\hat{y}_i$ is its estimation by the model, assuming there are n samples to test in total. The RMSE is more sensitive to large errors, because all errors are squared before taking the average and therefore RMSE is more suitable when large errors are not present.

The MAE gives the same weight to all errors, and it is expressed as the average of the absolute of estimation errors, calculated by:

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2.13)$$

where the symbols have the same meaning as in RMSE. Although the MAE is more easily to understand, because it describes average errors alone, the use of absolute value is undesirable in many mathematical calculations. For example, it would be

difficult to calculate the gradient or sensitivity of the MAEs with respect to certain model parameters.

The MAPE is simple and it is formulated as:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2.14)$$

where the symbols have the same meanings as in RMSE and MAE. In practical application, however, there are several issues with MAPE. For example, it cannot be used if there exist $y_i = 0$, or it can become “unstable” and significantly exceed 100 % if there are very small $|y_i|$, but relatively large $|y_i - \hat{y}_i|$.

2.3.2 Probabilistic Error Metrics

For evaluating probabilistic errors, pinball loss [10], Winkler score [10] and continuous ranked probability score (CRPS) [61] are also relatively widely used in the existing literature.

Pinball loss is designed to evaluate the overall performance of a probabilistic forecasting model for a set of quantiles, and it is calculated as:

$$\text{Pinball Loss} = \begin{cases} (y_i - \hat{y}_{i,q})q & \hat{y}_{i,q} < y_i \\ (\hat{y}_{i,q} - y_i)(1 - q) & y_i \geq \hat{y}_{i,q} \end{cases} \quad (2.15)$$

where: y_i is the actual observation at i , and $\hat{y}_{i,q}$ is its q -th quantile estimation by the model. Usually, q is an array covering range 0 to 1 and with a step equal 0.1, 0.05 or 0.01. Pinball loss is closely related to MAE, and when $q = 0.5$, its calculation is identical to MAE, but pinball loss assigns different weights (penalty factors) to other quantile values. For example, when $q > 0.5$, the model underestimation ($\hat{y}_{i,q} < y_i$) will have stronger impacts on the pinball loss and vice versa for the case when $q < 0.5$.

Winkler score is a metric focusing on prediction intervals instead on a few quantiles and is formulated as:

$$\text{Winkler Score} = \begin{cases} 2(\hat{y}_{i,\min} - y_i)/\alpha + (\hat{y}_{i,\max} - \hat{y}_{i,\min}) & y_i < \hat{y}_{i,\min} \\ 2(y_i - \hat{y}_{i,\max})/\alpha + (\hat{y}_{i,\max} - \hat{y}_{i,\min}) & y_i > \hat{y}_{i,\max} \\ (\hat{y}_{i,\max} - \hat{y}_{i,\min}) & \text{Otherwise} \end{cases} \quad (2.16)$$

where: α is the significance level (usually 0.1 or 0.05) and $\widehat{y_{t;\min}}$ and $\widehat{y_{t;\max}}$ are the minimum and maximum values provided by the probabilistic forecasting model at α , which forms the prediction interval. For the actual realised recording observed within the prediction interval, the Winkler score is simply the length of the interval; otherwise, it will add a penalty proportional to the distance of the observation outside the interval.

The CRPS assess the probabilistic model by considering the distribution as a whole and it can be calculated as:

$$\text{CRPS} = \int_{-\infty}^{y_i} F(z)^2 dz + \int_{y_i}^{\infty} (F(z) - 1)^2 dz \quad (2.17)$$

where: $F(z)$ is the cumulative distribution function (CDF) estimated by the probabilistic forecasting model. This equation consists of two integrals on y_i as the boundary, which in practice can be estimated by Monte Carlo. It can be interpreted as the quadratic measure of the difference between forecasted CDF and the empirical CDF of the observation.

2.4 Load Modelling and Forecasting Review

This PhD started the load research with load forecasting, which was based on and an extension of the internship as a big data analyser at China Electric Power Research Institute, from June to July 2018. The related work and outcome will be discussed in more detail in Chapter 3. The load modelling, more specifically, load disaggregation (a hindcasting problem) is researched in a later stage and is presented in Chapter 4. Therefore, in this chapter, the load forecasting literature is reviewed first, followed by the review of load disaggregation.

2.4.1 Load Forecasting

Load forecasting is generally aimed at providing an accurate estimation of energy consumption based on historical recordings and other relevant data. It is becoming increasingly important for the operation and planning of existing electricity network and future smart grids, which will feature much higher ranges of uncertainties, larger variations of power flows and increased levels of interactions between supply and demand sides. Long-term forecasting is important for power system planning, while

short-term forecasting, especially day-ahead forecasting, is essential for operational studies [6].

There are many load forecasting approaches in existing literature, including time series analysis [62], Gaussian process [41], neural network [6, 42], fuzzy regression [63], tree-based model [40], support vector regression [39], etc. Typical load profiles exhibit periodicity, but also feature strong hourly, daily, weekly and monthly/seasonal variations. Weather (meteorological) factors, such as temperature, precipitation, solar irradiation, etc., as well as socio-behavioural factors, all have impact on variations of demands. Inclusion of these factors improves forecasting accuracy, but also increases the dimension of the forecasting problem, as several synchronous series (i.e., calendar and weather variables) should be analysed together with demand series (“the curse of dimensionality”).

Additionally, many current approaches lack flexibilities in terms of performing forecasting over different period lengths (i.e., day-ahead forecasting vs week-ahead forecasting). Regardless of the technique used to model load profiles and provide predictions, model selection is an important step, which relies on the understanding of the inherent features and characteristics of different models.

2.4.2 Load Disaggregation

Due to the recent deployment of demand-side management and demand-responsive load control schemes, load disaggregation is becoming increasingly important for balancing renewable generation and provision of various system support services [7, 8]. Two additional drivers for a significant increase of interest in load disaggregation are the much higher availability of demand datasets from advanced and intelligent monitors, e.g., smart meters, as well as the development of efficient neural/recurrent network and deep learning modelling methods. Load disaggregation denotes a general process in which specific load types or load categories are identified in the measured aggregated (i.e., total) electricity demands. Typically, load types and categories of interest for disaggregation are appliances, or groups of appliances that share similar electrical and/or consumption characteristics, e.g., demand-manageable loads, such as heating or cooling loads.

In cases where disaggregation is performed on total demands measured by a single electricity meter (i.e., for a single customer), the common term is “nonintrusive appliance/load monitoring” (NIALM/NILM), as opposed to the monitoring of individual appliances/loads by separate meters, which is denoted as “intrusive appliance/load monitoring” (IALM/ILM) [64, 65]. Load disaggregation is readily available for ILM, as the loads of interest are directly monitored, while NILM-based load disaggregation faces several challenges. The main idea behind NILM is that different load types will have different energy consumption “signatures” or “patterns”, which will then allow for their identification in the total demand, assuming measured data are available with a sufficient resolution. However, the energy consumption signatures/patterns may strongly vary, depending on the appliance-specific characteristics, weather conditions, user settings and habits, or wider socio-economic factors [9], making the identification of the distinctive load types difficult. A significantly more challenging problem is load disaggregation at a distribution network substation level, where available measurements are total demands of a large number of customers, whose individual smart meter data may not be accessible, e.g., due to data privacy/ownership concerns.

Direct load disaggregation approaches aim at detecting specific changes in total demand that can be allocated to the use of individual appliances, e.g. their on/off switching [66]. Various clustering and classification algorithms are widely applied to identify unique load signatures in demand time series, such as support vector machine [67], decision trees [68] and combinatorial optimisation (CO) [69]. If the sufficiently labelled datasets for both total demands and individual load type demands are available, as in ILM-based approaches, hidden Markov models (HMM) and their variants (e.g., factorial hidden Markov model, FHMM) can be used to explore patterns of individual appliances and their combinations [70, 71]. Frequency-based analysis methods, including Fourier [72] and wavelet [73] transforms, are also used to capture on/off switching events, or to identify load types with specific harmonic emission signatures. Usually, these methods are inaccurate when there is high noise in the measured data and their complexity and computational times significantly increase when the numbers of appliances and/or their operating states increase.

As mentioned, deep learning models are being increasingly used in many areas and in various applications, including load disaggregation [9, 64, 74, 75], where improvement in accuracy and performance for identifying the underlying load profiles and demand patterns have been reported. However, these approaches are often criticised as “black box” methods, because they are difficult to interpret and their results are sensitive to the settings of hyperparameters [10]. Accordingly, the majority of the deep learning models in current literature are simply built and tuned using the available data, without performing further exploration or analysis of the modelled processes. For example, the accuracy of the built models will increase if domain expert knowledge is used to analyse and extract specific characteristics from available data. This is usually denoted as “feature engineering”, as the extracted features may be essential for understanding the modelled process, and as the selection of the most representative data properties may result in performance improvement. Ref. [76] presents a review of features for appliance classification in load disaggregation (low and high frequency, steady-state and transient categories), showing that systematic feature selection can improve accuracy. Ref. [77] considers and compares electrical and statistical features for load disaggregation, confirming the importance of selecting both appropriate features and algorithms.

2.5 Wind Turbine and Wind Farm Modelling and Forecasting Review

This PhD also put great efforts into wind energy modelling (Chapter 6 and Chapter 7) and forecasting (Chapter 8 and Chapter 9), which leads to various important innovative WT-level and WF-level modelling and prediction approaches. This section presents a review of the related methods in the current literature. In addition, because it is important to have accurate information on their operational characteristics, which are typically obtained by processing large amounts of measurements from the individual WTs and the whole WFs, it is also crucial to identify and remove abnormal data before implementing the models. Therefore, a comprehensive WT and WF-levels data quality literature review is presented in a separate chapter, Chapter 5.

2.5.1 Overview of Wind Turbine Topologies and Control

Over the recent years, there was a significant increase in numbers and sizes of wind-based electricity generation systems, which are expected to grow even further in the future. There are four main electrical designs and topologies for the wind turbine (WT), which are Type 1, Type 2, Type 3, and Type 4. Type 1 WT is simple but robust, where there is an asynchronous generator directly connecting the turbine and the grid. The asynchronous generator rotates at a speed determined by the grid frequency, which does not need to be controlled. In addition, capacitor banks are required in this design, because the asynchronous generator consumes the reactive power and the capacitor banks can compensate for the consumption and stable the voltage. Type 2 topology extends the Type 1 design by installing a rotor variable resistance. And the rotor speed can be changed by adjusting the variable resistance. However, some heat loss is generated in the resistance.

Type 3 WT is more controllable than Type 2 WT. The stator of the generator is directly connected to the grid, and it connects the rotor through a back-to-back converter instead of a variable resistance, which reduces heat loss and the speed can be controlled in a much wider range. And the converter can also control the active and reactive power flows, thus there is no need for capacitor banks anymore. This type of WT is also known as doubly-fed induction generator (DFIG), because there are two separately connected three-phase windings, and the stator windings are connected to the grid and the rotor windings are connected to the converter. Since there is only a partial converter used, the power passing through the converter is not maximised, and the converter size can be smaller which means lower capital cost, despite that the maintenance cost may be more expensive.

In Type 4 topology, a full back-to-back power converter is utilised to connect the generator fully to the grid and transmit all the power. Therefore, the maximum rated power of Type 4 WT is scalable and is higher than Type 3 WT. Although the converter size is larger and the cost of power electronics is increased, one advantage of Type 4 WT is that it is isolated from the grid, which can be very useful in the case that the grid side has disturbances.

The purpose of controlling the wind turbine is to optimise its power output. When the wind speed is low, maximum power point tracking is implemented to maximise the power produced by the wind turbine. And if the wind speed increases, the control will limit the power output to a rated value and may even shut down the WT for protection. The control of the WT also ensures it meets compliance with grid connection requirements, including control of the active and reactive power (frequency and voltage control), fault-ride-through and power quality (e.g., harmonic emission and flicker). In normal operating conditions, the yaw angle of the WT is controlled to point the WT up to the wind so the mechanical load is reduced and the extracted power is maximised. The WT is also controlled by either stall control or pitch control. The blade of the stall-controlled WT has a fixed angle which relies on the aerodynamic design that allows the WT can still run at a relatively constant speed when the wind speed is high and limit the power generation. In the pitch-controlled WT, the controller monitors the power output of the wind turbine and adjusts the blade pitch angles to increase or decrease the power output.

2.5.2 Wind Turbine and Wind Farm Modelling

The most common approach in assessing outputs of wind-based electricity generation systems is to correlate wind turbine power output with only one parameter: input WS. A widely-used examples are power curves (PCs) provided by WT manufacturers, which are obtained under predefined standard conditions and therefore cannot correctly represent many site-specific and application-specific factors, such as the wake or terrain effects, or WT dynamics due to wind speed (WS) and/or wind direction (WD) variations [12, 13]. In deterministic power curve modelling, there are two main types of methods. One is parametric model, such as linearised segmented model, polynomial model, and logistic functions [78], while another is nonparametric model, such as neural network, random forest, support vector machine (SVM), and k -nearest neighbour (k -NN) [79]. The nonparametric model does not explicitly provide the mathematical expression between wind speed and power output; therefore, these methods (especially NN-based) may lose insight into the problem [16].

In the parametric deterministic power curve modelling, the five-parameter logistic function (5PLF) usually has higher accuracy compared to other models [80-82].

Specifically, the reduction of errors using 5PLF over four-parameter logistic function (4PLF) is significant, as the output of 4PLF is symmetric curve, which is seldom the case in real applications, while 5PLF could address asymmetric representations [83]. In addition, power curves using six-parameter logistic function (6PLF) could be found in [78, 84]. However, the results show that their errors are almost similar to those based on 5PLF, but 6PLF needs one additional parameter. The equation of 5PLF model is:

$$5PLF(x) = d + (a - d) \times \left(1 + \left(\frac{x}{c}\right)^b\right)^{-g} \quad (2.18)$$

where: x is the independent variable, a and d represent two asymptotes of 5PLF, $c > 0$ is the inflection point, b is the hill slope and $g > 0$ is asymmetry factor (if $g = 1$, 5PLF is reduced to 4PLF, which is symmetrical around the inflection point). More detailed discussions and comparisons about the properties of these five parameters could be found in [83].

As illustrated above, the conventional power curve model only considers WS as the predictor variable, and the impact of WD is often neglected in current studies, [14, 15, 80, 85], assuming that, for the WS between the cut-in and cut-out speed, yawing control will orient WT so that the WD (“attacking angle”) is perpendicular to the rotor blades plane, [86, 87]. However, yawing mechanisms are relatively slow (e.g., 0.5 °/s in [87]), so frequent WD changes typically result in a reduced P_{out} of a WT. Furthermore, the conventional WT models do not provide information on the impact of WD and although [88] and [89] demonstrated that WT model can be improved by considering both WS and WD, this work mostly concentrated on assessing impact of stochastic WS and WD variations, not on correlating P_{out} uncertainties for specific WS and WD data (e.g., due to wake or terrain effects in a WF).

Compared to the modelling of individual WT, the WF modelling is more complex. However, a WF is essentially an aggregation of multiple WTs, and not all of them always have the same operating condition simultaneously. Therefore, the power output of WF will be further impacted by the WF operating regimes (ORs), i.e., by the operational states of individual WTs in the WF. However, the WF ORs are mostly neglected in wind energy research. To the best of the PhD student’s knowledge, the

only attempt of modelling WF ORs could be found in [17], where wind power fluctuation is taken as a reference to compute regimes by observable or non-observable processes. In Chapter 9, this PhD thesis will use two general cases to assess the influence of WF ORs on the WF P_{out} forecasting: planned, and therefore known outages, and unplanned, and therefore forecasted outages.

2.5.3 Wind Turbine and Wind Farm Forecasting

Wind energy is a “clean” and renewable energy resource, which however exhibits strong and stochastic spatio-temporal variations, therefore resulting in uncertain and difficult to predict outputs of WT and WF. The assessment and quantification of uncertainties in power outputs of WT and WF are important not only for wind farm owners, but also for system operators, due to their impact on the variations of network power flows and related uncertainties in network operating conditions [56, 90].

Forecasting of WF P_{out} should correctly account for stochastic spatio-temporal variations of wind energy resource, which in turn depend on a number of weather variables. For example, Chapter 8 and [16] show that besides uncertainties in WS, variations in AD and WD may also have strong impact on the WF P_{out} , which will be further impacted by the WF ORs, i.e., by the operational states of individual WTs in the WF.

The conventional forecasting techniques are deterministic, outputting either point values, or single sequences, depending on the prediction horizons [91, 92]. The difference or error between the predicted deterministic values and realised future values may have strong impact on unit commitment, operating reserve setting and operation of electricity market [93-95].

Probabilistic forecasting is increasingly used in recent years, since it can provide information on not only the deterministic average/mean (or most likely/expected) values, but also on the uncertainties of deterministic forecasts. There are two general types of probabilistic forecasting models: those that provide explicit distributions to represent uncertainties, where Normal, Beta and Logit-Normal distributions are the most used [61, 96, 97], and those that provide non-parametric models, where quantile regression is commonly applied [98, 99]. Markov chain (MC) models are also widely

used for probabilistic forecasting. In [100], a support vector machine-enhanced MC model for a short-term wind power forecasting is developed using only historical data of wind power outputs, without considering exogenous (weather) variables. In [101], relations between the wind, PV power output and heat index are modelled by a multivariate MC model to forecast 15-min ahead PV and wind power outputs, without explicitly forecasting WS, AD and WD.

The ensemble learning and hybrid-models are also actively researched, since a single model may not be adequate in all scenarios, so combining of multiple models may improve the overall forecasting performance [61, 96, 102]. Furthermore, the dynamics and non-stationarity of the meteorological variables, especially WS, which has the greatest impact on WF power outputs, make forecasting by using single specific distribution difficult: the types of the WTs installed in the WF, the impact of the terrain, wake effects and aging effectively require to use multiple distributions to cover all scenarios. For example, the outputs of the two classes of models in this thesis (see Chapter 9), i.e., the NN-based forecasting model and Copula-based correlational model, are assumed to be mixtures of distributions.

2.6 Review of Classification of Uncertainty

One important aspect of mathematical modelling, which is particularly important for modelling stochastic/random processes by statistical or probabilistic models, is the evaluation of factors contributing to the uncertainties in the model outputs [103]. Some of these are briefly reviewed below, with examples in wind farm power output forecasting for illustration purposes:

- ***Parameter uncertainty:*** The selected or assumed model inputs (predictor variables, or model parameters/coefficients) cannot exactly describe the considered process, or cannot be inferred by the applied model, or are simply unknown. An example in the analysis presented in this PhD thesis is uncertainty in input WS, WD and AD predictor variables for modelling WF-level P_{out} , as the considered WFs consist of a number of individual WTs, each with generally different WT-level values of WS, WD and AD. Therefore, the observed input WF-level predictor variables are inherently uncertain representations of a more

complex “wind energy conversion process” performed by the individual WTs. Another example is the assumed uncertainty of the internal parameterisation of the implemented model. For example, in a simple linear model, $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, the weights β and bias ε are also denoted as the model “parameters”, e.g., in computer science research [45], while they are often called model “coefficients” in statistics [104], or quantitative finance [105]. In this thesis (see Chapter 9 for Bayesian neural network), the true values of weights parameters and bias parameters of the NN models cannot be inferred by limited training samples, so they are represented non-deterministically, by assumed or calculated distributions

- **Parametric variability:** The selected predictor variables of the model exhibit stochastic variations, or their values cannot be accurately predicted for evaluating future states of the modelled process. This is illustrated in the thesis by considering ranges of forecasted input WS, WD and AD values, i.e., their distributions, as the model inputs to forecast the values and uncertainties of the WF power outputs.
- **Structural uncertainty:** It is also known as model inadequacy, which is due to the lack of a full understanding of the underlying physics. For example, in the field measurements of an operational WF, consisting of only wind speed and wind farm power output, the same input WS values usually do not map to the same P_{out} values, especially if their relationship is nonlinear. A simple WS to P_{out} model based on this dataset may be biased or inaccurate, but the accuracy can be improved, and the uncertainty can be reduced if a higher dimensional model with air density and wind direction as additional predictor variables is applied. The challenges of adding AD and WD into the model include 1) not all datasets have those recordings; 2) some existing widely used P_{out} -WS models are inherently difficult to be extended (e.g., linearised segmented model, polynomial model, and logistic functions reviewed in Section 2.5.2). In this thesis, AD and WD values are either recorded in the WF dataset or obtained from an external data source for the same geographical location. Copula theory and neural networks are applied. Since they are naturally capable of modelling multi-dimensional process, they can be feasibly extended to analyse higher-dimensional data.
- **Algorithm/Code uncertainty:** The computational solver or optimiser, built as the algorithmic/coded implementation of the mathematical model, has numeric

limitations, or cannot take into account all states/values of predictor variables. For example, this is reflected in the presented analysis through the implementation of numerical integration, instead of analytical solutions, for the unknown or intractable Copula distribution functions in the Vine Copula model. Another example is about the uncertainty in the hyperparameters. The most appropriate model structure is *a priori* (or generally) not known. For example, this thesis involves deep learning LSTM-based model, Markov chain model, Copula-based model and some of their combinations in both homostructure and heterostructure modelling approaches, but makes no further conclusions about their suitability, other than through comparison of their performance by several evaluation metrics. Furthermore, even if only a single homostructure model is used, settings of its hyperparameters, e.g., the discretisation step in Markov chain model, or the number of neurons/learning rates in LSTM model, will impact changes, sometimes strongly, in the forecasted outputs for the same predictor variables.

- **Observation/Measurement uncertainty:** This is caused by the inevitable limited accuracy of measurements and other experimental means of capturing/measuring physical quantities. In many applications, these uncertainties are assumed to be noise.
- **Interpolation uncertainty:** The available observations and measurements do not include all process states with the required resolution/step, or there are missing values. This is illustrated in the thesis by cross-correlational analysis of geographically close WFs, to extend the available datasets for the considered individual WF.

One intuitive way to deal with the above uncertainties is to assume continuous or discrete distributions for all model parameters, rather than using their single deterministic values. In that way, model parameter uncertainties can be dealt with by, first, assigning prior distributions to model parameters, and then posterior distributions could be inferred or updated through Bayesian analysis, e.g., by commonly used Markov chain Monte Carlo (MCMC) and variational inference (VI) approaches. The parametric variability could be approached by integration over the input dimensions, which could be numerically estimated through suitable sampling. Structural uncertainty can be reduced by exploring the latent variables and adding other

correlated dimensions of measurements into the model. Algorithm/code uncertainty can be reduced by using more precise data structures or fine-tuning techniques, but it often comes at a cost of longer training times and higher memory requirements. The uncertainty caused by hyperparameter settings can be resolved at least to some extent by applying ensemble learning, or hyperparameter optimisation techniques. Observation/measurement uncertainty is unavoidable, and in many applications, they are assumed to follow Gaussian distribution. Finally, interpolation uncertainty in terms of the missing values could be resolved by imputing missing data in some way, e.g. probabilistically, or through auto-regression, or by cross-correlation, before the data are being used for the analysis, which essentially leads to multiple training datasets.

2.7 Discussions and Conclusions

This chapter firstly reviews two important modelling approaches: neural network-based models and Copula correlation models. They have satisfactory performance in analysing time series data and handling complex dependency among different variables. Then, the widely used deterministic and probabilistic error metrics are formulated and compared in terms of handling small and large errors, as well as reflecting various statistical properties. Afterwards, the literature review on modelling and forecasting of load, wind turbine and wind farm is presented.

Although there is an enormous number of models in load modelling and forecasting areas, there is still a lack of a flexible model, which can provide high accuracy and good interpretability. In the wind turbine and wind farm modelling and forecasting literature, many previous models neglect the impacts of wind direction and temperature (air density) on power output, and there is almost no research on including wind farm operating regimes in either hindcasting or forecasting models. Finally, this chapter reviews some causes and examples of various classifications of uncertainties. It also discusses the potential methods to handle these uncertainties, which are mainly related to use of probabilistic models, sampling, inference, and optimisation techniques.

Chapter 3

Weather Based Day-Ahead and Week-Ahead Load Modelling and Forecasting

3.1 Introduction

The main purpose of load forecasting is to provide an accurate estimation of the future electricity demands, which is becoming increasingly important for the operation and planning of existing electricity network and future smart grids. The load values feature short and long periodicity and seasonality, and their variations are also contributed by calendar and socio-behavioural aspects as well as meteorological variables. This chapter firstly introduces three datasets from household, substation and city levels, then it investigates the load variations using Markov chain (MC) methods. Afterwards, it implements and compares three models for day-ahead total load forecasting. Finally, the model with the best performance is extended to week-ahead total load forecasting and multi-dimensional total and heating load forecasting.

3.2 Description of Available Datasets

This chapter uses three datasets to illustrate obtained results on the load profile variations and load forecasting, which are discussed first.

Scotland Medium Voltage (MV) Substation Dataset

This dataset is from a medium voltage (MV) substation near Glasgow, Scotland (mainly residential), for which six-year demand and weather data recordings are available [106, 107]. Demand measurements are average 30-minute values of active power (AP), while meteorological data (MD) are: average 60-minute recordings of temperature (TEMP), solar irradiation (SI), precipitation (PPT) and wind speed (WS). Since temporal resolutions of load and meteorological data are different, the meteorological series are all up-sampled to 48 values per day using the spline interpolation method, in order to make them synchronous with demand data [108].

Demand recordings have strong autocorrelations and temporal dependencies, where actual variations of considered electrical load are influenced by standard calendar variables (SCV), and this dataset contains year, month, day of the month, hour and day

of the week (YY, MM, DD, HH, DW) and indicators of public holidays (HLD) and daylight saving time (DST) [109, 110]. In addition, solar analemma (Sun azimuth angle, SA, and Sun elevation angle, SE) can be used to precisely represent time of the day and season of the year variations, and they are calculated as in [111]. Finally, the Moon also has impacts on earth and human behaviours [112], and Moon phase (MP) is added into the dataset as the last additional variable following [113].

China City Dataset

This dataset contains five-year measurements of average 15-minute total active power demands from an anonymous city in the south of China and synchronous daily minimum ($TEMP_{min}$) and maximum ($TEMP_{max}$) temperature recordings in the same area. The timestamp is also provided, so the calendar variables can be extracted. SA, SE and MP are calculated similarly as in the Scotland MV substation dataset. Because of different temporal resolutions from load data, temperature data are firstly up-sampled to 96 values per day using spline interpolation method [108], assuming that $TEMP_{min}$ and $TEMP_{max}$ occur around 5 am and 2 pm, respectively [114].

The Almanac of Minutely Power Dataset (Version 2) [115]

The Almanac of Minutely Power dataset Version 2 (AMPds2) comes from one household located in Burnaby, Canada. It has active power and reactive power measurements for the whole household and active power recordings of the heat pump from 1 April 2012 to 31 March 2014. Calendar variables are also known since the timestamp is given. The original temporal resolution is 1 minute, and they are aggregate into 30 minutes resolution by averaging. Temperature data in the same area and for the same period are obtained from an external data source, The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) [116].

This chapter firstly uses the Scotland MV substation dataset to explore the typical load profiles and the variations at the same time on different days. Afterwards, it analyses three state-of-the-art methods for day-ahead load forecasting and compares their performance on the Scotland MV substation dataset and the China city dataset. Then, it selects the method with the best performance and extends it as a week-ahead load forecasting model and tests it on the Scotland MV substation dataset. Finally, this

chapter gives an example case study on AMPds2 to show that the proposed model can also forecast both the total load and the individual appliance load (heating) simultaneously.

3.3 Time Series and Variation Analysis for Load Profiles

Typical daily load profiles of Scotland MV substation dataset are presented in Fig. 3.1, showing calculated average daily load profiles of the same day types among the six years. It shows the difference between weekday, weekend and holiday, suggesting the importance of considering the day types in the model.

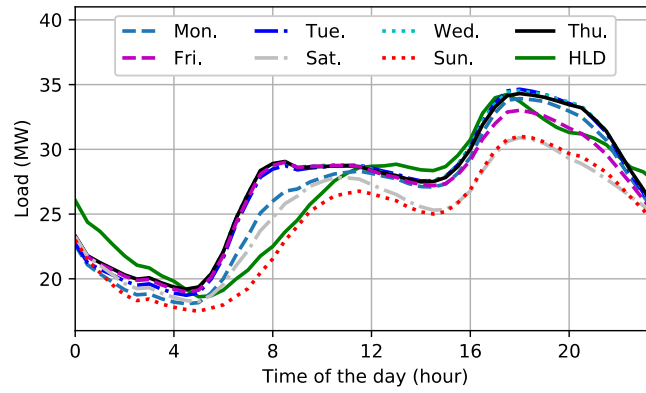


Fig. 3.1. Average daily load profiles of different types of days (Scotland MV substation dataset).

Two methods of assessing the load profiles variation ranges are compared: sliding window approach and second-order Markov chain analysis. For both of them, the six-year dataset is assumed to be a circular dataset, i.e., the last recording will be followed by the first recording, and the whole dataset repeats itself like a circle. The maximum value of load recording is selected to represent 1.0 p.u., which is 50.81 MW.

3.3.1 Sliding Window Approach

Demands on weekdays and weekends are quite different and load profiles are therefore classified into two categories: weekday and weekend. A sliding window is implemented for each record in the active power measurements series, consisting of all recordings from the same hour of an identical category day (weekday or weekend) within ± 14 days interval. For an individual weekday record, there are $21 = 4 \times 5 + 1$

values in the sliding window, while for every weekend record, there are $9 = 4 \times 2 + 1$ sliding windows entries.

The variations of recorded demands in the sliding window represent the variation associated with the actual recording (i.e., centre position value in the window), which can be expressed by either maximum or minimum, or quantile points, as shown in Fig. 3.2.

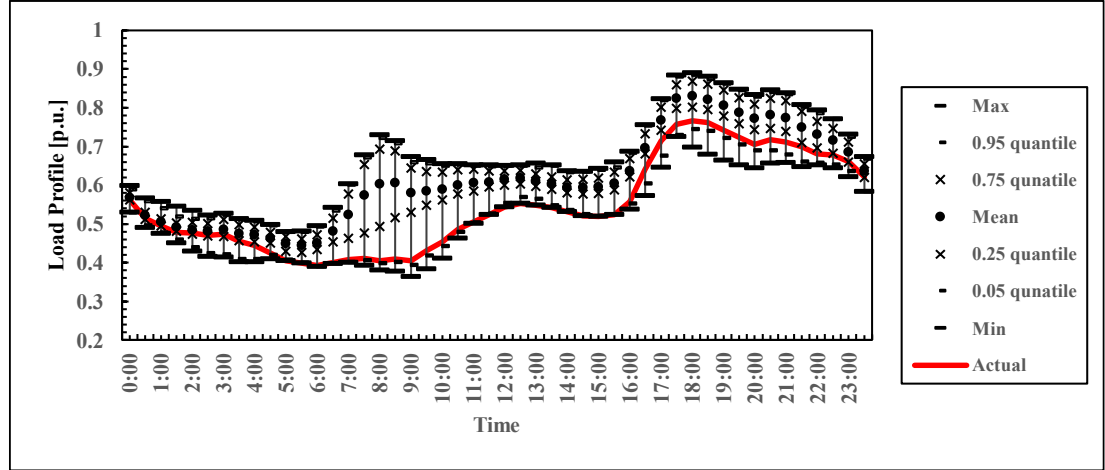


Fig. 3.2. Evaluation of demand variation with sliding window approach (Scotland MV substation dataset, Tuesday, 2 January 2007).

The example of the ± 14 days window in Fig. 3.2 shows how the actual recordings (red solid line) are around the mean value for some hours, but for other hours are also close to the minimum values recorded in the window. The results show that for the same hour of the same type of day, the load values have a wide range of variations.

3.3.2 Second-Order Markov Chain Analysis

The second-order MC model, in which the probability of transition from one state to the next state depends only on the two successive previous states, is also used to analyse variations in the weekday and weekend load profile time series:

$$P(x_{n+1}|x_1, x_2, \dots, x_{n-1}, x_n) = P(x_{n+1}|x_{n-1}, x_n) \quad (3.1)$$

where $x_1, x_2, \dots, x_n$ is a sequence of n random variables, and $P(\cdot|\cdot)$ represent conditional probability. The recorded active power data are fuzzily approximated with a resolution of 0.5 MW as:

$$AP_{\text{new}} = \left\lfloor \frac{AP_{\text{old}}}{0.5} \right\rfloor \times 0.5 + \frac{0.5}{2} \quad (3.2)$$

where: AP_{old} is the original recorded active power, $\lfloor \cdot \rfloor$ is floor function, and AP_{new} is the discretised active power, which transforms the continuous load measurements to discrete states, so that the MC model could handle data more efficiently. The number of MC model states is equal to the unique counts of discretised active power, with transition probability matrix, $\mathbf{P}$, giving probabilities of state-to-state transfers; and cumulative transition probability matrix $\mathbf{C}$ represents cumulative transfer probabilities:

$$\mathbf{P} = \begin{bmatrix} p_{111} & p_{112} & \cdots & p_{11s} \\ p_{121} & p_{122} & \cdots & p_{12s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1s1} & p_{1s2} & \cdots & p_{1ss} \\ \vdots & \vdots & \ddots & \vdots \\ p_{ss1} & p_{ss2} & \cdots & p_{sss} \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} c_{111} & c_{112} & \cdots & c_{11s} \\ c_{121} & c_{122} & \cdots & c_{12s} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1s1} & c_{1s2} & \cdots & c_{1ss} \\ \vdots & \vdots & \ddots & \vdots \\ c_{ss1} & c_{ss2} & \cdots & c_{sss} \end{bmatrix} \quad (3.3)$$

where: s is the number of states, and p_{ijk} is the probability of transferring from state i (at $t = t - 1$) and following state j (at $t = t$) to the state k (at $t = t + 1$). The elements of $\mathbf{P}$ are calculated with the MLE method from [117] and elements of $\mathbf{C}$ are then calculated as $c_{ijk} = \sum_{x=1}^k p_{ijx}$. The empirical mean, variance, median, maximum and minimum values, as well as cumulative probability boundaries of quantiles (denoted as $(1 - Q_x)$ and Q_x , where Q_x is quantile value) of transferring from states i and j are:

$$\begin{cases} \text{mean}_{ij} = \sum_{k=1}^s k \times p_{ijk} \\ \text{variance}_{ij} = \sum_{k=1}^s p_{ijk} \times (k - \text{mean}_{ij})^2 \\ \text{median}_{ij} = k \quad \text{s.t.} \quad |c_{ijk} - 0.5| = \min(|c_{ijk} - 0.5|) \\ \text{max}_{ij} = k \quad \text{s.t.} \quad c_{ijk} = 1, c_{ijk+1} = 1, \dots, c_{ijs} = 1 \\ \text{min}_{ij} = k \quad \text{s.t.} \quad c_{ijk} \neq 0, c_{ijk-1} = 0, \dots, c_{ij1} = 0 \\ (1 - Q_x)_{ij} = k \quad \text{s.t.} \quad |c_{ijk} - (1 - Q_x)| = \min(|c_{ijk} - (1 - Q_x)|) \\ (Q_x)_{ij} = k \quad \text{s.t.} \quad |c_{ijk} - Q_x| = \min(|c_{ijk} - Q_x|) \end{cases} \quad (3.4)$$

For median_{ij} , if there exist multiple solutions of k , then their mean value is chosen. If $(1 - Q_x)_{ij} = k$ has multiple solutions, the smallest k is selected, and for $(Q_x)_{ij} = k$, the largest k is selected, if it has more than one solution. An example of the second-order MC model for assessing variations in hourly demands is shown in Fig. 3.3 (the same

day as in Fig. 3.2). The second-order MC model is based on the whole recorded dataset, where two successive previous recordings are considered as the condition for the next estimated value.

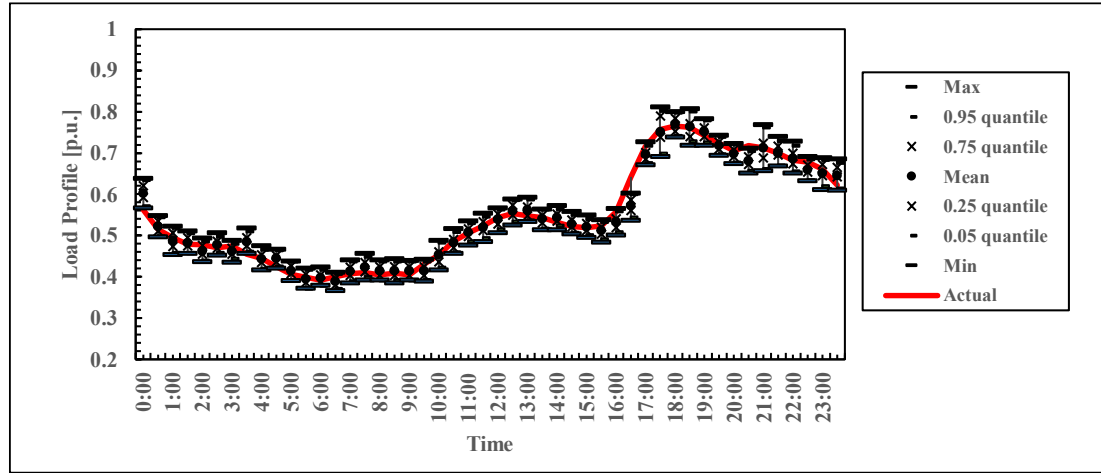


Fig. 3.3. Demand variation evaluation with the second-order MC model (Scotland MV substation dataset, Tuesday, 2 January 2007).

The second-order MC model provides much smaller variations than the longer-term day type and time specified sliding window approach, which is due to “second-order” model putting more strict conditions on the conditional distribution and limited number of samples at every row in $\mathbf{P}$ and $\mathbf{C}$. The conclusion for both sets of results is that there exists a range of load variations from either point of view.

3.4 Day-Ahead Load Forecasting

This section analyses three state-of-the-art methods (MLP, SB-LSTM, and gradient boosting regression trees, GBRT) for the day-ahead load forecasting and compares their performance on the Scotland MV substation dataset and the China city dataset. It shows the implementation of the specific model instances, their inputs and outputs, hyperparameter settings and performance. The review of the MLP and LSTM models is provided in Section 2.2.1, and for tree-based model, GBRT, it is introduced in Section 3.4.3.

3.4.1 Data Pre-processing: Dataset Splitting, One Hot Encoder, and Data Normalisation.

Each of the whole Scotland MV substation dataset and the whole China city dataset are divided into training set, validation set and test set, where the minimum demand week, the average demand week and the maximum demand week in the last year are selected for testing, while all previous recordings exactly before the test set are used as training/validation set. The training set and validation set are then distinguished by splitting the training/validation set by the ratio of 0.8 : 0.2 in contiguous blocks for every 100 days. That is, in training/validation set, day $1 + 100 \times i$ to day $80 + 100 \times i$ are used for training, and day $81 + 100 \times i$ to day $100 + 100 \times i$ are validation set, where i is a non-negative integer.

Standard calendar variables are inherently categorical features and the overall accuracy of neural network-based models can be improved by firstly encoding categorical variables using one hot encoder [118]. Solar analemma, Moon phase, meteorological and load recordings are linearly scaled and mapped into a bounded range using the min-max normalisation:

$$x_{\text{new}} = \min(x_{\text{new}}) + \frac{\max(x_{\text{new}}) - \min(x_{\text{new}})}{\max(x_{\text{old}}) - \min(x_{\text{old}})} \times (x_{\text{old}} - \min(x_{\text{old}})) \quad (3.5)$$

where: x_{old} and x_{new} are the original and transformed data series, respectively. The minimum and maximum values of the original data series are assumed to be the minimum and maximum recordings in the training/validation set, which does not consider any information from the test set. More specifically, in this day-ahead load forecasting problem, $\min(x_{\text{new}})$ and $\max(x_{\text{new}})$ are set as -1 and 1, respectively, so x_{new} is within the range of [-1, 1].

3.4.2 Instantiation of MLP and SB-LSTM Model

In MLP model, to forecast the day-ahead load profile, multiple load values at all moments in the next day (i.e., 96 values for the China city dataset and 48 points for the Scotland MV substation dataset) are predicted separately. For every point of forecasted load value, the model uses following inputs: 1) calendar variables (YY, MM, DD, HH, DW, HLD, DST, SA, SE and MP) of the forecasted point/moment; 2)

meteorological data of the forecasted point/moment (assuming they can be forecasted with 100 % accuracy): for the China city dataset, only temperature (TEMP), while full meteorological data (TEMP, SI, PPT, WS) for the Scotland MV substation dataset; 3) load recordings at the identical moment from the past 28 days. The model output is the predicted load value at the forecasted moment, where the full day-ahead forecasting is obtained by applying the model multiple times for all moments of the next day.

Three hidden layers and three dropout layers are implemented in the MLP model (each hidden layer consists of 256 neurons). Leaky rectified linear units (LReLU) is selected as activation function following each hidden layer, due to its good performance in deep learning and advantages over normal ReLU in handling negative values [119]. L2 regularisation (factor = 0.0005) is added to prevent overfitting and Adam optimiser algorithm is used [120]. Objective is to minimise mean absolute error between model outputs and actual recordings [121].

One advantage of SB-LSTM is that the characteristics of the load series need not to be artificially extracted, so this class of model can process sequence data and learn its deep features directly. Therefore, there will be no need to forecast different moments separately. In SB-LSTM model, the inputs are the 28 daily sequence of load measurements from the previous 28 days, the meteorological data in the predicted day (assuming they can be forecasted with 100 % accuracy) and the calendar variables in the forecasted day. The output is the load profile in the forecasted day. Like the MLP model, one hot encoder and min-max normalisation are used to pre-process the data before inputting into the model. The SB-LSTM model is implemented with three bidirectional LSTM layers and three dropout layers, while L2 regularisation (factor = 0.0005) and Adam optimiser are again used to minimise mean absolute error.

The SB-LSTM model is implemented with three stacked bidirectional LSTM layers (512 neurons are used for every BiLSTM layer) and three dropout layers (dropout rate = 0.5). This SB-LSTM architecture is illustrated in Fig. 3.4.

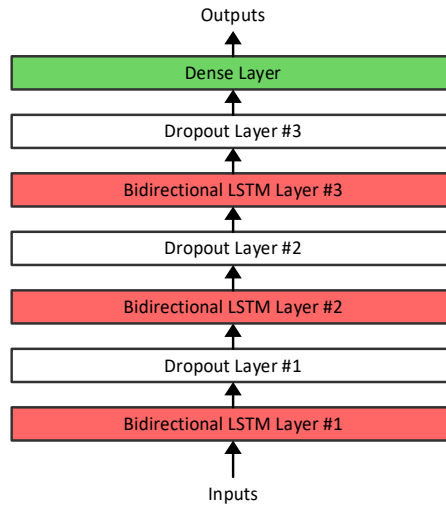


Fig. 3.4. The architecture of SB-LSTM model.

3.4.3 Benchmarking NN-Based Model with Tree-Based Model: Gradient Boosting Regression Trees

Unlike neural network models, tree-based models are white box models, rather than black box models. A regression tree can be grown through an iterative process that recursively divides data into partitions, which is inherently a statistical method based on conditional probabilities in features' class space. Since the splitting allocates the data to small groups according to the given predictors, regression trees are usually well explainable.

A typical regression tree is a top-down structure, where at each step constructing algorithm chooses a variable that can most efficiently split the dataset; efficiency can be evaluated by various metrics, e.g. information gain and variance [122]. A fully-grown regression tree is very likely to suffer overfitting issues (shallow tree) and generally cannot perform well on any other datasets, except the training set.

Ensemble learning can improve the flexibility of regression tree model by growing multiple regression trees to obtain better prediction, instead of relying only on one shallow tree [123]. Gradient boosting can be used for creating accurate ensembles that are typically formed of tree-based models. At each boosting stage, the algorithm will try to identify weak regression trees that have the maximum predicting errors and then improve these learners by adding extra estimators fitted to residuals. At the next boosting stage, the new regression trees will attempt to reduce the errors of their

predecessors. As residuals are negative gradients of the squared errors, gradient boosting is actually a gradient reduce algorithm [124].

The Gradient Boosting Regression Trees (GBRT) model uses the same inputs and outputs as the MLP model, but it does not need to use one hot encoder, or min-max normalisation to pre-process the data before feeding the model, because they can be regarded as individual features with specific physical meanings in the tree-based model. There are 512 learners implemented in the GBRT and the minimum leaf size is set as 5, which means there should be at least 5 responses of each leaf node. Least-square boosting (LSBoost) algorithm is adopted to fit the regression [125].

3.4.4 Example Case Study 3.1: Comparison of Day-Ahead Load Forecasting Results by Three Considered Models

The examples of forecasting results of day-ahead load profiles are shown in Fig. 3.5 and Fig. 3.6 for the China city dataset and Scotland MV substation dataset, respectively. They are both day-by-day (one day after another) day-ahead forecasting results for Monday to Sunday in the same weeks. Table 3.1 to Table 3.6 provide numerical results for quantifying the benefits of different models, which should be evident through the comparisons of MAE and MAPE [121] (see Section 2.3.1). The absolute and percentage errors in the total, overestimated and underestimated energy consumptions (denoted as E_T , E_O and E_U respectively) are also compared with the actual demands, assuming that the mean load values are demanded over each averaging window (15 minutes for China city dataset or 30 minutes for Scotland MV substation dataset) in the considered dataset. The column named “Total Energy” presents the realised energy consumption, thus the values are the same for different models. More specifically, if ΔDEM is defined as the difference between the forecasted demand, $\widehat{DEM}$, and actually measured demand, DEM , of the substation at a time stamp t_i :

$$\Delta DEM(t_i) = \widehat{DEM}(t_i) - DEM(t_i) \quad (3.6)$$

the estimation of the total energy demand (as algebraic sum and sum of absolute values) is calculated for $\Delta t = 15$ minutes (or 30 minutes) by trapezoidal integration.

Indicating as $\Delta E(t_i)$ the infinitesimal area calculated between the two values, it is possible to distinguish total overestimated and underestimated demand by defining two different sets of time stamps: T_o time stamps when forecasted $\widehat{DEM}$ is greater than actual DEM, and T_u in the opposite case.

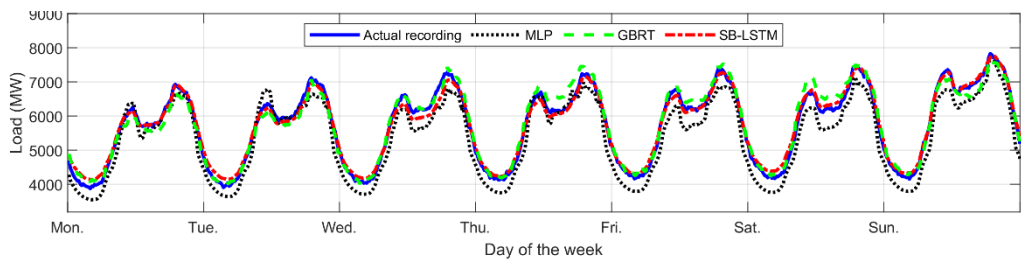
$$E_o = \sum_{i \in T_o} \Delta E(t_i) \quad (3.7)$$

$$E_u = \sum_{i \in T_u} \Delta E(t_i) \quad (3.8)$$

$$E_T = E_o + E_u \quad (3.9)$$

and their percentage values are defined as $\frac{E_o}{\sum_{i \in T_o} DEM(t_i)}$, $\frac{E_u}{\sum_{i \in T_u} DEM(t_i)}$ and $\frac{E_T}{\sum_{i \in (T_o \cup T_u)} DEM(t_i)}$, respectively. Only the time stamps when the reconstructed demand is higher or lower than the actual recordings are considered in the calculations of the amount of overestimated/underestimated energy and their percentages. Therefore, in some cases, there are higher percentages of overestimated or underestimated energy while their absolute values are lower, and vice versa, since different methods may have different time stamps for overestimated or underestimated energy.

The variations of load profiles on different days may be caused by a lot of factors, such as weather changes and day-of-the-week effects [110]. All models have appreciated the impacts of meteorological factors and different days of the week, and they all can capture the main shapes and trends of load profiles. The errors of GBRT and SB-LSTM methods are both lower than MLP method errors regarding MAE and MAPE. From the presented results, SB-LSTM method can provide the most accurate predictions in the two datasets, and is more flexible in analysing sequenced load data.



(a)

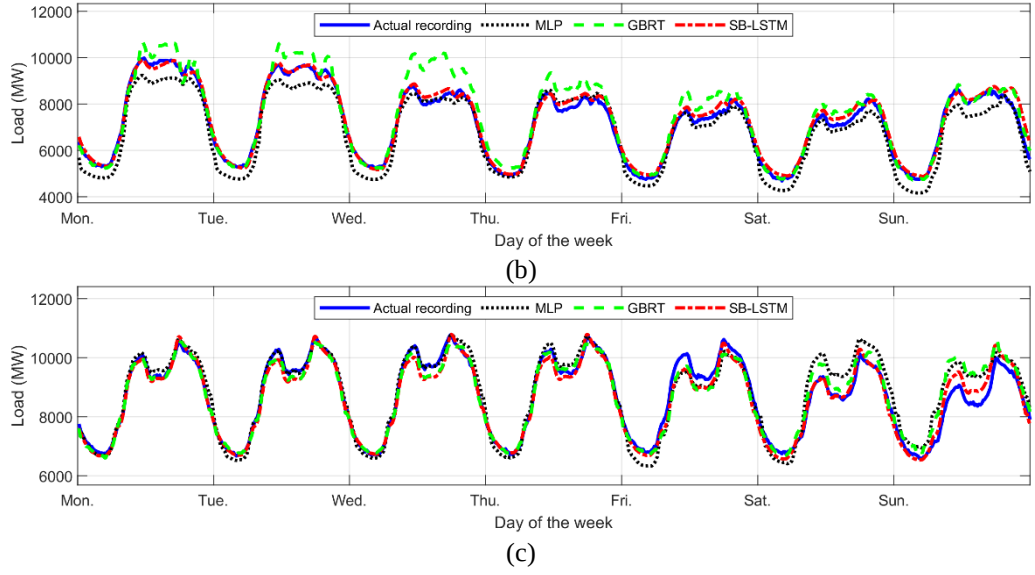


Fig. 3.5. Examples of day-ahead load forecasting in the China city dataset: (a) The minimum demand week. (b) The average demand week. (c) The maximum demand week.

Table 3.1 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.5(a))

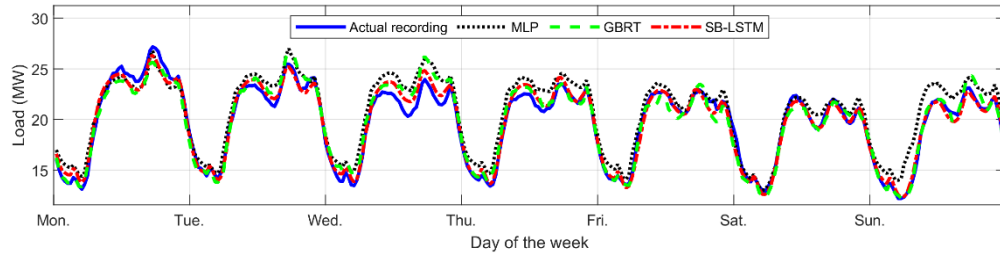
	MAE (MW)	MAPE (%)	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)	Total Energy (GWh)
MLP	370.174	6.751	1.733	2.738	60.456	6.718	-58.723	-6.096	963.262
GBRT	151.039	2.575	16.762	2.785	8.613	2.383	8.149	0.846	963.262
SB-LSTM	107.915	2.027	10.973	2.014	7.157	1.711	3.817	0.396	963.262

Table 3.2 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.5(b))

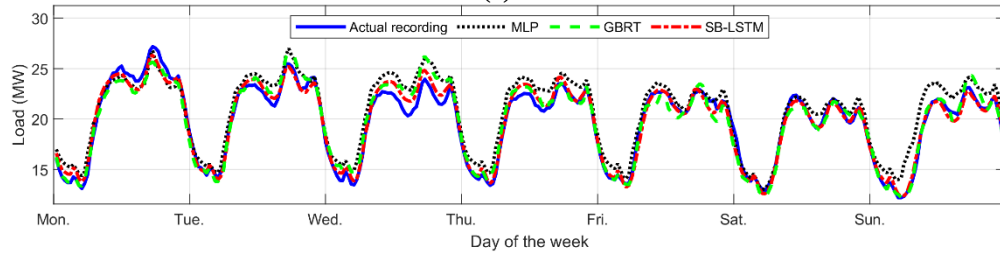
	MAE (MW)	MAPE (%)	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)	Total Energy (GWh)
MLP	412.577	6.060	4.417	2.269	64.896	6.396	-60.480	-5.001	1209.353
GBRT	382.546	4.952	59.838	6.418	4.430	1.599	55.408	4.582	1209.353
SB-LSTM	165.138	2.356	22.534	2.694	5.209	1.397	17.325	1.433	1209.353

Table 3.3 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.5(c))

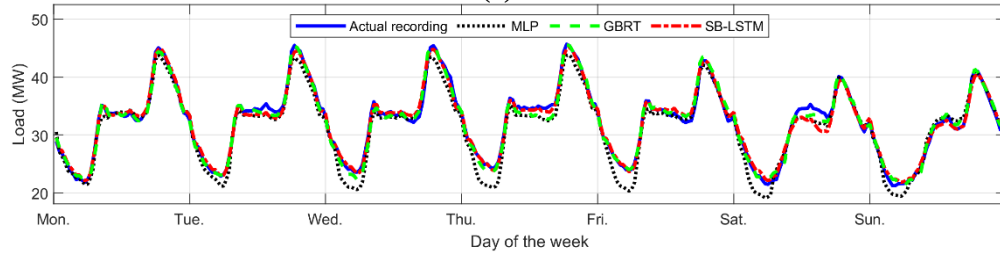
	MAE (MW)	MAPE (%)	E_O (GWh)	E_O (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)	Total Energy (GWh)
MLP	299.849	3.548	35.189	4.049	15.185	2.561	20.004	1.368	1462.043
GBRT	222.017	2.543	22.031	3.550	15.268	1.815	6.763	0.463	1462.043
SB-LSTM	165.530	1.864	8.757	1.817	19.052	1.944	-10.296	-0.704	1462.043



(a)



(b)



(c)

Fig. 3.6. Examples of day-ahead load forecasting in the Scotland MV substation dataset: (a) The minimum demand week. (b) The average demand week. (c) The maximum demand week.

Table 3.4 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.6(a))

	MAE (MW)	MAPE (%)	E_O (MWh)	E_O (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)	Total Energy (MWh)
MLP	1.145	6.125	178.218	6.371	14.191	2.760	164.026	4.953	3311.580
GBRT	0.647	3.223	63.895	3.652	44.728	2.864	19.167	0.579	3311.580
SB-LSTM	0.448	2.296	45.073	2.402	30.143	2.100	14.930	0.451	3311.580

Table 3.5 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.6(b))

	MAE (MW)	MAPE (%)	E_O (MWh)	E_O (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)	Total Energy (MWh)
MLP	0.823	3.432	66.817	3.208	71.461	3.413	-4.644	-0.111	4176.947
GBRT	0.514	2.066	34.939	1.768	51.441	2.337	-16.502	-0.395	4176.947
SB- LSTM	0.411	1.635	38.432	1.717	30.688	1.583	7.744	0.185	4176.947

Table 3.6 Comparison of Model Results and Actual Measurements in the China City Dataset (For 7 Days in Fig. 3.6(c))

	MAE (MW)	MAPE (%)	E_O (MWh)	E_O (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)	Total Energy (MWh)
MLP	1.215	4.080	23.567	1.859	180.554	4.325	-156.987	-2.884	5442.538
GBRT	0.551	1.763	37.532	1.584	55.046	1.791	-17.514	-0.322	5442.538
SB- LSTM	0.525	1.660	44.182	1.553	44.028	1.695	0.154	0.003	5442.538

The MLP is a relatively straightforward and naïve model, in which input signals can only travel in one direction to the output layer with no feedback and memory structure. It is suitable for modelling the nonlinear relationship between its input and output, but selections of input features are crucial and pattern extraction is difficult in complex problems; in presented load forecasting study, MLP fails to capture the full temporal correlations and changes in the daily load profiles.

The GBRT is an ensemble model, which combines many regression trees to learn the characteristics of load profiles. Although every individual regression tree is weak and shallow, together they provide more accurate predictions. Another reason why GBRT is better than MLP is that it can handle categorical features naturally and there is no need to use one hot encoder, or data normalisation. Since the inputs of GBRT are also artificial extracted features (i.e., recordings at separate moments), many hidden but important patterns may be lost and its performance is lower than SB-LSTM method.

In SB-LSTM, it is not necessary to transfer the dynamic sequence modelling problem to the static correlation modelling problem. Therefore, its input can naturally preserve the complete temporal correlation information. For load profile modelling, SB-LSTM not only considers load data at the same moment of past days, but analyses all load recordings at other moments and their hidden correlations and deep patterns. Therefore, it is the most robust and promising model among all three models.

There are days where MLP, GBRT and SB-LSTM all fail to have more accurate predictions, because load profiles in these days are inherently and significantly different from other days learned by the models. This may be due to more complex factors, planned/unplanned network maintenance and possibly demand-side-management actions, especially during the hot summer days, when temperatures are very high in the China city dataset. These all bring higher uncertainty to the network operating conditions and make the forecasting less accurate.

3.5 Week-Ahead Load Forecasting

This section extends the model with the best performance in day-ahead load forecasting, SB-LSTM, to be a week-ahead load forecasting model, and the performance is tested and compared on the Scotland MV substation dataset.

The inputs of the SB-LSTM models are very similar for the day-ahead and week-ahead forecasting problems, and the extra six days' data can be simply treated as additional features. In summary, the SB-LSTM models use the following sequence data as inputs:

- 1) Calendar variables (including standard calendar variables, solar analemma (sun azimuth and elevation angles) and moon phase) of the forecasted day/week, i.e., $(SCV_{+1 \text{ or } +1\sim7}, SA_{+1 \text{ or } +1\sim7}, SE_{+1 \text{ or } +1\sim7}, MP_{+1 \text{ or } +1\sim7})$, where $SCV_{+1 \text{ or } +1\sim7} = (YY_{+1 \text{ or } +1\sim7}, MM_{+1 \text{ or } +1\sim7}, DD_{+1 \text{ or } +1\sim7}, HH_{+1 \text{ or } +1\sim7}, DW_{+1 \text{ or } +1\sim7}, HLD_{+1 \text{ or } +1\sim7}, DST_{+1 \text{ or } +1\sim7})$;
- 2) Meteorological data (temperature, solar irradiation, precipitation and wind speed) of the forecasted day/week (assuming they can be forecasted with 100 % accuracy), i.e., $MD_{+1 \text{ or } +1\sim7} = (TEMP_{+1 \text{ or } +1\sim7}, SI_{+1 \text{ or } +1\sim7}, PPT_{+1 \text{ or } +1\sim7}, WS_{+1 \text{ or } +1\sim7})$;
- 3) Load recordings from past 28 days before the forecasted day/week, i.e., $AP_{-27 \sim 0} = (AP_{-27}, AP_{-26}, AP_{-25}, \dots, AP_0)$.

The outputs of the day-ahead forecasting and week-ahead forecasting of SB-LSTM models are AP_{+1} and $AP_{+1\sim7} = (AP_{+1}, AP_{+2}, AP_{+3}, AP_{+4}, AP_{+5}, AP_{+6}, AP_{+7})$, respectively. The hyperparameters (e.g., dropout rate, number of neurons, and L2 regularisation factor) and the architecture (see Fig. 3.4) are all the same.

3.5.1 Example Case Study 3.2: Comparison of Day-Ahead and Week-Ahead Load Forecasting Results by SB-LSTM

The examples of forecasting results of day-ahead and week-ahead load profiles are shown in Fig. 3.7, Fig. 3.8 and Fig. 3.9 for the test set of the minimum demand week, the average demand week and the maximum demand week, respectively. The error metrics for these results are shown in Table 3.7, Table 3.8 and Table 3.9. For day-ahead forecasting, the weekly results are obtained as day-by-day predictions (one day after another, in a rolling forecasting manner) for Monday to Sunday in the same weeks, while for week-ahead forecasting, the SB-LSTM model gives the predictions for the whole week directly in one run as a single forecasting.

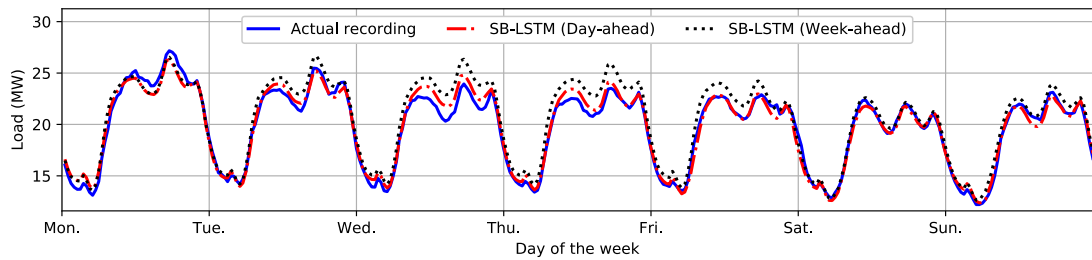


Fig. 3.7. Comparison of day-ahead and week-ahead load forecasting results for the week of minimum demand.

Table 3.7 Evaluation of SB-LSTM Performance (For 7 Days in Fig. 3.7)

	MAE (MW)	MAPE (%)	E _O (MWh)	E _O (%)	E _U (MWh)	E _U (%)	E _T (MWh)	E _T (%)	Total Energy (MWh)
Day-ahead	0.448	2.296	45.073	2.402	30.143	2.100	14.930	0.451	3311.580
Week-ahead	0.933	4.721	147.168	5.163	9.505	2.060	137.663	4.157	3311.580

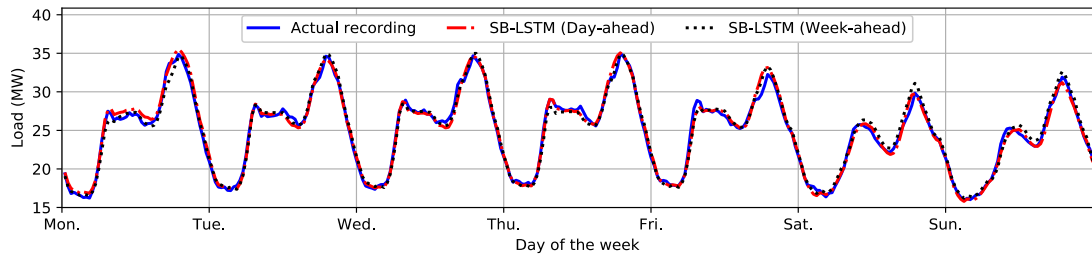


Fig. 3.8. Comparison of day-ahead and week-ahead load forecasting results for the week of average demand.

Table 3.8 Evaluation of SB-LSTM Performance (For 7 Days in Fig. 3.8)

	MAE (MW)	MAPE (%)	E_O (MWh)	E_O (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)	Total Energy (MWh)
Day-ahead	0.411	1.635	38.432	1.717	-30.688	-1.583	7.744	0.185	4176.947
Week-ahead	0.495	1.973	50.619	1.963	-32.622	-2.041	17.998	0.431	4176.947

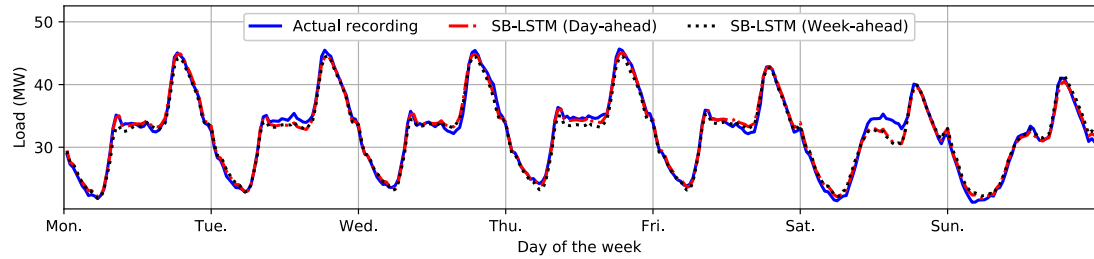


Fig. 3.9. Comparison of day-ahead and week-ahead load forecasting results for the week of maximum demand.

Table 3.9 Evaluation of SB-LSTM Performance (For 7 Days in Fig. 3.9)

	MAE (MW)	MAPE (%)	E_O (MWh)	E_O (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)	Total Energy (MWh)
Day-ahead	0.525	1.660	44.182	1.553	44.028	1.695	0.154	0.003	5442.538
Week-ahead	0.708	2.218	41.718	2.050	77.274	2.267	-35.555	-0.653	5442.538

From the presented results, it can be seen that SB-LSTM method can provide accurate results for both day-ahead and week-ahead load forecasting, although the error is greater in the latter case. As illustrated, the day-ahead forecasting SB-LSTM model and the week-ahead forecasting SB-LSTM model are quite similar, which means that SB-LSTM approach is very flexible, and it can capture deep dependencies and features of the analysed data sequences with different lengths.

Although the presented results show that day-ahead load forecasting has higher accuracy than week-ahead forecasting, there are days where both day-ahead and week-ahead forecasting fail to provide accurate predictions. The load profiles in these days are significantly different from the other days that were used during the learning process by the models. The main reason for that may be due to the impact of “hidden” (i.e., not considered) variables, e.g., information on planned or unplanned network

maintenance, possible demand-side-management actions and similar, which all increase complexity and bring higher uncertainty to the network operating conditions, making resulting forecasting less accurate.

3.6 Forecasting of both the Total and Disaggregated Load

In this section, the SB-LSTM model is further extended and applied on AMPds2 dataset, to provide day-ahead forecasting for both total and heating load. AMPds2 is from a household as previously described, and it only has two years of measurements. For this forecasting problem, the first year is used as the training/validation set, and the second year is the test set.

The same SB-LSTM architecture as in the previous day-ahead and week-ahead load forecasting problems (see Fig. 3.4) is used, but the number of neurons in each BiLSTM layer is doubled and increases to 1024, since the household level data are more complex. All other hyperparameters for SB-LSTM remain unchanged.

The inputs include: 1) standard calendar variables (only MM, HH, DW, HLD and DST) of the forecasted day. DD, SA, SE and MP are not considered for simplicity, and YY is not included since the data are all in the same year; 2) meteorological data (only TEMP) from one day before the forecasted day. No assumption that the forecasted day's weather can be predicted accurately is made, as this model is built to learn the relationship between historical temperature and future load. Although the accuracy may not be as good as using the forecasted day's weather information, it is a more realistic/practical model; 3) total load recordings from past continuous 14 days before the forecasted day. No historical heating information is inputted into the model to make this forecasting more challenging, which means the network needs to estimate the heating demand by capturing information only from total load recordings, time information and weather data.

The outputs of the SB-LSTM model are day-ahead forecasted total load and heating load, which is a high-dimensional, rather than single-dimensional response variable.

3.6.1 Example Case Study 3.3: Multi-dimensional Day-Ahead Load Forecasting Results by SB-LSTM

The examples of rolling day-ahead forecasting results for weeks with the minimum, average and maximum total demand are shown in Fig. 3.10, Fig. 3.11 and Fig. 3.12, with their error metrics are calculated and listed in Table 3.10, Table 3.11 and Table 3.12.

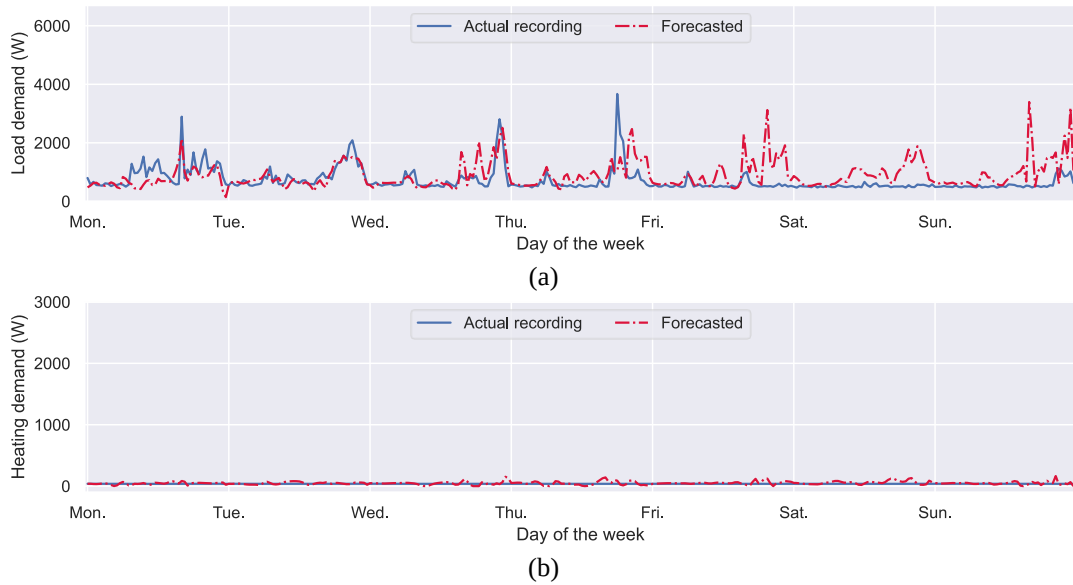
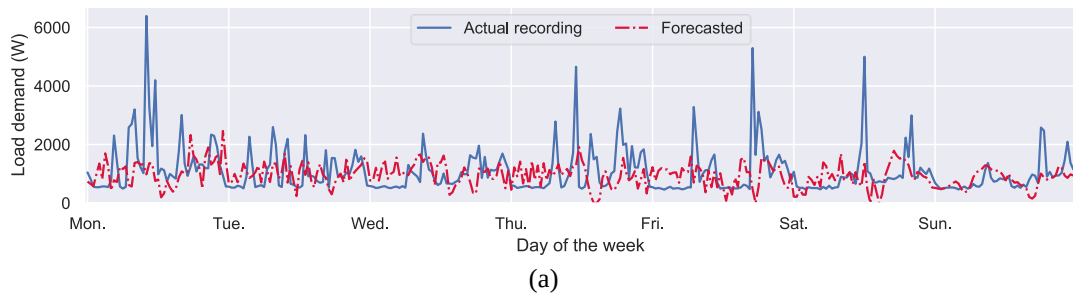
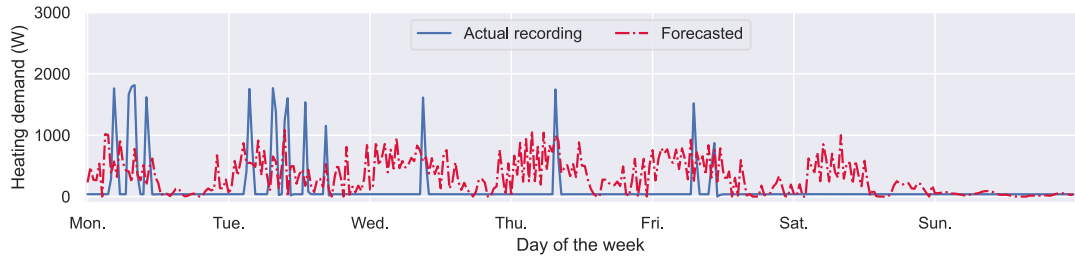


Fig. 3.10. The comparison of day-ahead forecasting of total demand and heating demand for the week of minimum total demand.

Table 3.10 Performance of Day-ahead Forecasting of Total Demand and Heating Demand (For 7 Days in Fig. 3.10)

	MAE (W)	MAPE (%)	E_O (kWh)	E_O (%)	E_U (kWh)	E_U (%)	E_T (kWh)	E_T (%)	Total Energy (kWh)
Total Demand	333.917	53.265	43.019	59.124	13.079	27.639	29.941	24.934	56.098
Heating Demand	20.113	50.658	2.612	58.965	0.767	39.836	1.845	29.039	3.379



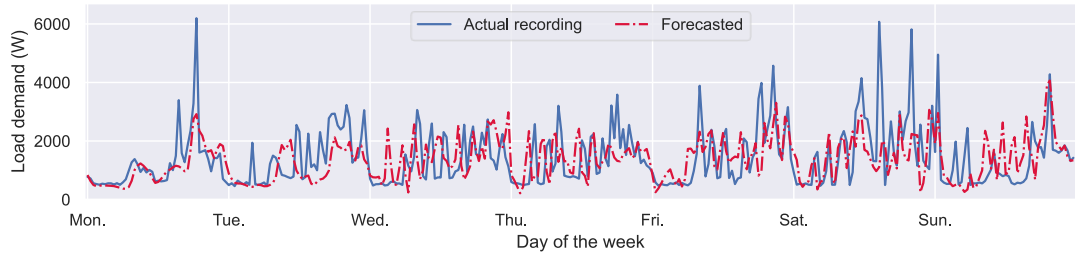


(b)

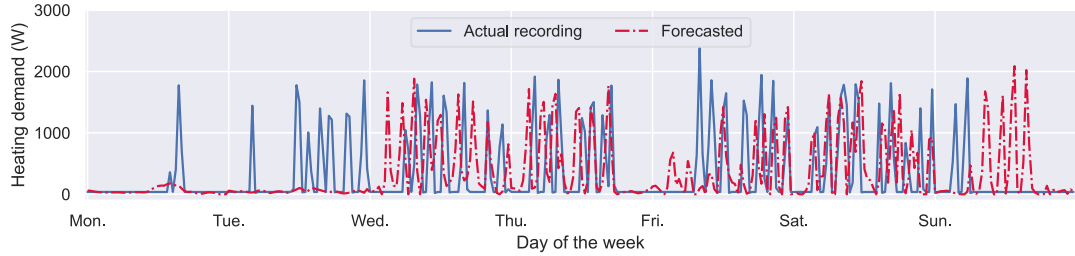
Fig. 3.11. The comparison of day-ahead forecasting of total demand and heating demand for the week of average total demand.

Table 3.11 Performance of Day-ahead Forecasting of Total Demand and Heating Demand (For 7 Days in Fig. 3.11)

	MAE (W)	MAPE (%)	E_o (kWh)	E_o (%)	E_U (kWh)	E_U (%)	E_T (kWh)	E_T (%)	Total Energy (kWh)
Total Demand	578.864	56.398	37.261	61.615	59.988	49.293	- 22.727	-12.476	97.249
Heating Demand	301.077	720.500	40.607	683.294	9.974	64.934	30.633	143.799	50.581



(a)



(b)

Fig. 3.12. The comparison of day-ahead forecasting of total demand and heating demand for the week of maximum total demand.

Table 3.12 Performance of Day-ahead Forecasting of Total Demand and Heating Demand (For 7 Days in Fig. 3.12)

	MAE (W)	MAPE (%)	E_o (kWh)	E_o (%)	E_U (kWh)	E_U (%)	E_T (kWh)	E_T (%)	Total Energy (kWh)
Total Demand	660.752	54.863	48.938	61.722	62.069	38.630	- 13.131	-5.472	111.006

Heating Demand	425.871	558.990	36.031	456.531	35.515	74.511	0.516	0.929	71.546
----------------	---------	---------	--------	---------	--------	--------	-------	-------	--------

The results contain two parts: one for total load demand and another for heating demand forecasting. In Fig. 3.10 (Table 3.10), both forecasting and observed heating demand are very low and near zeros. It is because that the dataset is collected from a household in Burnaby, Canada. In Burnaby, no heating devices are required during summer. Therefore, for the minimum demand week (which happens in July), the values of heating demand are quite small. In the weeks with the average and maximum total demand, the forecasting results are not as accurate as in the Scotland MV substation and China city dataset. The household level time series is more variable (complex and less-smooth), and the length of the available dataset is shorter. However, the general trends are still captured, in particular, for Fig. 3.12 (Table 3.12), SB-LSTM successfully captures the changes in the total load. The heating load are forecasted with higher accuracy than expected because the heating load is not included as predictor variable in the model. Furthermore, the MAPE values for heating demand in the average and maximum demand weeks exceed 100 %, because there are some cases where the actual measurements are low (near zero), but the forecasted values are very high (1000 – 2000 W), which confirms the issues of MAPE as discussed in Section 2.3.1.

3.7 Discussions and Conclusions

This chapter firstly compares two methods in assessing the load variations: sliding window approach and second-order Markov chain analysis. The sliding window method is based on ± 14 days observation window centred with the measurement and also considers different day types. The second-order Markov chain method gives much smaller ranges of variations, which approaches the problem by conditioning on two previous consecutive measurements. And both methods imply the existence of load variation.

The forecasting analysis begins with the comparison of day-ahead load forecasting with MLP, GBRT and SB-LSTM methods, based on historical load recordings and meteorological (weather) data. The SB-LSTM is found to be the most accurate and flexible method and it is further extended to provide weekly-ahead load forecasting

and multi-dimensional day-ahead total load and heating forecasting. In the Scotland and China city datasets case studies, SB-LSTM achieves satisfactory accuracy. For AMPds2 dataset, the forecasting accuracy decreases and SB-LSTM struggles to capture the general time series trends.

The SB-LSTM method is a type of deep recurrent neural network model, which is capable of resolving the gradient vanishing and gradient exploding issues of the traditional recurrent neural network models. In SB-LSTM method, the input can naturally preserve the full correlational information on temporal dependencies. For load forecasting problems, SB-LSTM analyses the previous days' load recordings and the meteorological forecast information recursively, capturing their hidden correlations and deep patterns, and providing robust and promising predictions.

Chapter 4

Heating and Lighting Load Disaggregation Using Frequency Components and Convolutional Bidirectional Long Short-Term Memory Method

4.1 Introduction

Load disaggregation for the identification of specific load types in the total demands (e.g., demand-manageable loads, such as heating or cooling loads) is becoming increasingly important for the operation of existing and future power supply systems. This chapter aims to contribute to the existing research in the general area of load disaggregation by proposing a novel frequency domain decomposition and correlation method, which requires only synchronised time series of active power demands and weather variables (temperature and solar irradiance). It introduces an approach in which periodical changes in the total demands (e.g., daily, weekly, and seasonal variations) are disaggregated into corresponding frequency components and correlated with the same frequency components in the meteorological variables (e.g., temperature and solar irradiance), allowing to select combinations of frequency components with the strongest correlations as the additional explanatory variables.

The three main contributions of this chapter are: a) introduction of a new Fourier series regression method for decomposition of synchronised demand and weather time series into target frequency components, b) formulation of an approach for evaluating correlations between obtained frequency components, allowing to select components with strongest correlations as the additional explanatory variables for the load disaggregation, and c) comparison of performance of deep learning model in which correlated frequency components are used as the additional explanatory variables with a conventional “black box” deep learning model (without frequency components), showing that the frequency component-based model provides a more accurate load disaggregation.

The chapter first presents a novel Fourier series regression method for obtaining target frequency components, which is illustrated on two household-level datasets and one

substation-level dataset. These results show that correlations between selected disaggregated frequency components are stronger than the correlations between the original non-disaggregated data. Afterwards, convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) methods are used to represent dependencies among multiple dimensions and to output the estimated disaggregated time series of specific types of loads, where Bayesian optimisation (BO) is applied to select hyperparameters of CNN-BiLSTM model. The CNN-BiLSTM and other deep learning models are reported to have excellent performance in many regression problems, but they are often applied as “black box” models without further exploration or analysis of the modelled processes. Therefore, the chapter compares CNN-BiLSTM model in which correlated frequency components are used as the additional explanatory variables with a naïve CNN-BiLSTM model (without frequency components). The presented case studies, related to the identification of electrical heating load and lighting load from the total demands, show that the accuracy of disaggregation improves after specific frequency components of the total demand are correlated with the corresponding frequency components of temperature and solar irradiance, i.e., that frequency component-based CNN-BiLSTM model provides a more accurate load disaggregation. Obtained results are also compared/benchmarked against the two other commonly used models, confirming the benefits of the presented load disaggregation methodology.

Fig. 4.1 below provides a “block diagram” illustration of the proposed load disaggregation approach, which also shows a layout of the overall analysis presented in the chapter. In Fig. 4.1, “Freq.”, “w/o” and “w/” are the abbreviations of “frequency”, “without” and “with”, respectively.

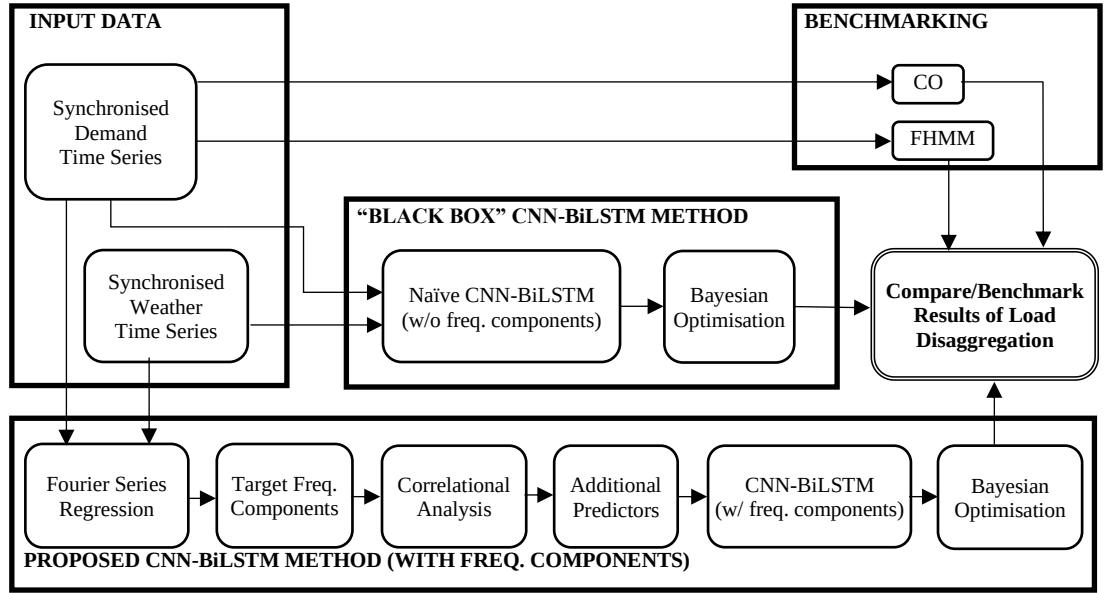


Fig. 4.1. A block diagram illustration of the proposed load disaggregation method and overall analysis presented in Chapter 4.

4.2 Selection of Target Frequency Components for Fourier Series Decomposition and Correlational Analysis

Periodical changes in weather conditions (e.g. daily and seasonal variations in temperature and solar irradiation) have strong impact on the periodical changes in residential electricity demands, which are also influenced by socio-economic factors, e.g., weekly schedule of working days and weekend days [9]. For example, space heating appliances in a cold-climate location are mainly used in winter and spring/autumn seasons, when ambient temperature is low, while demand for lighting is also usually higher in winter, because of the shorter duration of daylight hours. This results in a strong negative correlation of heating load demand with temperature and similarly strong negative correlation of lighting load with solar irradiance. As the temperature and solar irradiance also exhibit periodical changes on both longer-term scale (e.g., seasonal and annual variations) and shorter-term scale (e.g., diurnal or daily variations), the actual variations in electricity demands for heating and lighting are generally expected to follow these periodical changes. Accordingly, the correlational analysis can be used to select the most strongly correlated frequency components, which then can indicate portions of the disaggregated demands that contain higher

contributions from specific load types and therefore can serve as additional explanatory variables for a more accurate load disaggregation model.

This section first describes the available residential demand datasets. Afterwards, it presents a novel Fourier series regression (FSR) method with different window sizes, where each window allows to extract specific single-frequency component from the total demand, temperature and solar irradiance time series data, demonstrating that these frequency components can relatively accurately reconstruct the original dataset. Spearman correlation coefficients are used to quantify correlations, in order to identify combinations of specific disaggregated frequency components that have stronger negative correlations than the original non-disaggregated data.

4.2.1 Description of Available Datasets

To illustrate the proposed methodology, this chapter uses three available datasets, which are either measured or pre-processed to have temporal resolution of 30 minutes. The first is AMPds2 from [115] (the same as in Chapter 3), which contains total active and reactive power demand measurements, as well as measurements of active power of the heat pump of a single household located in Burnaby, Canada, from 1 April 2012 to 31 March 2014. The second dataset is UK Domestic Appliance-Level Electricity (UK-DALE) [126], consisting of measurements of total active and reactive power demands and measurements of active power of lighting load for one household in London, UK, from 18 March 2013 to 17 March 2016. For both AMPds2 and UK-DALE datasets, the synchronous data for meteorological variables (temperature and solar irradiance) in the same geographical location are obtained from an external data source, MERRA-2 [116]. Unlike the first two household-level datasets, the third dataset is related to an MV substation-level measurements of predominantly residential total active and reactive power demands in a city in Scotland, UK, together with temperature and solar irradiance recordings from 1 January 2007 to 31 December 2012 (i.e., the same Scotland MV substation dataset as in Chapter 3). In this case, the measurements of disaggregated loads, including heating and lighting loads, are not available. All three datasets are used as the case studies for Fourier series regression and correlation analysis, where in the next sections AMPds2 is used for the identification of heating load from the total demand using correlations with

temperature, UK-DALE is analysed for the identification of lighting load in the total demand using correlations with solar irradiance, while MV level dataset is used for considering both the correlation between the total demand and temperature, and correlation between the total demand and solar irradiance.

4.2.2 Disaggregation through Targeted Single-Components Fourier Series Decomposition

Fourier series decomposition is a commonly used approach for representing periodical function or “signal” with a sum of ideally sinusoidal components and one constant term, usually written as:

$$f(n) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi k}{N}n\right) + \sum_{k=1}^{\infty} b_k \sin\left(\frac{2\pi k}{N}n\right) \quad (4.1)$$

where: $f(n)$ is the input periodical function/signal with period N ; a_k and b_k are Fourier coefficients calculated by integrals $\frac{2}{N} \int_0^N f(n) \cos\left(\frac{2\pi k}{N}n\right) dn$ and $\frac{2}{N} \int_0^N f(n) \sin\left(\frac{2\pi k}{N}n\right) dn$, respectively; and $\frac{a_0}{2} = \frac{1}{N} \int_0^N f(n) dx$ represents the non-periodical, i.e., constant DC component.

For extracting the single frequency components, as it is done in this chapter for the target periodical components with specific window lengths, the following relation can be used:

$$f(n) = c_x + a_x \cos\left(\frac{2\pi}{N_x}n\right) + b_x \sin\left(\frac{2\pi}{N_x}n\right) \quad (4.2)$$

where: c_x is the DC component if $f(n)$ is fitted and determined by Fourier coefficients a_x and b_x , and the frequency is $1/N_x$.

If Equation (4.2) is applied with the multiple window lengths, where each window represents one identified or dominant periodical component in the analysed time series, and where for each component related window length is applied in successive steps, the sum of the considered frequency components is effectively an alternative to the fitting by Equation (4.1). The advantage of Equation (4.2) is that it allows for extracting different frequency components in a computationally efficient way by simply applying different observation window sizes, where the length of each window

defines the fundamental period of each extracted component. For example, if Equation (4.1) is applied on a time series (whose sampling frequency is, e.g., per 30 minutes) length of one calendar year, it will provide annual, seasonal, monthly, weekly and daily components, where annual component is fundamental, and other mentioned components are harmonics. However, this approach will result in equal daily component for all days, which is fitted based on the average of all days. On the other hand, Equation (4.2) can be applied separately for each day (window length of one day), then for the week to which that day belongs (window length of one week), then for the corresponding month (window length of one month), then corresponding season (window length of one season) and finally for year to which that day belongs (window length of one year), providing considered frequency components for the day of interest. Afterwards, this procedure can be repeated for the next day, and so on. A comparison between Fourier series decomposition by Equation (4.1) and by applying selected components using Equation (4.2) is provided later in this section.

If the exact decomposition through the infinite Fourier series in Equation (4.1) is not of interest, successive application of Equation (4.2) allows to obtain selected or targeted frequency components, e.g., to disaggregate total demand in periodically changing daily, weekly and seasonal components. In a matrix notation, it can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \Rightarrow \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(N) \end{bmatrix} = \begin{bmatrix} 1 & \cos\left(\frac{2\pi}{N_x} \times 1\right) & \sin\left(\frac{2\pi}{N_x} \times 1\right) \\ 1 & \cos\left(\frac{2\pi}{N_x} \times 2\right) & \sin\left(\frac{2\pi}{N_x} \times 2\right) \\ \vdots & \vdots & \vdots \\ 1 & \cos\left(\frac{2\pi}{N_x} \times N\right) & \sin\left(\frac{2\pi}{N_x} \times N\right) \end{bmatrix} \begin{bmatrix} c_x \\ a_x \\ b_x \end{bmatrix} \quad (4.3)$$

which is equivalent to a regression problem. The value of $\boldsymbol{\beta} = [c_x, a_x, b_x]^T$ can be estimated using ordinary least squares (OLS), whose objective is to minimise the sum of squared residuals. The final results of OLS for $\boldsymbol{\beta}$ can be derived as $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. Refs [127, 128] also apply regression techniques in period analysis, i.e., the least absolute shrinkage and selection operator (LASSO) as the regression regularisation. Also, Goertzel algorithm [129], which is a type of discrete Fourier transform (DFT), may be

used if only certain frequencies should be extracted, as in, e.g., real-time applications where short computational times are of high importance.

In terms of disaggregating measured total demand, temperature and solar irradiance time series, only seven predefined frequency components are selected in this chapter: half-daily, daily, two-daily (for 2 weekend days, if considered day is on weekend), five-daily (for 5 working days, if considered day is working day), weekly (for 7 days of a week), monthly (for 28 days of a lunar month), seasonal (for 91 days of a season) and annual (for 364 days of a year, as 365 days will make most of the other frequency components to be interharmonics). The chapter explicitly considers disaggregation of daily (24-hour) demands, where the length of the output window is one day (see Section 4.3) and corresponding daily time series is denoted as D_i . The frequency components are obtained for the considered day of interest, including components with shorter and longer periodicity than one day. The half-daily component is calculated for the considered day as the 2nd harmonic of the Fourier series decomposition applied on the window length of one day (for which daily component is fundamental frequency component). The extension of the observation window allows for the identification of other (lower) frequency components. In that way, all considered frequency components for D_i are extracted, so that the estimations on the considered day are based on the most relevant observations. This is illustrated in Fig. 4.2.

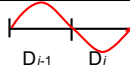
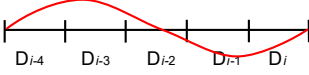
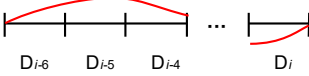
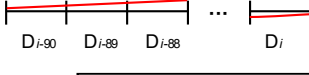
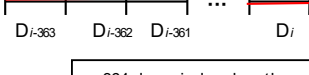
Half-daily and daily frequency components	 1-day window length	Fit with $c_{hd} + c_d + a_h \cos\left(\frac{2\pi}{N_h} n\right) + b_h \sin\left(\frac{2\pi}{N_h} n\right)$ $+ a_d \cos\left(\frac{2\pi}{N_d} n\right) + b_d \sin\left(\frac{2\pi}{N_d} n\right)$	Extracted half-daily and daily frequency components
2-day frequency component	 2-day window length, if considered day is on weekend	Fit with $c_{2d} + a_{2d} \cos\left(\frac{2\pi}{N_{2d}} n\right) + b_{2d} \sin\left(\frac{2\pi}{N_{2d}} n\right)$	Extracted 2- day frequency component
5-day frequency component	 5-day window length, if considered day is working day	Fit with $c_{5d} + a_{5d} \cos\left(\frac{2\pi}{N_{5d}} n\right) + b_{5d} \sin\left(\frac{2\pi}{N_{5d}} n\right)$	Extracted 5- day frequency component
Weekly frequency component	 7-day window length	Fit with $c_w + a_w \cos\left(\frac{2\pi}{N_w} n\right) + b_w \sin\left(\frac{2\pi}{N_w} n\right)$	Extracted weekly frequency component
Monthly frequency component	 28-day window length	Fit with $c_m + a_m \cos\left(\frac{2\pi}{N_m} n\right) + b_m \sin\left(\frac{2\pi}{N_m} n\right)$	Extracted monthly frequency component
Seasonal frequency component	 91-day window length	Fit with $c_s + a_s \cos\left(\frac{2\pi}{N_s} n\right) + b_s \sin\left(\frac{2\pi}{N_s} n\right)$	Extracted seasonally frequency component
Annual frequency component	 364-day window length	Fit with $c_a + a_a \cos\left(\frac{2\pi}{N_a} n\right) + b_a \sin\left(\frac{2\pi}{N_a} n\right)$	Extracted annual frequency component

Fig. 4.2. Window lengths for Fourier decomposition into seven predefined considered frequency components.

It should be noted that there are several ways how observation windows could be extended: the one illustrated in Fig. 4.2 is by going backwards in the historical data, as this is generally suitable for load forecasting applications, where the future realisations of the process are unknown and where using the future data to extend the observation window may cause a “look-ahead bias” [130]. For general analysis and characteristics or features extraction, as well as hindcasting applications, observation windows can be extended differently, e.g., by either putting day of interest in the centre of the historical data, or at the beginning of the historical data.

Examples of the results of the frequency component disaggregation of the AMPds2 total demand and temperature data are shown in Figs. 4.3 – 4.5. The “rank” of the x-

axis label in those figures represents the index of the time series data regarding the current unit. For example, in Fig. 4.3(c), the data point with x-axis value equal to 36 means it is the 36th hour of a 5-daily window, and in Fig. 4.3(f), the recording whose x-axis value is 1 represents the 7th day of a season. The first eight plots in Fig. 4.3(a)-(h) show the selected frequency components, following the procedure illustrated in Fig. 4.2. The DC component from Fig. 4.3(h) is calculated as a difference between the mean value of the original daily time series and the mean value of the summed those frequency components (without DC components) in Fig. 4.3(a)-(g). It can be seen that frequency components of total demand and temperature exhibit different correlations, both in terms of the full-length observation window, but also in terms of the increasing or decreasing trends/gradients on the considered day, which are discussed in more detail in the following section. Fig. 4 depict the magnitudes and phase angles of the considered frequency components of the total demand and temperature time series, while Fig. 6 shows the reconstructions of the non-disaggregated total demands and temperatures using only considered frequency components, i.e., by applying Equations (4.2)-(4.3) in Fig. 6(a), and by applying Equation (4.1) in Fig. 6(b).

Fig. 6(a) combines and sums up only considered frequency and DC components in Fig. 4.3(a)-(h) (denoted as $c_h, c_d, \dots, c_a$ in Fig. 4.2). The blue solid line and red dashed line represent original and reconstructed power time series, respectively. The black dotted and green dash-dot lines are original and reconstructed temperature data, respectively. It can be seen in Fig. 6(a) that the original non-disaggregated time series cannot be reconstructed with 100 % accuracy, which is expected, as only a limited number of frequency components is used for the analysis. The considered frequency components can capture general increasing and decreasing changes during the considered day, therefore preserving an important information in the approximation of the original time series. By comparing results for considered day in Fig. 6(a) and Fig. 6(b), it can be seen that the Fourier series decomposition that extracts all frequency components by Equation (4.1) results in a less accurate reconstruction of the original time series than the proposed method. The biggest differences are in capturing morning peak in demand and unusual temperature increase during the evening time.

A numerical comparison between Fig. 6(a) and Fig. 6(b) is presented in Table 4.1. The MAE and RMSE, as well as the absolute and percentage errors between the actually measured demands and reconstructed demands in terms of the overestimated, underestimated and total energy (denoted as E_O , E_U and E_T , respectively) are used to evaluate the performance of the models [131], with their calculation formulas and discussions given in Section 2.3.1 and Section 3.4.4. In assessing the energy-based indices, it is assumed that the average half-hourly demand values are constant over the period of 30 minutes, so that the energy can be calculated as the power multiplied by the time. The results obtained by both conventional Fourier transform Equation (4.1) and proposed FSR method provide accurate estimations for mean values, and their total errors are zero. However, the proposed FSR method have lower MAE, RMSE, and absolute overestimated and underestimated energy values.

Similarly, the results for the total demand and temperature recordings of the MV level datasets are shown in Figs. 4.6 – 4.7, while Table 4.2 shows the comparisons between the proposed method and the conventional Fourier transform.

The same Fourier series regression method is also applied to analyse the total load and solar irradiance time series, as shown in examples in Figs. 4.9 – 4.11 and Figs. 4.12 – 14 with Table 4.3 and Table 4.4 listing the comparison between the FSR method and the conventional Fourier transform. The definitions for the energy-based metrics, E_O , E_U and E_T can be found in Section 3.4.4. The presented results suggest that the proposed FSR method can reconstruct the original demand more accurately than the conventional Fourier transform, while also providing more informative results for the considered frequency components of interest.

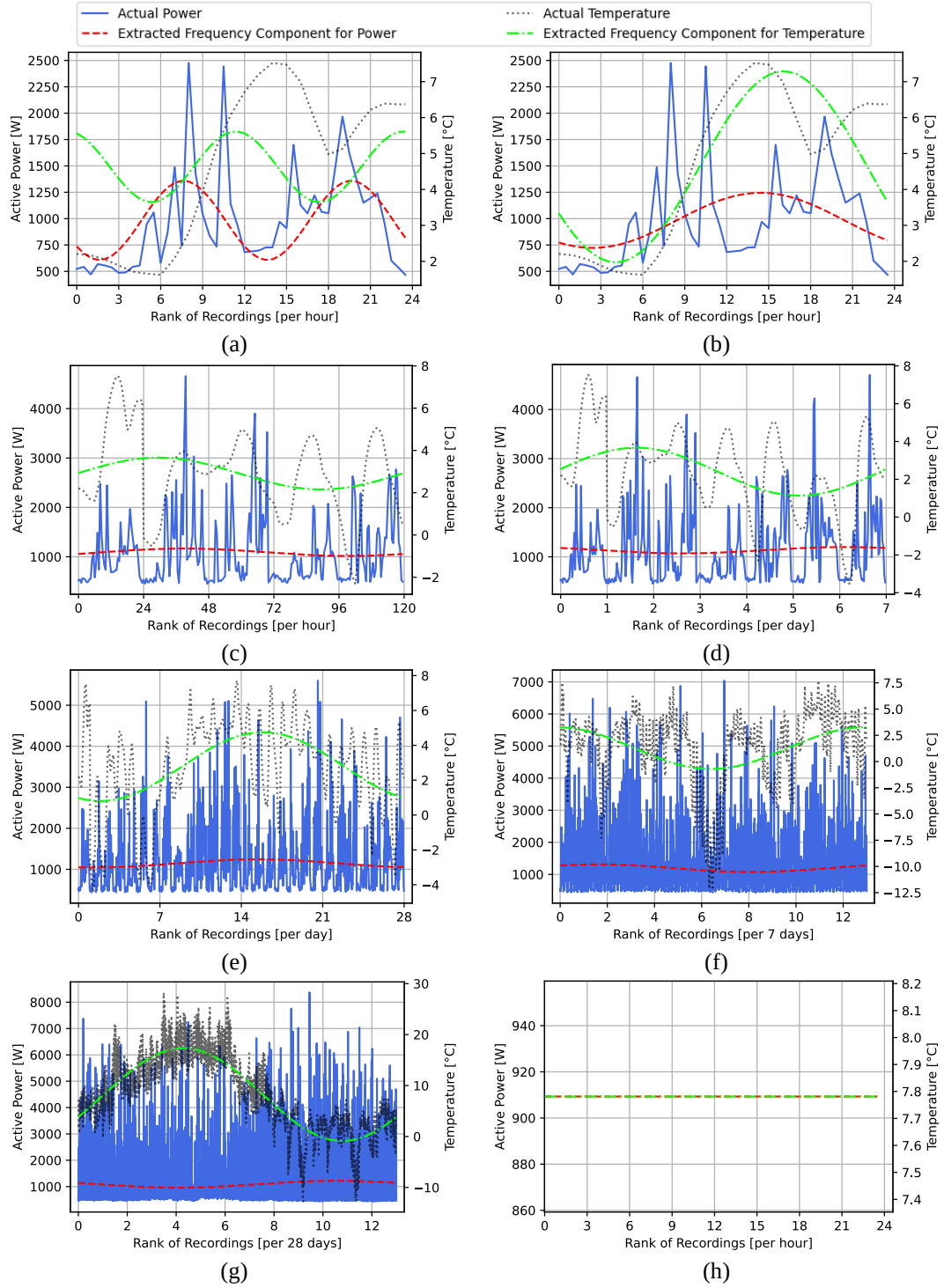


Fig. 4.3. Components of the proposed Fourier series regression on AMPDs2 (Monday, 24 March 2014) for load and temperature recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.

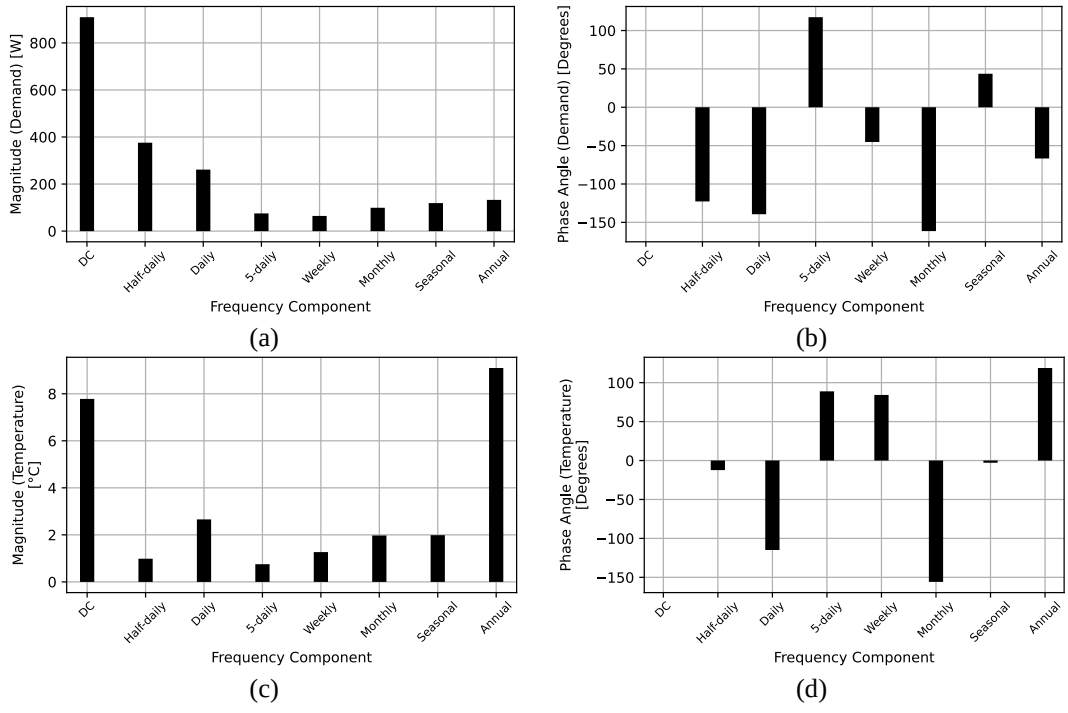


Fig. 4.4. Results of the proposed Fourier series regression on AMPds2 (Monday, 24 March 2014) for load and temperature recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of temperature frequency components, (d) phase angles of temperature frequency components.

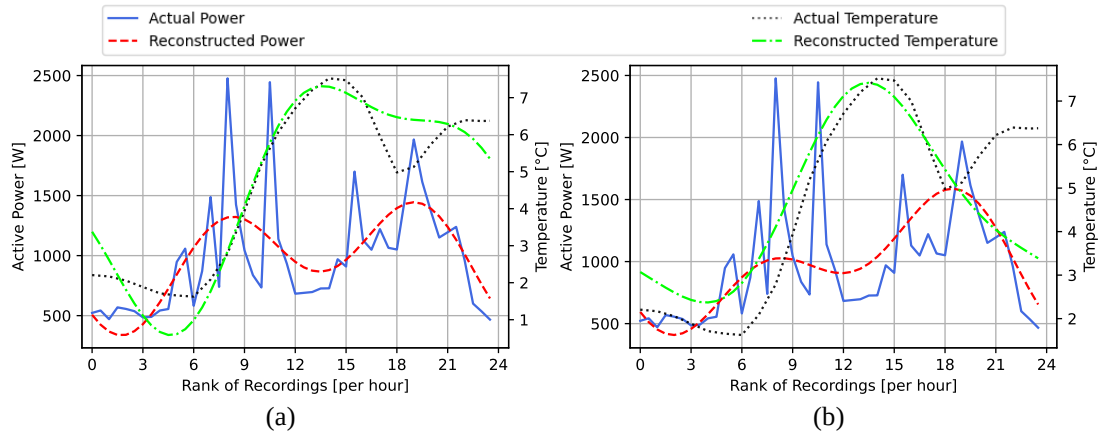


Fig. 4.5. Reconstructions using the proposed Fourier series regression on AMPds2 (Monday, 24 March 2014) for load and temperature recordings: (a) reconstruction of non-disaggregated demand and temperature daily profiles using considered frequency components; (b) reconstruction of non-disaggregated demand and temperature daily profiles using conventional Fourier transform.

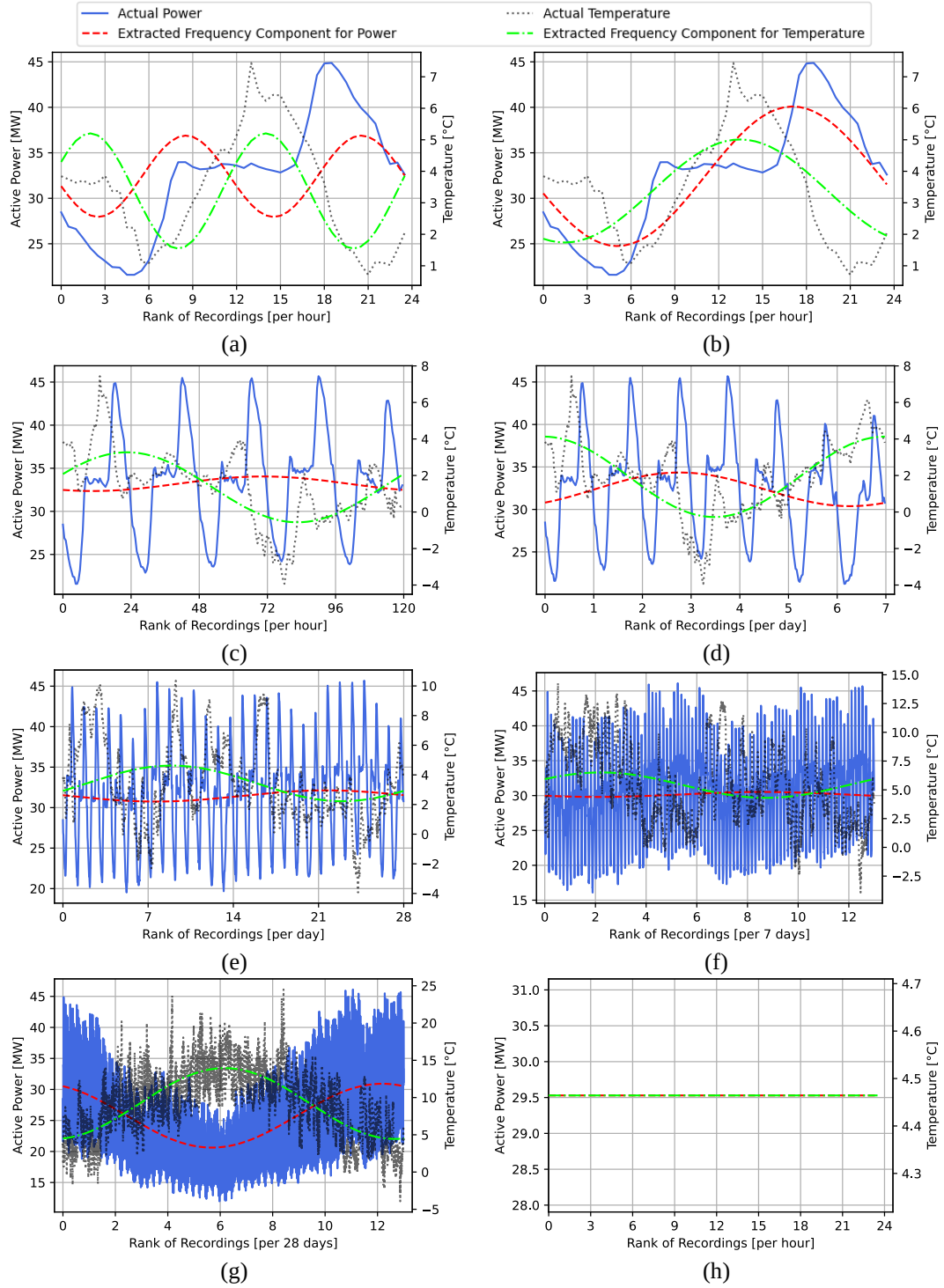


Fig. 4.6. Components of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and temperature recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.

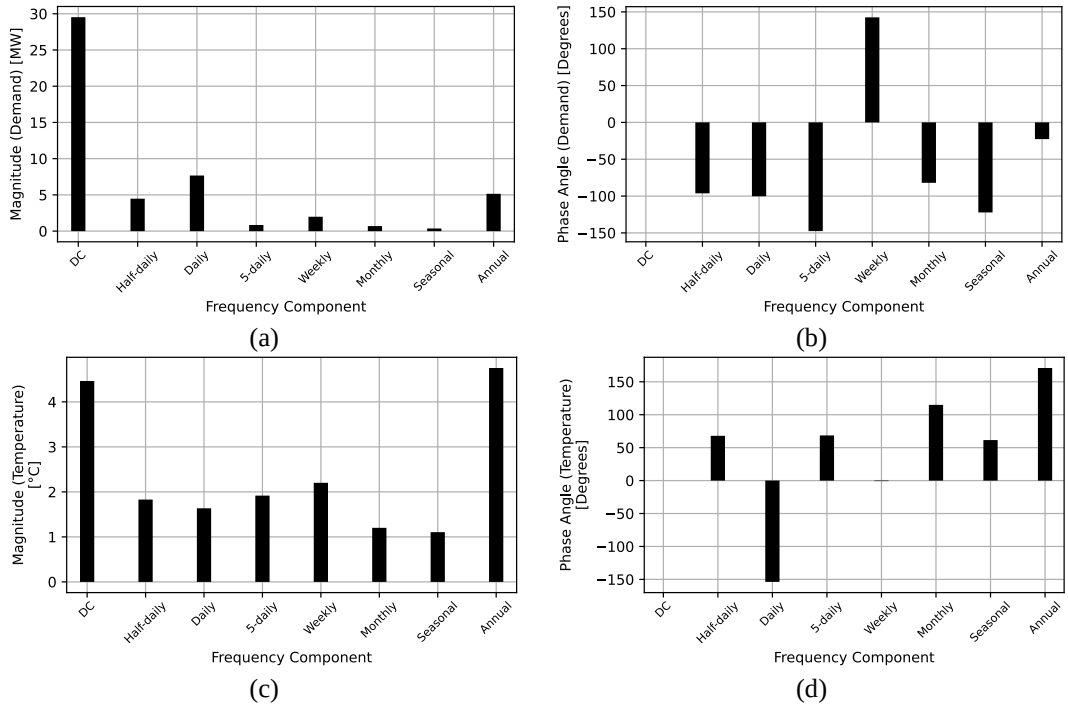


Fig. 4.7. Results of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and temperature recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of temperature frequency components, (d) phase angles of temperature frequency components.

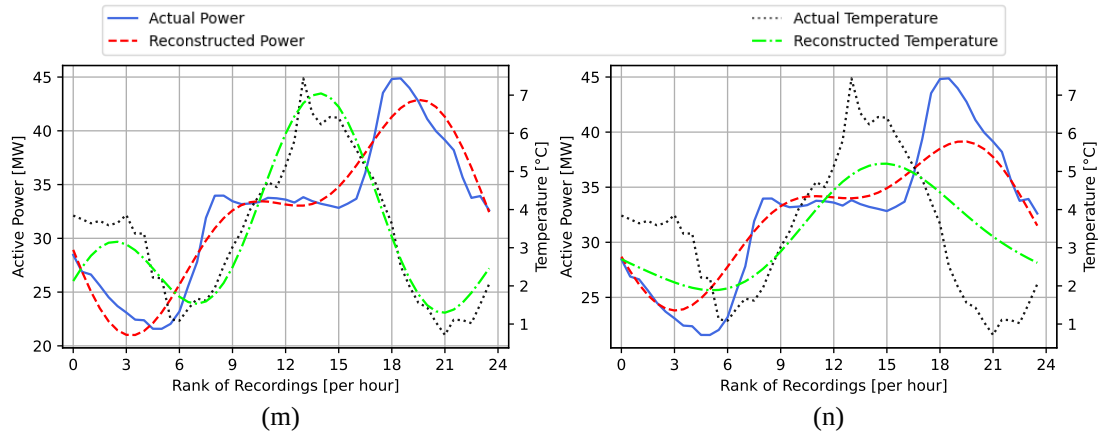


Fig. 4.8. Reconstructions using the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and temperature recordings: (a) reconstruction of non-disaggregated demand and temperature daily profiles using considered frequency components; (b) reconstruction of non-disaggregated demand and temperature daily profiles using conventional Fourier transform.

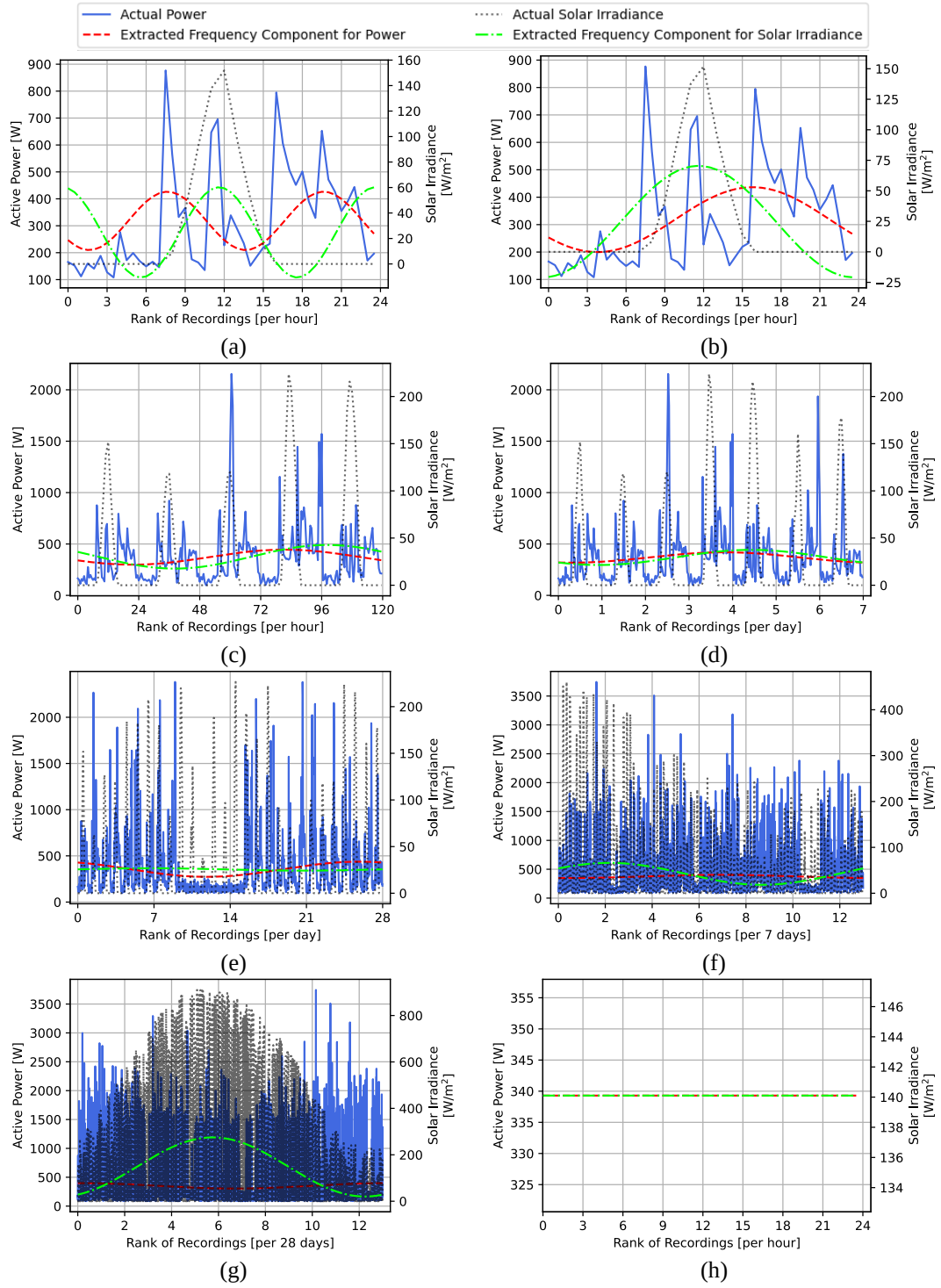


Fig. 4.9. Components of the proposed Fourier series regression on UK-DALE (Monday, 11 January 2016) for load and solar irradiance recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.

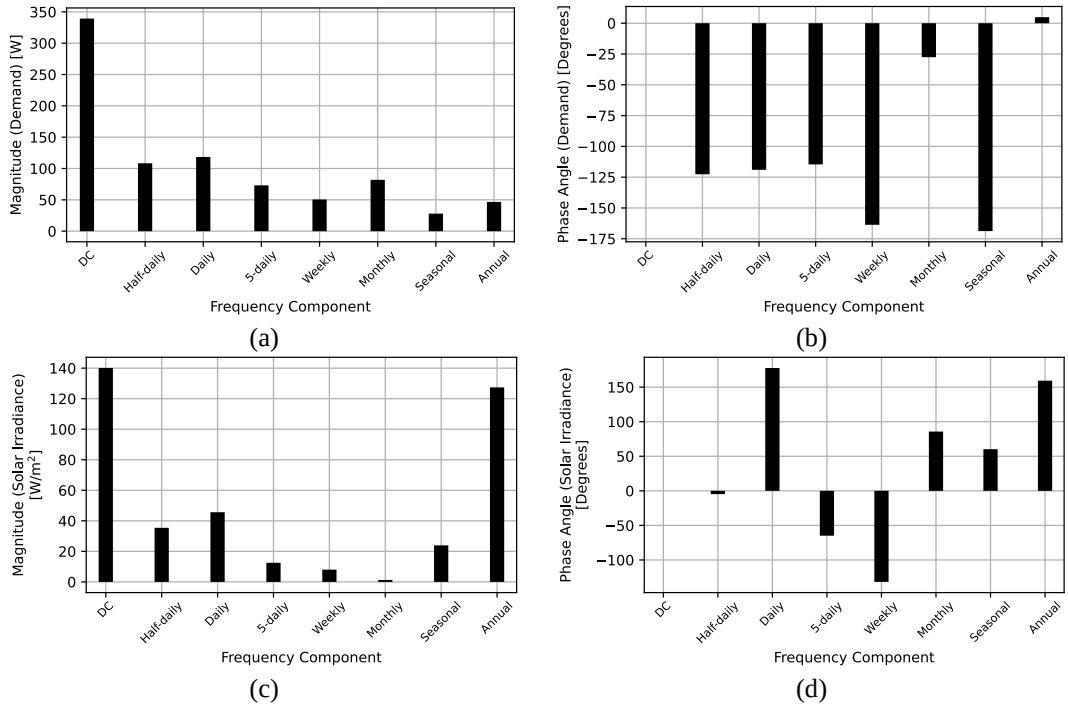


Fig. 4.10. Results of the proposed Fourier series regression on UK-DALE (Monday, 11 January 2016) for load and solar irradiance recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of solar irradiance frequency components, (d) phase angles of solar irradiance frequency components.

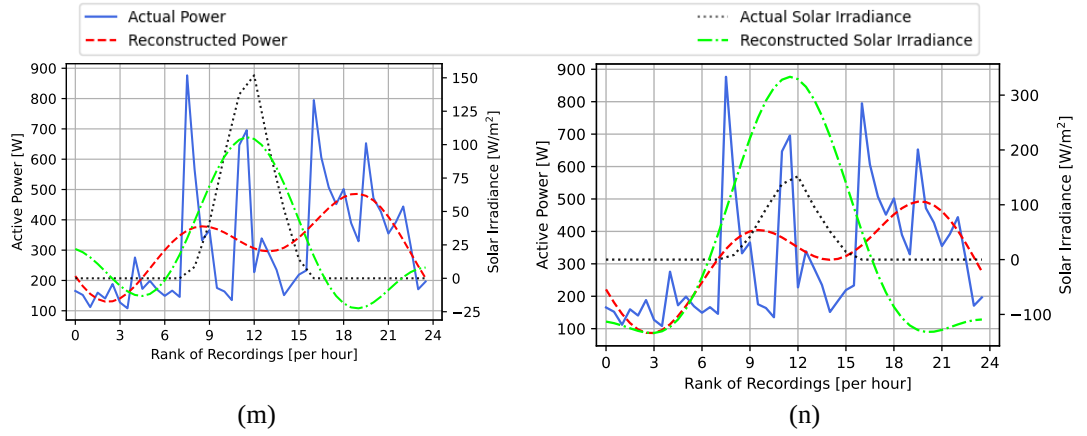


Fig. 4.11. Reconstructions using the proposed Fourier series regression on UK-DALE (Monday, 11 January 2016) for load and solar irradiance recordings: (a) reconstruction of non-disaggregated demand and solar irradiance daily profiles using considered frequency components; (b) reconstruction of non-disaggregated demand and solar irradiance daily profiles using conventional Fourier transform.

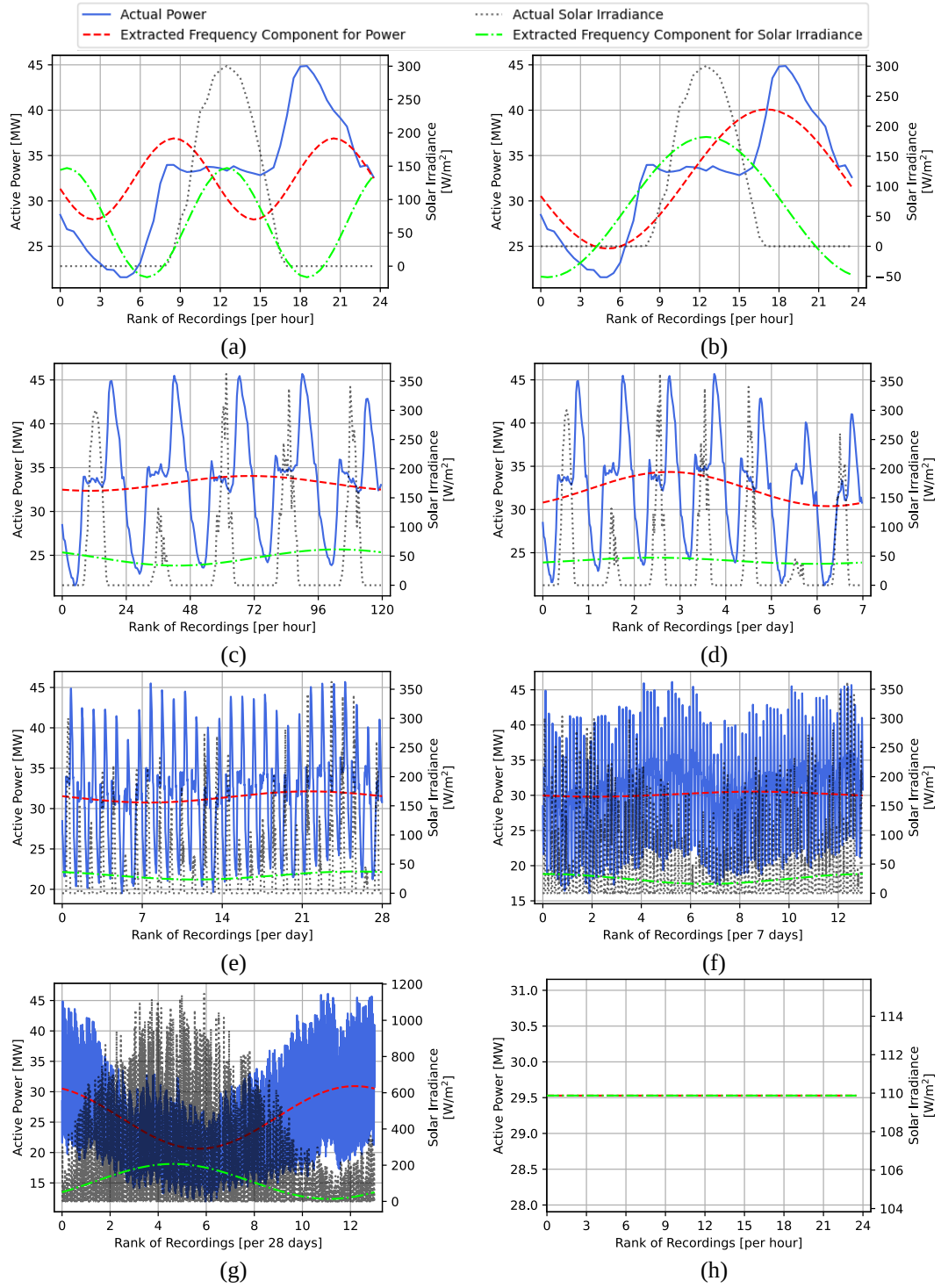


Fig. 4.12. Components of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and solar irradiance recordings: (a) half-daily, (b) daily, (c) 5-daily, (d) weekly, (e) monthly, (f) seasonal, (g) annual, (h) re-estimated DC component.

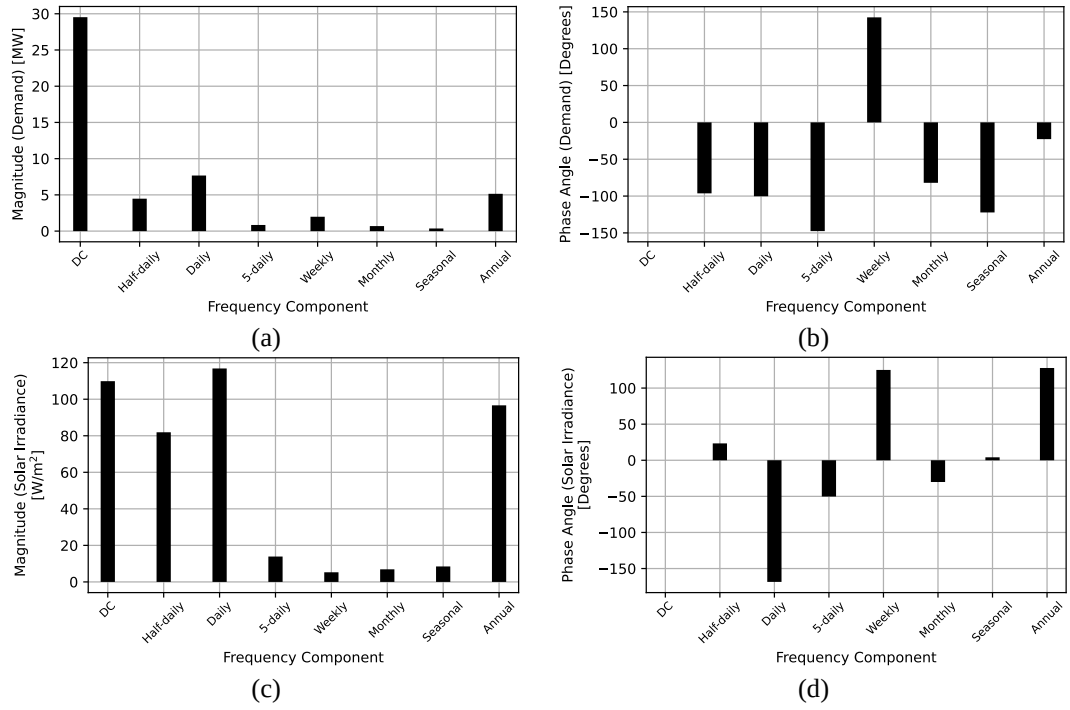


Fig. 4.13. Results of the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and solar irradiance recordings: (a) magnitudes of demand frequency components, (b) phase angles of demand frequency components, (c) magnitudes of solar irradiance frequency components, (d) phase angles of solar irradiance frequency components.

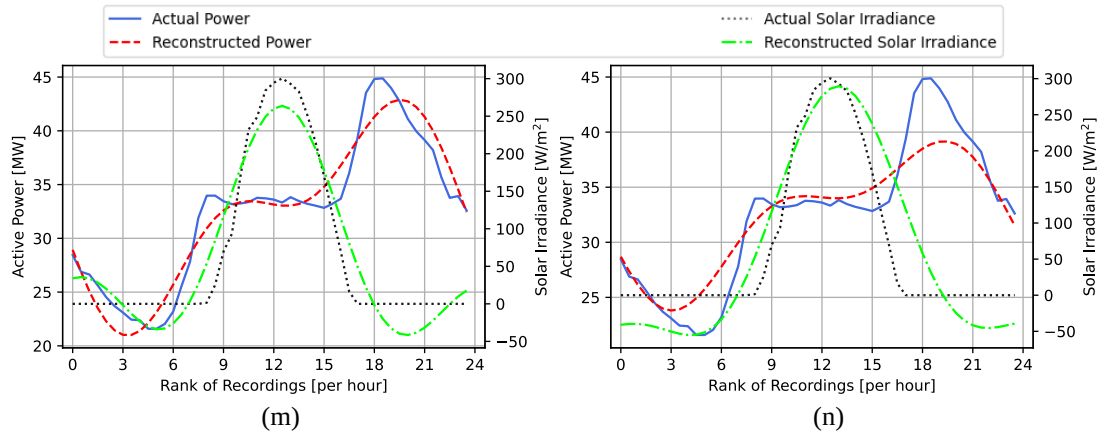


Fig. 4.14. Reconstructions using the proposed Fourier series regression on MV level dataset (Monday, 6 February 2012) for load and solar irradiance recordings: (a) reconstruction of non-disaggregated demand and solar irradiance daily profiles using considered frequency components; (b) reconstruction of non-disaggregated demand and solar irradiance daily profiles using conventional Fourier transform.

Table 4.1 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.3 – 4.5

Temperature Reconstruction MAE [°C] RMSE [°C]			Demand Reconstruction MAE [W] RMSE [W] E _O [Wh] E _O [%] E _U [Wh] E _U [%] E _T [Wh] E _T [%]								
Conv. Fourier	0.901	1.182	Conv. Fourier	253.281	383.512	3039.368	24.736	3039.368	26.850	0.000	0.000
Proposed FSR	0.506	0.652	Proposed FSR	233.751	347.824	2805.014	25.009	2805.014	22.638	0.000	0.000

Table 4.2 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.6 – 4.7

Temperature Reconstruction MAE [°C] RMSE [°C]			Demand Reconstruction MAE [MW] RMSE [MW] E _O [MWh] E _O [%] E _U [MWh] E _U [%] E _T [MWh] E _T [%]								
Conv. Fourier	1.085	1.287	Conv. Fourier	1.766	2.427	21.192	5.131	21.192	5.806	0.000	0.000
Proposed FSR	0.464	0.567	Proposed FSR	1.447	1.768	17.361	4.507	17.361	4.419	0.000	0.000

Table 4.3 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.8 – 4.11

Solar Irradiance Reconstruction MAE [W/m ²] RMSE[W/m ²]			Demand Reconstruction MAE [W] RMSE [W] E _O [Wh] E _O [%] E _U [Wh] E _U [%] E _T [Wh] E _T [%]								
Conv. Fourier	121.469	131.259	Conv. Fourier	117.872	160.147	1414.469	41.507	1414.469	33.464	0.000	0.000
Proposed FSR	15.713	19.009	Proposed FSR	110.304	155.998	1323.643	39.530	1323.643	30.881	0.000	0.000

Table 4.4 Comparison of Errors of Time Series Reconstruction using Conventional Fourier Transform (“Conv. Fourier”) and Proposed Fourier Series Regression (“Proposed FSR”) for the Same Day in Figs. 4.12 – 4.14

Solar Irradiance Reconstruction	MAE [W/m ²]	RMSE [W/m ²]	Demand Reconstruction	MAE [MW]	RMSE [MW]	E _O [MWh]	E _O [%]	E _U [MWh]	E _U [%]	E _T [MWh]	E _T [%]
Conv. Fourier	43.117	51.570	Conv. Fourier	1.766	2.427	21.192	5.131	21.192	5.806	0.000	0.000
Proposed FSR	27.481	31.472	Proposed FSR	1.447	1.768	17.361	4.507	17.361	4.419	0.000	0.000

4.2.3 Extraction of the Strongly Correlated Components

From the previous frequency component disaggregation results, it is clear that for a particular day, D_i , the correlations between the total demand and temperature/solar irradiance are different for different frequency components. The heating load is assumed to be negatively correlated with the temperature, because when the temperature decreases (e.g., in winter), there is expected to be an increase in demand of heating devices, and when the temperature becomes higher (e.g., in summer), the heating demand decreases. Similarly, the lighting load and solar irradiance are assumed to follow a negative correlation. As a comparison, other types of load may not change too much with the changes of temperature or solar irradiance. As the non-disaggregated total demand consists of the contributions of different types of the load, including heating load and lighting load, then disaggregated frequency components of the total demand with such strong negative correlations should contain significant part of heating and lighting load. In other words, frequency components with strong negative correlations should help in the identification of heating and lighting loads. Accordingly, further text is related to the analysis of these correlations.

The two most common metrics for correlational analysis are Pearson and Spearman correlation coefficients. As Pearson coefficient is a measure of linear correlation, this chapter uses Spearman coefficient, ρ_{xy} , which can capture non-linear dependencies [132] to calculate and compare the correlations between total demand and temperature/solar irradiance:

$$\rho_{xy} = \frac{\text{cov}(r_x, r_y)}{\sigma_{r_x} \sigma_{r_y}} \quad (4.4)$$

where: r_x and r_y are the ranks of the two analysed data/time series, respectively; $\text{cov}(\cdot)$ means covariance; and σ represents the standard deviation. For the example results shown in Fig. 4.3, Fig. 4.6, Fig. 4.9, and Fig. 4.12, the Spearman correlation coefficients between the measured demand and temperature/solar irradiance time series on the considered days are calculated as 0.2323, -0.0579, 0.1791 and 0.1168, respectively. Note that the longer-term frequency components (with lower than daily frequencies) are obtained by extended observation windows, but only the

considered/first days of longer windows are used in the calculations. The results for the correlational analysis of considered frequency components for the same days in Fig. 4.3, Fig. 4.6, Fig. 4.9, and Fig. 4.12 are presented in Table 4.5, where bold values indicate stronger negative correlations. Since the original time series and the FSR results may be quite different on various days/datasets, and one identified more negatively correlated frequency component may be different, or may even increase the positive correlation on the other days, it is important to analyse every individual day and dataset separately.

Table 4.5 Spearman Correlation Coefficients between the Frequency Components of Total Load and Temperature/Solar Irradiance in Example Results (D_i only)

	Half Daily	Daily	5 Days'	Weekly	Monthly	Seasonal	Annual
Fig. 4.3	-0.3677	0.8849	1.0000	-1.0000	1.0000	-1.0000	-1.0000
Fig. 4.6	-0.9145	0.5363	0.2553	-1.0000	-1.0000	-1.0000	-1.0000
Fig. 4.9	-0.4814	0.4251	0.9946	-0.7414	-1.0000	-1.0000	1.0000
Fig. 4.12	-0.4816	0.3672	-0.2561	1.0000	1.0000	-1.0000	-1.0000

It can be seen that the frequency components may have stronger negative correlations, i.e., the correlation value of a specific frequency component is smaller. For example, for the day in Fig. 4.3, the correlation between the original demand and temperature time series is 0.2323, but the correlations between their half daily, weekly, seasonal and annual components are -0.3677, -1.0000, -1.0000 and -1.0000, respectively. Summing up those components (the DC components are the same, as if using all frequency components) on the considered day will result in a new time series, whose characteristics, e.g., increasing or decreasing trends, can imply the changes in heating and lighting loads, with example results shown in Fig. 4.15 and Fig. 4.16, where “actual power” and “actual temperature/solar irradiance” are original aggregated and total power and temperature/solar irradiance, respectively. Although the reconstructed daily load profiles are not as accurate as if all considered frequency components are used (e.g., these with positive correlations), the Spearman correlation coefficients between the time series reconstructions using the combination of the selected frequency components with negative correlations are now -0.2210, -0.9679, -0.4798 and -0.4568 in Fig. 4.15(a), Fig. 4.15(b), Fig. 4.16(a) and Fig. 4.16(b), respectively, i.e., all feature stronger negative correlations compared to the results using original/measured daily time series (i.e., 0.2323, -0.0579, 0.1791 and 0.1168). As the

starting assumption/hypothesis is that the combination of the selected demand frequency components with stronger negative correlations with temperature and solar irradiance will contain significant portions of the heating and lighting loads, respectively, the new time series of these negatively correlated frequency components are used as an additional information for the heating and lighting load disaggregation models presented in the next section.

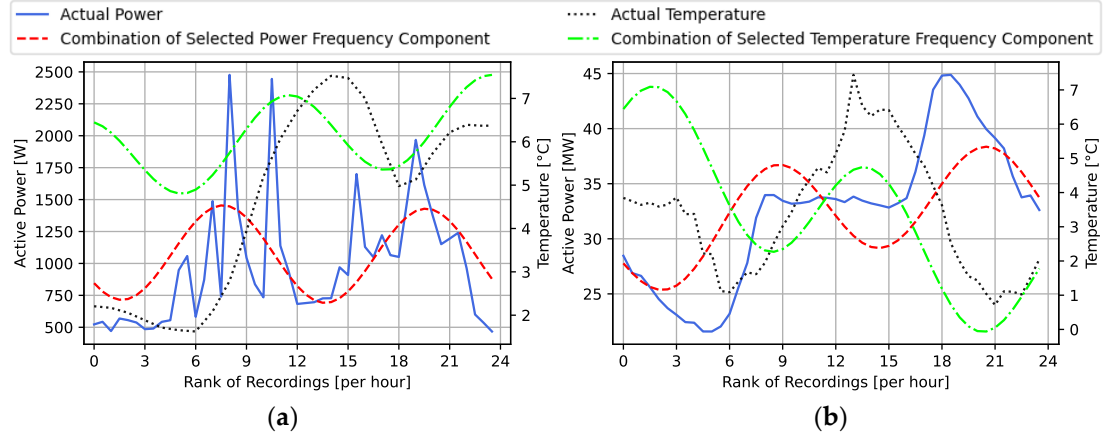


Fig. 4.15. Combinations of more negatively correlated frequency components for demand and temperature: (a) AMPds2 (Monday, 24 March 2014). (b) MV level dataset (Monday, 6 February 2012).

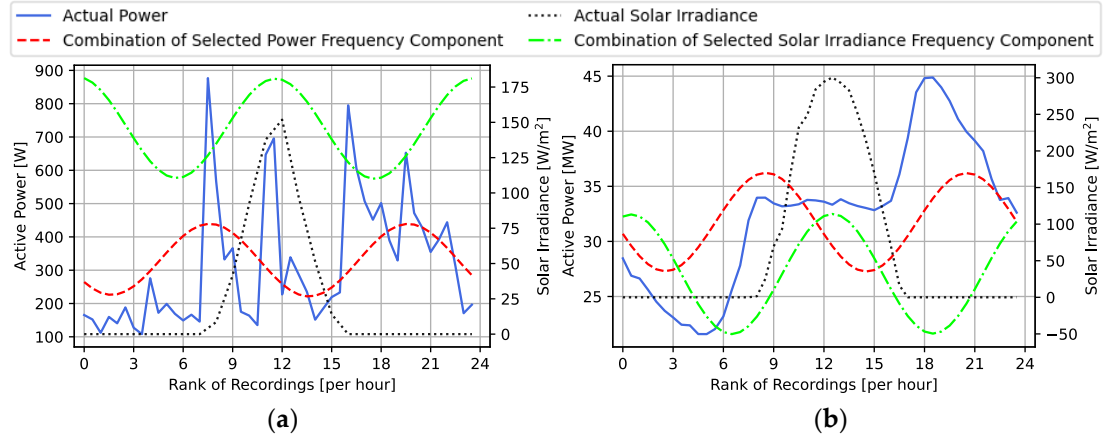


Fig. 4.16. Combinations of more negatively correlated frequency components for demand and solar irradiance: (a) UK-DALE (Monday, 11 January 2016). (b) MV level dataset (Monday, 6 February 2012).

4.3 Neural Network-Based Load Disaggregation Model

This section presents the load disaggregation model used in this chapter, which is a neural network model, combining two widely used NN approaches: 1) recurrent neural

network; and 2) convolutional neural network. The detailed theoretical description about the RNN and CNN can be found in Section 2.2.1.

4.3.1 Dataset Splitting and Model Input and Output Variables

Before entering the time series predictor variables into the model, the frequency domain information is firstly extracted and analysed, followed by the Fourier series regression method from Section 4.2. The NN model utilises the embedded patterns from both time domain and frequency domain and provides the final estimations for the disaggregated load. In addition, since hyperparameters of NN model usually have strong impact on the performance, but their optimal values are not known a priori, Bayesian optimisation is used to guide the tuning of model hyperparameters.

In available datasets, AMPds2 has actual measurements for total load and heating load, while UK-DALE has measurements for total load and lighting load. Therefore, these two datasets are used to test the load disaggregation NN model. Each dataset is divided into training, validation and test sub-sets. More specifically, in AMPds2, the measurements between 23 February 2014 and 31 March 2014 are used as test sub-set, and in UK-DALE, the recordings between 5 January 2016 and 17 March 2016 are used as test sub-set. For both datasets, the rest of available data are used as the training and validation sub-sets, in the ratio 90 % to 10 %.

The inputs of the NN models are: 1) daily time series of the total active and reactive power demands; 2) synchronous daily time series of temperature and solar irradiance; 3) corresponding time and calendar variables, including month, day of week, holiday indicator (Boolean variable) and daylight-saving time indicator (Boolean variable); and 4) synchronous daily time series of combinations of more negatively correlated frequency components of demand, temperature and solar irradiance. All individual input features are treated as column vectors, i.e., columns that form the input data matrix. Adding of any additional predictor variable (e.g., synchronous combinations of more negatively correlated frequency components) is simply performed by appending a new column to the existing input matrix. The outputs of the load disaggregation models are heating load profile and lighting load profile on the considered day, which is the same day as the inputs for AMPds2 and UK-DALE,

respectively. All continuous features are linearly scaled to a range between 0 and 1, and all categorical features are transformed using one hot encoder.

As previously mentioned in Section 2.4.2, deep learning and BiLSTM models are increasingly used for load disaggregation, as they can process long time series of data with complex patterns and identify target load profiles. As these models are often approached as “black box” modelling methods, the CNN-BiLSTM model proposed in this chapter uses FSR-based frequency component decomposition as the additional explanatory variable/feature, in order to provide more accurate load disaggregation.

In the proposed CNN-BiLSTM load disaggregation model, the features in the predictors are firstly extracted with the CNN architecture, and outputs are padded with zeros evenly to the left and right, so that they have the same width dimension as the inputs, before they enter into the two stacked layers of BiLSTM. The convolution operation could reduce the variation in the input feature, what will then increase the temporal learning accuracy of the BiLSTM model [133]. Every layer is followed by a dropout layer to increase the generalisation ability and reduce overfitting, which refers to random dropping-out of hidden and visible neurons in a certain layer [43]. The architecture of the CNN-BiLSTM is depicted in Fig. 4.17, which also shows the hyperparameters (HPs) within the topology, which will be disused/optimised further in Section 4.3.2.

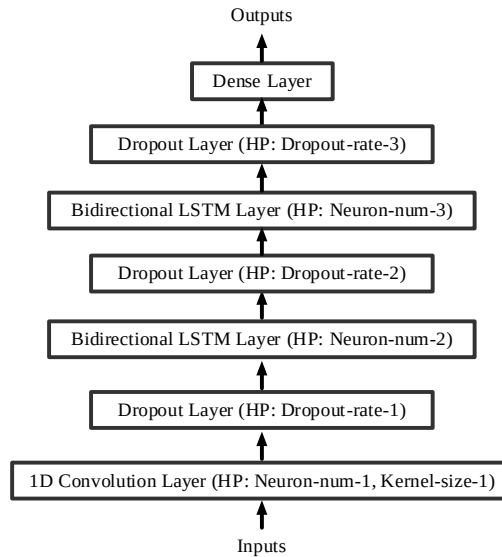


Fig. 4.17. Proposed CNN-BiLSTM architecture and hyperparameters.

4.3.2 Model Hyperparameter Tuning: Bayesian Optimisation

The optimal values of the NN model parameters are unknown usually and may change in different cases, even with only small difference in the available training set. Furthermore, the final results may also be sensitive to the selected hyperparameters, which brings additional uncertainty. To build and begin training the model, the hyperparameters must be specified, but the values are usually unknown and most of them are defined in a too wide range and lead to a big uncertainty. For example, the dropout rate, in theory, can be any numeric value from 0.0 to 1.0, and the number of neurons may vary from several to thousands or even more).

As the hyperparameters cannot be optimised through the training process, an optimisation algorithm should be implemented. In this chapter, Adam optimiser [120] is used to find the best model hyperparameters (i.e., weight and bias values for every neuron) minimising the error function, since they are fixed once the NN model is instantiated. There are many types of hyperparameters in the proposed CNN-BiLSTM model, including the number of neurons (neuron-num) for trainable layers (CNN and BiLSTM), size of the kernel in CNN (kernel-size), and dropout rates for the dropout layer (dropout-rate), which represents the probability of dropout for the neurons in the previous layer. In addition, the learning rate for Adam optimiser (Adam-learning-rate) is also one of the model hyperparameters.

Random search or brute force looping for the combinations of different hyperparameter values can find the most appropriate hyperparameters in theory, but it requires significant computational times, since it does not use the information of previous trials and due to the “curse of dimensionality”. Instead, Bayesian optimisation is used, which is a heuristic searching method that uses Gaussian process as a surrogate model and provides exploitation-exploration trade-off through an acquisition function [134]. Bayesian optimisation results provide a set of hyperparameter candidates that can make the model performs relatively well, therefore it reduces the uncertainty and selection space of hyperparameters (used to be a possible infinite range).

For the NN model hyperparameters denoted as $\mathbf{x}$, and for the task of searching the best $\mathbf{x}$ for the NN model denoted as $f(\mathbf{x})$, where $f(\cdot)$ is an objective/cost black box function,

its evaluation requires actual training and validation of the NN model, and the optimisation is to find $\mathbf{x} = \text{argmax}_{\mathbf{x}}(f(\mathbf{x}))$. From the view of Bayesian optimisation, $f(\cdot)$ can be described as a Gaussian process (GP), i.e., a distribution over functions $y = f(\mathbf{x}) \sim \text{GP}$, appreciating the uncertainty, which can be formulated as:

$$P(y_{t+1}|D_{1:t}, \mathbf{x}_{t+1}) = N(\mu_t(\mathbf{x}_{t+1}), \sigma_t^2(\mathbf{x}_{t+1})) \quad (4.5)$$

$$\mu_t(\mathbf{x}_{t+1}) = \mathbf{k}^T \mathbf{K}^{-1} \mathbf{y}_{1:t} \quad (4.6)$$

$$\sigma_t^2(\mathbf{x}_{t+1}) = k(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \quad (4.7)$$

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & k(\mathbf{x}_1, \mathbf{x}_t) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \dots & k(\mathbf{x}_2, \mathbf{x}_t) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_t, \mathbf{x}_1) & k(\mathbf{x}_t, \mathbf{x}_2) & \dots & k(\mathbf{x}_t, \mathbf{x}_t) \end{bmatrix} \quad (4.8)$$

$$\mathbf{k} = [k(\mathbf{x}_{t+1}, \mathbf{x}_1) \quad k(\mathbf{x}_{t+1}, \mathbf{x}_2) \quad \dots \quad k(\mathbf{x}_{t+1}, \mathbf{x}_t)]^T \quad (4.9)$$

where: $P(y_{t+1}|D_{1:t}, \mathbf{x}_{t+1})$ is the posterior GP given observations $D_{1:t} = ((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_t, y_t))$ and $\mathbf{x}_{t+1}$. $\mu_t(\cdot)$ and $\sigma_t^2(\cdot)$ are the mean and variance functions, respectively; $k(\cdot)$ is kernel function modelling the dependence among GP random variables, and $\mathbf{K}$ is the kernel matrix (i.e., covariance matrix), which should be positive definite. The choice of kernel is not a trivial problem, as it determines almost all generalisation properties of a GP. Commonly used kernel functions include squared exponential kernel, rational quadratic kernel, and matern kernel (used in this chapter), while further discussions and comparisons can be found in [134]. The next trial of hyperparameters, $\mathbf{x}_{t+1}$, is desired to lead to $y_{t+1} = f(\mathbf{x}_{t+1})$ to have both higher mean (exploitation) value and higher variance (exploration). Since they may not be simultaneously improved, a balance is needed and an additional criterion considering both μ_t and σ_t (known as acquisition function) is required. This chapter selects confidence bounds, $\text{UCB}(\mu_t, \sigma_t)$, criterion proposed in [135] to determine the next trial of hyperparameters. In summary, Bayesian optimisation can be described iteratively for $t + 1$ trials as:

1. Fit the GP using $D_{1:t} = ((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_t, y_t))$;
2. Determine $\mathbf{x}_{t+1} = \text{argmax}_{\mathbf{x}_{t+1}} (\text{UCB}(\mu_t(\mathbf{x}_{t+1}), \sigma_t(\mathbf{x}_{t+1})))$;
3. Evaluate $y_{t+1} = f(\mathbf{x}_{t+1})$;

4. Insert $(\mathbf{x}_{t+1}, y_{t+1})$ into $D_{1:t}$ and obtain $D_{1:t+1}$.

The BO method is applied to four instances of load disaggregation models separately and run 32 times each, which are: 1) AMPds2, heating load disaggregation, without FSR and correlation analysis results (denoted as “AMPds2, w/o FR”); 2) same as 1), but with FSR and correlation analysis results (denoted as “AMPds2, w/ FR”); 3) UK-DALE, lighting load disaggregation, without FSR and correlation analysis results (denoted as “UK-DALE, w/o FR”); 4) same as 3), but with FSR and correlation analysis results (denoted as “UK-DALE, w/ FR”). In every trial, the BO is executed 3 times and average of error metrics is used, which can reduce the uncertainty and variance due to training and obtain a more reliable assessment of the considered settings of hyperparameters (thus, there are total $32 \times 3 = 96$ times of BO executions for each model instance). Each BO execution requires maximum 25,000 epochs on the training set, with minimised MSE as objective function. It assesses model performance with current hyperparameter settings on the validation set, which guides the searching direction of BO, so that the next groups of hyperparameters could potentially lead to a model with higher accuracy.

The set constraints of hyperparameters and the results of optimisation of hyperparameters by BO for all model instances are listed in Table 4.6. The constraints format $[start:step:stop]$ means all values between *start* (inclusive) and *stop* (inclusive) with a distance of *step*, and the format $[ele_1, ele_2, ele_3]$ represent predefined set of possible values. It can be seen that in most cases, obtained optimised hyperparameters are different for different model instances. In particular, there are notable differences in the numbers of neurons for the 1D convolution layer. Even though the neuron numbers for two BiLSTM layers are quite different for two AMPds2 heating load disaggregation models, the dropout rates of the followed dropout layers are closer to each other, and similar analysis also applies to UK-DALE lighting load disaggregation models. For all four model instances, the identified most proper kernel size in the 1D convolution layer and the learning rate for Adam optimiser are the same, which are 5 and 0.0005, respectively. The load disaggregation results based on the models with the optimised hyperparameters are shown and compared in Section 4.3.3.

Table 4.6 Selection of Optimal Settings of Hyperparameters by BO for All Load Disaggregation Models

	Constraints	AMPds2, w/o FR	AMPds2, w/ FR	UK-DALE, w/o FR	UK-DALE, w/ FR
Neuron-num-1	[8:1:16]	13	8	9	12
Kernel-size-1	[3:2:7]	5	5	5	5
Dropout-rate-1	[0.05:0.05:0.5]	0.2	0.15	0.25	0.4
Neuron-num-2	[48:8:128]	88	120	80	56
Dropout-rate-2	[0.05:0.05:0.5]	0.45	0.45	0.3	0.3
Neuron-num-3	[48:8:128]	64	80	128	120
Dropout-rate-3	[0.05:0.05:0.5]	0.2	0.2	0.1	0.35
Adam-learning-rate	[0.0005, 0.001, 0.005]	0.0005	0.0005	0.0005	0.005

4.3.3 Load Disaggregation Results and The Benefits of Using Fourier Series Regression

As in previous comparisons of errors of time series reconstruction using conventional Fourier transform and proposed Fourier series regression (see Table 4.1 to Table 4.4), the same error metrics are used to assess the load disaggregation models. The example results of load disaggregation for AMPds2 and UK-DALE are shown in Fig. 4.18, which are for the same days as in Figs. 4.3 – 4.5 and Fig. 4.15(a) (Monday, 24 March 2014) and Figs. 4.9 – 4.11 and Fig. 4.16(a) (Monday, 11 January 2016), respectively. Fig. 4.19 shows the results for the two corresponding weeks. Table 4.7 and Table 4.8 show calculated error indices for these days/weeks, while Table 4.9 shows the overall model performance on the whole test set (i.e., from 23 February 2014 to 31 March 2014 in AMPds2, and between 5 January 2016 and 17 March 2016 for UK-DALE, which are both around 10% of the whole datasets, as explained in Section 4.3.1). In all comparisons, the results obtained by the proposed frequency component-based CNN-BiLSTM model (denoted as “NN w/ FR”) are compared with a naïve CNN-BiLSTM model (without frequency components, denoted as “NN w/o FR”) and additionally benchmarked against two other commonly used NILM models from [136]: combinatorial optimisation and factorial hidden Markov model algorithms (more detailed discussions about CO and FHMM can be found in [136]).

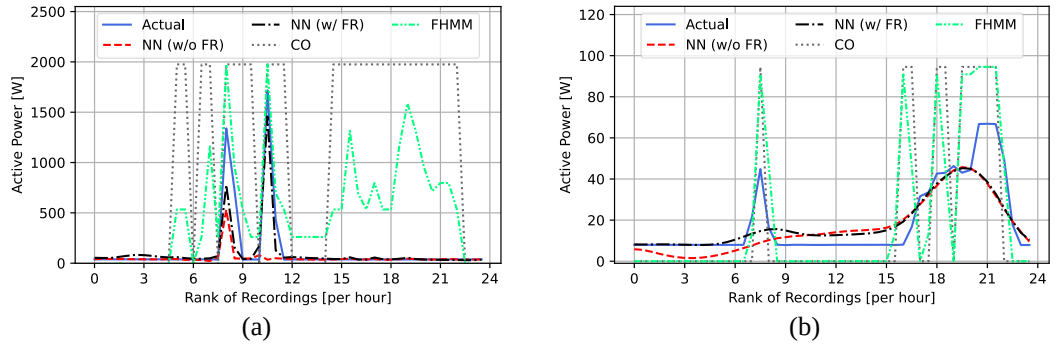


Fig. 4.18. Example load disaggregation results. (a) AMPds2, heating load disaggregation, Monday, 24 March 2014. (b) UK-DALE, lighting load disaggregation, Monday, 11 January 2016.

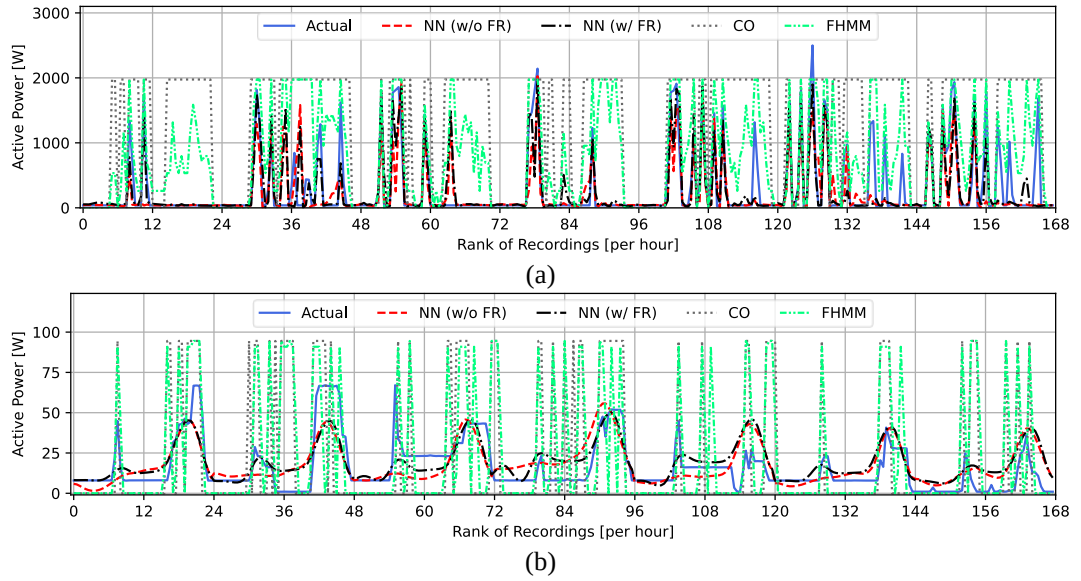


Fig. 4.19. Example load disaggregation results. (a) AMPds2, heating load disaggregation, 24 - 30 March 2014. (b) UK-DALE, lighting load disaggregation, 11 - 17 January 2016.

Table 4.7 Comparison of Performance of Load Disaggregation Models on Two Example Days

		MAE [W]	RMSE [W]	E_o [Wh]	E_o [%]	E_U [Wh]	E_U [%]	E_T [Wh]	E_T [%]
AMPds2	NN, w/o FR	78.295	292.132	80.115	15.857	1798.954	74.131	-1718.839	-58.624
	NN, w/ FR	50.393	131.612	342.627	53.778	866.815	37.772	-524.188	-17.878
	CO	1027.054	1375.749	24129.788	952.398	519.509	130.410	23610.280	805.276
	FHMM	414.752	562.774	9624.049	361.166	330.011	123.492	9294.038	316.992
UK-DALE	NN, w/o FR	11.213	13.616	267.908	75.557	1.213	12.301	266.695	73.18
	NN, w/ FR	6.996	8.77	159.78	52.472	8.126	13.558	151.654	41.613
	CO	23.810	33.603	268.930	434.256	302.508	100.000	-33.578	-9.214
	FHMM	19.924	27.290	203.949	243.845	274.236	97.663	-70.287	-19.287

Table 4.8 Comparison of Performance of Load Disaggregation Models on Two Example Weeks

		MAE [W]	RMSE [W]	E _O [kWh]	E _O [%]	E _U [kWh]	E _U [%]	E _T [kWh]	E _T [%]
AMPds2	NN, w/o FR	166.332	375.100	7.341	121.448	20.602	47.061	-13.261	-26.617
	NN, w/ FR	142.235	318.995	8.449	81.054	15.446	39.205	-6.997	-14.043
	CO	1005.172	1327.605	165.334	367.322	3.535	73.462	161.799	324.747
	FHMM	509.442	731.393	82.869	185.445	2.717	52.897	80.152	160.875
UK-DALE	NN, w/o FR	9.164	12.055	0.887	83.941	0.652	39.020	0.235	8.620
	NN, w/ FR	8.128	11.219	0.931	77.321	0.434	28.485	0.497	18.221
	CO	25.017	36.611	2.486	245.737	1.717	100.000	0.769	28.185
	FHMM	22.692	33.526	2.454	214.682	1.358	85.660	1.096	40.172

Table 4.9 Comparison of Performance of Load Disaggregation Models on the Whole Test Sets

		MAE [W]	RMSE [W]	E _O [kWh]	E _O [%]	E _U [kWh]	E _U [%]	E _T [kWh]	E _T [%]
AMPds2	NN, w/o FR	136.966	341.062	43.591	139.917	74.748	46.948	-31.157	-16.367
	NN, w/ FR	133.853	332.633	45.178	106.736	70.472	47.602	-25.294	-13.287
	CO	956.996	1306.700	805.931	507.231	20.914	66.433	785.017	412.364
	FHMM	465.163	699.053	385.334	240.339	16.567	55.149	368.767	193.711
UK-DALE	NN, w/o FR	9.911	13.077	11.729	91.101	5.640	34.214	6.089	20.722
	NN, w/ FR	9.233	12.298	11.528	89.319	4.649	28.268	6.879	23.433
	CO	23.889	34.634	21.635	236.850	20.219	100.000	1.417	4.826
	FHMM	22.039	32.162	22.021	204.723	16.591	89.216	5.430	18.499

From examples in Fig. 4.18, it is clear that the use of frequency components from the FSR-stage as additional predictors in the load disaggregation model can increase the accuracy of the estimations for both heating and lighting loads. Specifically, in the example of AMPds2, the identified more negatively correlated frequency components of demand capture peak during the morning time (Fig. 4.15(a)), and this pattern helps the CNN-BiLSTM model to assign more attention and weight to this period, thus leading to a more precise capture of the heating load (Fig. 4.18(a)). Similar comparisons can be made between Fig. 4.16(a) and Fig. 4.18(b) for the lighting load disaggregation from the UK-DALE dataset, as in this case lighting load has been more accurately estimated in the model with FSR results. In Fig. 4.19, the weekly results for the CNN-BiLSTM model with correlated frequency components are consistently better than the results without them.

The calculated error indices in Table 4.7 to Table 4.9 also demonstrate the benefits of using correlated frequency components, as these results have lower MAE and RMSE values for the example days, weeks, and in the whole test sets. In terms of the energy-

based errors, although the overestimated and underestimated energy consumptions may not be simultaneously improved, the total energy estimations are generally better. There are some exceptions, e.g., in the week and whole test set comparisons for UK-DALE, where E_T is seemingly worse if correlated frequency components are used, than if they are not used. This is because the E_O and E_U values increase or decrease by different levels, so the relative difference between them may be larger; in the weekly and whole test sets, the absolute (and percentage) E_O and E_U values are typically lower than in the cases when the correlated frequency components are not used.

The errors of the two NN models (frequency component-based CNN-BiLSTM model and naïve CNN-BiLSTM model) are all much smaller than the errors of CO and FHMM benchmarking models. The CO and FHMM seemingly correctly capture the phases of both morning and evening peaks, but the overall accuracy is not satisfactory. One reason is that both algorithms only consider the correlations between the total demand and individual appliances. Additionally, the performance of the CO model is impacted by treatment of each timestamp independently. However, both the CO and FHMM may perform better in the case of multiple appliance disaggregation, as they can analyse the combined demand of individual appliances. The CO usually contains a set of individual appliance models and aims to minimise the difference between the sum of estimations of all appliance demands and observed aggregated total demand. In FHMM, the separate appliances are also jointly considered, where each disaggregated load is modelled through one HMM and a hidden component of the HMM is the state of the individual appliance. In that way, all HMMs are combined to form one FHMM (or an equivalent HMM) in the final model to estimate the demand of all appliances.

All models are implemented using Python 3.8.5 and NumPy 1.8.5. The FSR-based decomposition and correlational analysis are built on Scikit-learn 0.23.2 and SciPy 1.5.4. The Bayesian optimisations and final CNN-BiLSTM models are implemented with KerasTuner 1.0.2 and TensorFlow 2.4.1, respectively. The CO and FHMM models are obtained from NILMTK 0.4.0. All calculations are done on a desktop running Microsoft Windows 10 21H1 operating system, with AMD Ryzen 1800X CPU, Nvidia GTX 1080ti GPU, and 32GB DDR4 RAM. The training times of all

models are summarised and compared in Table 4.10. The computational time for BO is very long, since there are 96 times of BO executions in each case. The required times for BO and CNN-BiLSTM are given within square brackets, where the first number is related to the case where the FSR results are not used as an additional predictor, while the second number indicates the case where the model was run with the FSR results. It can be seen that considering the frequency domain information can also reduce the computation time in most cases.

Table 4.10 Comparison of Training Time for All Models

	FSR [minute]	BO [minute]	CNN-BiLSTM [minute]	CO [second]	FHMM [second]
MV level dataset	59	-	-	-	-
AMPds 2	15	[1097, 934]	[15, 14]	5.4	5.1
UK-DALE	31	[1230, 750]	[6, 12]	12.1	11.8

4.4 Discussions and Conclusions

This chapter presents a novel Fourier series regression method to extract selected frequency components (half-daily, daily, 2-daily, 5-daily, weekly, monthly, seasonal, and annual) from periodically changing non-disaggregated time series of active power demands, temperature, and solar irradiance. The presented method is illustrated on the three study cases with actually measured demands, corresponding to two household-level and one network substation-level datasets. The reconstructions by selected frequency components show that the original time series can be reasonably accurately represented, i.e., that the selected frequency components preserve the information on the total demand.

Afterwards, the chapter presents the analysis of the correlations between different frequency components and demonstrates that specific individual frequency components of demands, temperature, and solar irradiance have stronger negative correlations than the original non-disaggregated time series. As the heating and lighting loads are expected to exhibit strong negative correlations with temperature and solar irradiance, the stronger negatively correlated frequency components are combined and used as the additional explanatory variables, i.e., as an additional feature of the CNN-BiLSTM load disaggregation model. The obtained results and comparisons demonstrate that the CNN-BiLSTM model in which correlated frequency

components are used as the additional explanatory variables is more accurate for load disaggregation than the conventional/naïve CNN-BiLSTM model without frequency components (“black box model”).

The presented CNN-BiLSTM model adopts Bayesian optimisation to select the most appropriate hyperparameters for the four separate load disaggregation models, showing that the heating and lighting load are identified with higher accuracy when the correlated frequency components are used as the additional information/input during the disaggregation. The presented frequency component-based CNN-BiLSTM model is additionally benchmarked against two other-widely used NILM models (CO and FHMM), which both achieve lower accuracy of the load disaggregation than the presented model.

Chapter 5

A Review of Approaches for the Detection and Treatment of Outliers in Processing Wind Turbine and Wind Farm Measurements

5.1 Introduction

Over the last few decades, there has been a significant increase in the penetration of wind-based electricity generation systems, with installed powers ranging from a few kilowatt (kW) individual wind turbines, to gigawatt (GW)-scale wind farms. As the number of wind energy conversion systems is increasing, so is the importance of having accurate information on their power outputs, which is difficult to obtain due to the strong and inherently stochastic variations of the wind energy resource, as well as the complexity of WT control systems and highly dynamic nature of WT operation. For example, the assessment of uncertainties in the outputs of wind-based generation systems [56] is of importance in a number of power system applications, including condition monitoring and fault diagnosis [137], unit commitment and economic dispatch [138], as well as the analysis of system security, reliability and voltage stability [139].

The common approach for modelling WTs and WFs is to use power curves (PCs) [16, 140-142], which typically provide information on the generated power output (P_{out}) as a function of the input wind speed (WS). These PC-based WT and WF models have numerous applications, e.g., in wind power forecasting [143, 144], WT condition monitoring [16, 145], WF condition monitoring [146, 147], wind energy resource assessment [82, 148, 149] and general networks analysis [150]. Manufacturers' power curves (Mfr-PCs) are typically used if no field (measured) data are available, as is the case in a planning phase, or general design phase of a WF. However, since Mfr-PCs are obtained under predefined standard conditions [11, 12], they cannot reflect many WT and WF application-specific factors, such as the impact of the terrain, wake effects and dynamics of input wind energy variations [13, 15, 151, 152]. Furthermore, after several years of operation, natural ageing will result in a deterioration of WT performance from the originally specified Mfr-PC, which may be further strongly

impacted by the meteorological conditions at locations where the WT is installed [15, 142, 153, 154]. Accordingly, the actual field performance of an operational WT can be more accurately assessed from the (equivalent) operational PC-based WT model, which is typically derived from the available measurement data. For that purpose, the corresponding IEC Standard [12] provides guidance for power performance measurements of an electricity-producing WT, including requirements for the test equipment, measurement procedures, derivation of results, data normalisation and reporting format. This standard is specifically intended for the analysis of individual WTs. Although it does not specify nor instruct how an equivalent PC model could, or should be obtained for the whole WF, it does specify the predictor variables impacting the wind power output. These predictor variables should be taken into consideration during the measurement and modelling, such as wind direction (WD), air density (mainly influenced by temperature, air pressure and relative humidity), turbulence intensity and some other conditions (e.g., precipitation, icing, dirt and insect accretions, etc.) [12, 151]. The PC-based model obtained from synchronous P_{out} and P_{in} measurements of an individual WT is in this thesis denoted as a WT-level model (see e.g., [16, 155]), while a PC-based model of the whole WF is denoted as a WF-level model (see e.g., [11]).

As with all measurement-based models and data-driven modelling approaches, the quality of the data used for both WT-level and WF-level modelling has to be checked [156]. The bad or abnormal data must be identified and removed (or suitably processed), as otherwise the obtained WT and WF models may be biased, or even inaccurate. There are wide ranges of both causes and manifestations of these bad/abnormal data, which are often denoted with a general term “outlier”. In the context of the PC-based WT and WF models, outliers are often identified as the specific measurement data that are far away from a much larger and closely clustered set of normal or regular measurement data, against which outliers are evaluated.

Correct identification and processing of outliers in the WT and WF measurements are particularly important in the condition monitoring and fault diagnosis applications [145, 152, 154, 157-162]. For example, the presence of outliers is one of the biggest difficulties in applying appropriate PC-based monitoring methods, as the outliers

usually result in unsatisfactory power curve limits and it is important that input data for the limit-setting algorithms do not include outliers [158]. Undetected or incorrectly processed outliers also have strong negative impact in applications related to wind power forecasting, power flow studies and economic dispatch analysis, since the models with outliers are likely to have lower accuracy, higher forecasting errors and suboptimal unit commitment results [163-166].

It can be generally concluded that accurate measurement-based models require to separate outliers (i.e., bad quality or abnormal data) from the regular data (i.e., high quality or normal data), in order to correctly analyse and represent the underlying relations among the all considered model parameters, variables and dimensions. The importance of outlier analysis is reflected in a recent increase of interest of both scientific and engineering communities in approaches and methods for detection and processing of outliers in measurement-based WT and WF models. Most of the proposed approaches are statistical [11, 15, 79, 82, 146, 147, 151-153, 156, 158, 160-162, 164, 166-175]. More specifically, [146, 162] have classification-based methods, [11, 147, 160, 166, 167, 170-173] present clustering-based models, [168, 169, 174] use correlation-based approaches, [15, 79, 82, 152, 156, 170] apply normal-based, quartile-based or mean absolute deviation-based criteria (where [156] also applies change point detection, and [170] also employs hypothesis test methods), [153, 158, 161] propose recursive or iteration-based algorithms, [151] distinguishes the outliers by statistically robust data fitting, [164] utilises data smoothing and averaging followed by data censoring rule, and [175] leverages computer visual technique. Some other methods are based on physical characteristics and constraints [16, 154, 176], or on combining both statistical approaches and physical constraints [177]. Finally, some of the work tried to develop so-called robust statistical models that could reduce the impact of outliers without their detection and elimination [157, 159, 178]. In terms of outlier treatment, in most of the work outliers are simply removed, e.g., [11, 15, 16, 79, 82, 146, 147, 151, 153, 154, 156, 158, 161, 162, 164, 167, 168, 171, 176, 177], while few references treated outliers as missing values and tried to recover them, e.g., [169, 174, 179]. However, and despite the significant work on the detection, treatment and analysis of outliers in WT and WF measurement applications, there is still lack of the commonly accepted, or recommended methodologies for their comprehensive and

systematic analysis, including clear definitions, detailed specification of their characteristics and confident identification of their causes and origins. This chapter provides a review of the approaches and methods reported in available literature for the detection and treatment of outliers when processing WT and WF measurement data, as well as further discussion of some general and specific aspects of the related analysis of outliers.

5.2 Wind Turbine and Wind Farm Measurements

Most of the previous related work focused on individual wind turbines (WT-level analysis), taking into consideration only the measurements from the sensors typically located at the WT nacelle. Some work, however, aimed to model the whole wind farm (WF-level analysis), when either: (a) data representing the whole WF are obtained directly, e.g., as a single-sensor measured total WF power outputs (P_{out}) data, typically obtained at the WF grid connection point, as well as a single-sensor measured WS and WD data, typically from the meteorological masts and towers [164, 167, 172, 174], or (b) available measurements from individual WTs in a WF are in some way aggregated to derive the information on the whole WF indirectly [146, 147]. The found literature are grouped by the levels of their presented analysis (e.g., WT-level and WF level), which are listed below:

- Analysis of individual WT performance, using single WT measurements: [15, 16, 79, 82, 145, 151-156, 158, 161, 162, 166, 168, 169, 171, 173, 175, 177, 179];
- Analysis of whole WF performance, using single-sensor WF measurements: [164, 167, 172, 174];
- Analysis of whole WF performance, using aggregation of measurements of all WTs in the same WF, i.e., not using directly measured single-sensor WF data [146, 147]. For example, [146] applies principal component analysis (PCA, [180]) on the measurements from 89 individual WTs in a WF, which are processed to reduce dimensionality and obtain an equivalent WS for the whole WF, while equivalent P_{out} WF values are calculated from the P_{out} measurements of 89 WTs. In [147], average values of P_{out} , WS, rotor speed and blade pitch angle are collected from 22 individual WTs and then used to obtain three PC

models for the whole WF: (i) power curve model (P_{out} -WS relationship), (ii) rotor curve model (rotor speed-WS relationship), and (iii) blade pitch curve model (blade pitch angle-WS relationship);

- Analysis of both WT and WF performance: [11, 176]. For example, a dataset consisting of 20 WTs in a WF is used in [11] to propose an outlier elimination approach for WT measurements, which is then applied to the WF's P_{out} -WS scatter-plot data, without specifying the method for obtaining the WF measurements (an example illustrating WF's P_{out} -WS scatter-plots can be found in Section 5.3.2). The measurements of both individual 274 WTs (P_{out} , WS and turbine availability binary data, which are sampled every 10 min) and of the whole WF (WS, WD, P_{out} , etc.) are used in [176], with a focus on the whole WF performance. Individual WT availability data are used to help identification of outliers in the WF measurements, underlining the difference between the aggregated WF P_{out} (from the individual WTs' P_{out} data) and single-point measured whole WF P_{out} , which is attributed to losses.

In the WT and WF outlier detection and treatment research in available literature, WS and P_{out} are the two most often considered variables, as the impact of other meteorological variables is less significant. In cases where measurements of other variables were available, these extra data were usually either removed, or used only for some limited analysis, often without explicitly checking the presence of outliers in the extra data. If other/extra variables are measured and considered in the outlier-related work, this was usually done at the data preprocessing stage, with examples that were found in a relatively smaller amount of literature discussed below:

- Measurements of rotor speed are added as an additional variable in order to help detection of outliers in WS and P_{out} measurements in [169], while WD is considered as an auxiliary variable in the model proposed for filtering outliers in P_{out} -WS data in [151];
- Measurements of additional variables are used for the transformation of P_{out} or WS data, e.g., based on guidance in [12] for data normalisation: in [15], temperature, humidity and air pressure are used to obtain the air density and then to transform P_{out} data; similar approach is applied in [176] to transform

WS data. In [151] and [161], the same WS or P_{out} corrections as in [15] and [176] are used to eliminate the impact of the variance of WS fluctuations by expanding the PC into a Taylor series with the higher order terms neglected. Another example is use of turbulence intensity, as defined in [12], which has significant impact on ten-minute averaged WS and P_{out} measurements (see Section 5.3.4). Typically, the increased turbulence intensity leads to an increase of power output at the lower wind speeds (above the “ankle” of the power curve), while at the higher wind speeds (below “knee” of the power curve), increased turbulence intensity results in a decrease of power output. Accordingly, a procedure for normalising PC data to a reference turbulence intensity from [12] is used in [181]. In that paper, a simulation-based approach and normal distribution (location parameter is the average wind speed within 10 minutes, scale parameter is the standard deviation of the wind speed within 10 minutes) of WS values within a 10-min window (and some other assumptions, for example, that at each instant the WT follows a zero turbulence PC) are used instead of the Taylor-series approach [151, 161] (both approaches are presented in [181, 182]);

- Examples of use of additional data where efforts are also made to detect potential outliers in these measurements are: [16], where the importance of ambient temperature and wind direction is emphasised, as temperature has the biggest influence on air density and, in turn, P_{out} , and [147, 166, 173] where various multivariate outlier detection approaches were applied on all considered variables (P_{out} , WS, rotor speed, blade pitch angle, etc.), as well as [154, 176], which tried to detect outliers by analysing them separately in one dimension and simultaneously in multi-dimensional space. Finally, [179] does not explicitly present the outlier detection method, but it employs several predictor variables (P_{out} , WS, WD, rotor speed, gear temperature, blade pitch angle) and treats the outliers in these variables as missing values.

In the above references, most of the measurements are sampled and collected at higher frequencies (usually 1 Hz), but are stored as 10-min average values [12], while only [82, 145], and [152, 153, 158] used 5-min average values and [167, 174] used 15-min average values.

All above discussed references use P_{out} and WS measurements and propose detection methods for potential outliers. However, five additional references are found with related, but inherently different analysis:

- The outliers are analysed in the measured high frequency vibration data (sampling frequency equal to 25.6 kHz) for health condition monitoring in [160], where they are collected via an accelerometer installed on the drivetrain structure of a WT;
- The outliers are considered for cyber-security analysis of false data injection attacks for two WFs (each with 18 WTs) using real-world hourly WS information in [170], but P_{out} values of WTs are calculated with a deterministic piecewise linear power curve;
- Robust statistical techniques that may reduce the effects of outliers and could be used as nominal baselines are proposed in [157, 159, 178], but the specific outlier detection methods are neither stated, nor used. Ref. [157] uses 12 kHz vibration data of WT bearings; Ref. [159] analyses 10-min average WS, P_{out} and blade bending moment data of WT; Ref. [178] builds models using 5-min average WF P_{out} data. These models are reviewed in more detail in Section 5.3.4.

The above discussion is summarised in Table 5.1 below.

Table 5.1 Summary of Wind Turbine and Wind Farm Measurements in Available Outlier-Related Literature

Ref.	Outlier-Related Research Focus	Measurement Data Considered in Outlier Analysis	Measurements Resolution
[156]	WT-level	WS, P_{out}	10-min
[168]	WT-level	WS, P_{out}	10-min
[169]	WT-level	WS, P_{out} , rotor speed	10-min
[16]	WT-level	WS, P_{out} , WD, ambient temperature	10-min
[162]	WT-level	WS, P_{out}	10-min
[79]	WT-level	WS, P_{out}	10-min
[179, 183]	WT-level	WS, P_{out} , WD, rotor speed, gear temperature, blade pitch angle	Unspecified
[15]	WT-level	WS, P_{out} , ambient temperature, relative humidity, air pressure	10-min
[155]	WT-level	WS, P_{out}	10-min
[145]	WT-level	WS, P_{out}	5-min
[153]	WT-level	WS, P_{out}	5-min
[175]	WT-level	WS, P_{out}	10-min
[171]	WT-level	WS, P_{out}	10-min
[152]	WT-level	WS, P_{out}	5-min
[151]	WT-level	WS, P_{out} , WD, ambient temperature, relative humidity, air pressure	10-min
[161]	WT-level	WS, P_{out} , ambient temperature, relative humidity, air pressure	10-min
[82]	WT-level	WS, P_{out}	5-min
[158]	WT-level	WS, P_{out}	5-min
[177]	WT-level	WS, P_{out}	10-min
[173]	WT-level	WS, P_{out} , ambient temperature, rotor speed, generator speed, blade pitch angle, etc.	Unspecified
[166]	WT-level	WS, P_{out} , blade pitch angle	10-min
[154]	WT-level	WS, P_{out} , generator speed, nacelle temperature, etc.	10-min
[167]	WF-level	WS, P_{out}	15-min
[174]	WF-level	WS, P_{out}	15-min
[164]	WF-level	WS, P_{out}	Unspecified
[172]	WF-level	WS, P_{out}	Unspecified
[146]	WF-level	WS, P_{out}	10-min
[147]	WF-level	WS, P_{out} , rotor speed, blade pitch angle	10-min
[11]	WT and WF-level	WS, P_{out}	Unspecified
[176]	WT and WF-level	WS, P_{out} , ambient temperature, relative humidity, air pressure	10-min

5.3 Definitions, Characteristics and Causes of Outliers

In related literature, different terms and terminology are used to refer to, or to denote an outlier. The following three general cases are found: 1) the authors indicate the alternative term by using an abbreviation for interchangeability “i.e.,” between it and the term outlier; 2) the authors explicitly state that the specifically termed set of data

are considered as the outliers; 3) the authors effectively treat the considered set of data as outliers by, for example, detecting them and either correcting or removing them from the further analysis. The alternative terms commonly found in the literature to denote an outlier are: “abnormal data” [79, 156, 171, 172, 175], “(data) error” [146, 172], “incorrect data” [167], “unnatural data” [167], “irrational data” [167], and “wrong measurement” [15]. Furthermore, the same term may be used in distinctively different contexts in different references. For example, the term “missing data” is distinguished from the term outlier in [167, 179], while the term “missing data” is used to refer to the data obtained when the communication system is faulty and when the last reading before the fault has occurred was recorded over a longer period, leading to the series of constant or zero values in the recorded data in [173].

5.3.1 Definitions of Outliers

Most of the references define outliers in a statistical sense, as the data that are simply different or stand apart from the other data points, and deviate from the normal or logistic trends. In terms of PC-based models, outliers are points assumed to be outside the range of normal output characteristics of the power curve. In related literature approaching outliers from the statistical point of view, outliers are data points which usually appear to be under-power points, stopping points, or sparsely located points (e.g., either low wind speeds, but high power outputs, or high wind speeds, but low power outputs), which are usually identified in the P_{out} -WS scatter plot [79, 156, 170, 172, 174]. Instead of relying on P_{out} -WS scatter plot, [178] defines outliers by looking at time series, considering the recurrent and large changes in wind power time series as outliers, although it also states that these recordings can be frequently found.

There are a few references that define outliers in different ways, e.g., looking at operational WT performance. For example, [161] defines outliers as the values that are not corresponding to normal operating conditions, or when the WT is not available over the whole 10-min averaging period. Reference [161] does not provide an explicit definition of normal operating conditions, but according to [151] asserts that the WT will work properly during the 10-min period if the power data points are averaged considering the power production; otherwise, the recorded data do not represent the normal operation of the WT. Although not giving clear definition of outliers, [154,

159] state that many outliers are particularly likely to be generated during the (extreme) transient operations, such as the start-ups and shut-downs of the WT. However, it is not uncommon for a WT to have several start-up and shut-down operations within a 10-min period, especially when the wind speed and/or wind turbulence is high, and such approaches will remove a number of rather usual operational points from the WT model.

The most general and intuitive approach in analysing outliers is that an outlier is denoted as the unusual or unexpected value, which is quite different from the set of values obtained or sampled under the same or very similar conditions (sometimes called inliers). If an outlier is related to a measured or in some other way observed value, its presence is typically assessed against a larger and clustered set of observations or measurement points obtained for the same or similar conditions. These points are used to both establish acceptable range(s) of variations (e.g., due to a limited measurement accuracy) and, through it, a statistically significant reference set(s) of values for the evaluation of outliers. Basically, this represents identification of outliers from a statistical point of view, where outliers are values that are at a large distance (or isolated) from the corresponding reference set of values (as denoted in e.g., [152, 177]). In a different, but still associated context, results of measurements and observations are often parts of experiments aimed at deriving unknown, or at confirming assumed physical (or at least causal) principles behind the observed process, when outliers are basically related to the observation points or measurement values that cannot be explained with the known or assumed physical principles. This represents identification of outliers from a physical point of view, where outliers are values that are physically implausible, or in some other ways violate underlying physical principles assumed to be valid under the considered conditions (see also [152, 177]).

It should be noted that analysing outliers from a statistical point of view is closely (but not exclusively) related to a concept of a visible outlier. Indeed, if statistical approaches are used for the identification of outliers, all identified outliers are by definition visible outliers, as they should be clearly distinguishable from the reference set of values to which they do not belong to. From the physical point of view, i.e., if

underlying physical principles are used for the identification of outliers, the outliers may or may not be visible. Specifically, the outliers may be hidden in the following cases: 1) underlying physical principles are not fully and correctly describing the observed process; 2) the observed process is not properly measured; 3) the process is stochastic and features strong uncertainties; 4) the process is highly dynamic and an outlier is related to a recorded transient value; 5) conditions simply cannot be properly reproduced or repeated in terms of both known and unknown variables. Typically, a hidden outlier is found within the otherwise regular reference set of data and an example are (random) de-synchronised WS- P_{out} measurements, when some outliers may be recorded outside the reference cluster (i.e., they will be visible), while other may be inside the reference cluster (i.e., they will be hidden). In such cases, none of them should be used for the analysis, although hidden outliers will seemingly be in accordance with the underlying physical principles, making their identification much more difficult.

Accordingly, this thesis uses a general but simple terminology, in which an outlier can have only two attributes: (a) true outlier, which means that this outlier should be removed (or suitably processed) from the further analysis, e.g., the abnormal data due to the errors of data acquisition, exchange, processing and management procedures, and (b) false outlier, meaning that this outlier should, or could be considered for the further analysis, assuming that there is a clear explanation or reasoning behind that decision, or that conditions under which this outlier is obtained are included in the model. An example are “mid-curve stacked outliers”, which might be due to wind curtailment control of a WT [156, 166]. In that case, the decision whether they will be processed as the true or false outliers basically depends on the inclusion, or not, of curtailment mode of operation in the WT model [11, 175]. More examples of true and false outliers are discussed in the next sections.

This approach can be applied to outliers processed from both statistical and physical points of view, as well as for both visible and hidden outliers. Although not necessarily using the same terminology, a similar discussion about true outliers and false outliers can be found in a few references. For example, although many of the observed outliers are due to problems with the meteorological tower anemometers, there are some

outliers that may not be errors, as individual WT sensors in general have measurements different from the meteorological tower measurements, which means they are generally false outliers [176]. “Isolated outliers” are one category of outliers discussed in [173], which may occur due to e.g., extreme weather conditions or due to human interference, and they may be either false (if extreme weather is to be included in the model), or true (if they are due to a human error). A clear distinction between the true outliers and false outliers is crucial in all wind energy applications, but especially in condition monitoring, where typically measurements from various sensors different from the measurements discussed in Section 5.2 are used. According to [145, 152, 154, 157-162], false outliers may exist in the recorded data and may be very important for condition monitoring, but further analysis of these references is deemed out of scope of this thesis, which focuses on the approaches for dealing with such outliers, not on the subsequent conditioning monitoring analysis.

5.3.2 Characteristics and Causes of Outliers in Available WT and WF Measurement Data

This section analyses the characteristics and causes of outliers reported in available WT and WF measurement data, where the current literature are also reviewed. Three representative P_{out} -WS scatter plots are depicted in figures below. Fig. 5.1 illustrates WT-level outliers using actual 3-year 10-min synchronous nacelle WS and P_{out} measurements of a single (on-shore) WT, which is pitch regulated, with cut-in, rated and cut-out wind speeds equal to 4 m/s, 18 m/s and 25 m/s, respectively [87]. Fig. 5.2(a) is based on the 3-year 10-min synchronous nacelle WS and P_{out} measurements of the same type of (but off-shore) WT [87]. Fig. 5.2(b) plots the P_{out} -WS relationship of a whole WF, which consists of the 36 WTs of the same type as in Fig. 5.2(a). In that case, directly measured WF values of single-sensor obtained WS and single-sensor obtained P_{out} data were not available, so whole-WF values are calculated as the averages of synchronous WS and P_{out} measurements from all WTs in the WF [142]. Finally, Fig. 5.3 shows available WF-level data obtained from direct single-sensor synchronous 10-min measurements of WS and P_{out} values over the period of one calendar year.

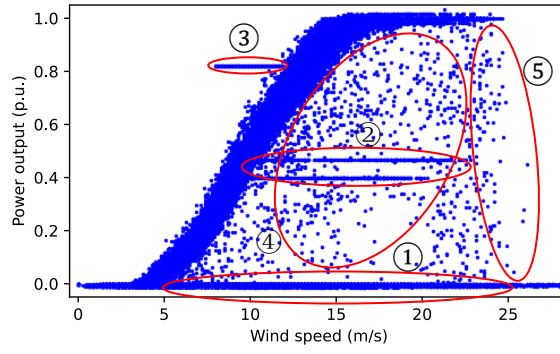


Fig. 5.1. Actual 3-year measurements with indicated different types of outliers for an individual WT.

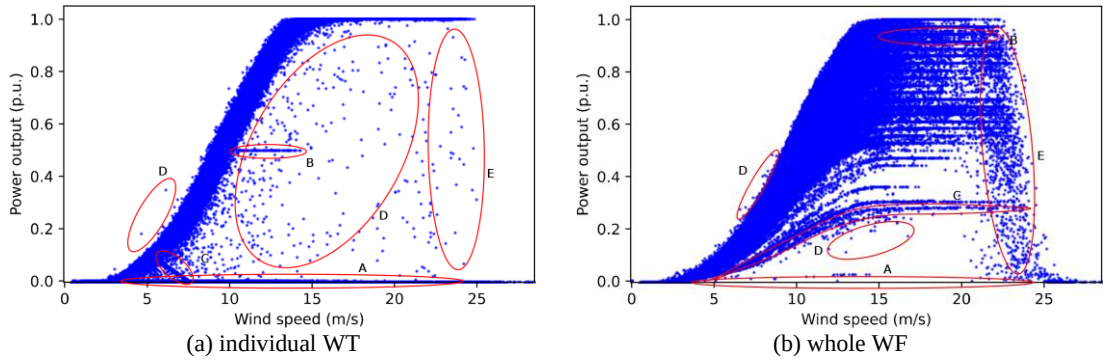


Fig. 5.2. Actual 3-year measurements with indicated different types of outliers for: (a) an individual WT, and (b) a whole WF (derived from the individual WTs in the same WF). (The A-E marks indicate characteristics of various types of WT outliers.)

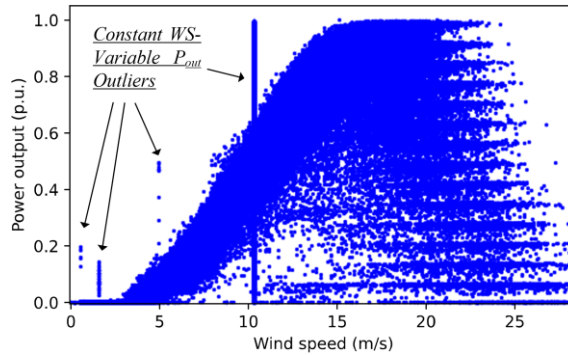


Fig. 5.3. Actual 1-year measurements with indicated constant WS-variable P_{out} outliers for a directly measured whole WF.

According to a previously introduced distinction between the WT-level and WF-level analysis of outliers, discussion of causes and characteristics of outliers is divided in outliers that occur at both WT-level and WF-level, and outliers that will be highly unlikely to occur in one of these. It should be noted that if the processed WF measurements are obtained directly (single-sensor measurements of WS and single-sensor measurements of P_{out}), the analysis of these WF-level outliers is similar (but

not identical) to the analysis and processing of outliers of individual WT measurements, i.e., WT-level outliers. In other words, some outliers at WT-level are highly unlikely to be found at WF-level. The greater differences from the WT-level outliers exist when WF-level data are obtained from the aggregation or averaging of measurements from individual WTs.

The characteristics and possible causes of various types of WT outliers are summarised as:

- Low P_{out} —High WS Outliers, e.g., these marked as ① and A in Fig. 5.1 and Fig. 5.2(b), respectively: At WT-level, these outliers are characterised with zero, or very low, or even negative P_{out} values when the corresponding/synchronous WS values are between the cut-in and cut-out wind speeds, i.e., when WT is expected to generate non-negligible output power. These outliers are denoted as “bottom-curve stacked outliers” and “stacked data at the bottom of the curve”, typically characterised by horizontal dense data band in the PC-based model [156, 166]. Their causes include WT failure or outage, unplanned WT maintenance, and faults of the WT measurement and/or communication systems [175]. Similarly, [15] stipulates that outliers representing P_{out} data close to, lower than, or equal to zero are exclusively due to wrong measurements or WT malfunctions (see Fig. 5.2(a) and outliers marked with A). If the WT does not rotate, P_{out} should be zero, but if the WT’s control system is kept energised, the P_{out} measurements might be negative [153]. At WF-level, when data are obtained directly, these outliers are most likely true outliers due to measurement errors or communication system errors [174], or disconnection of the whole WF due to activation of the protection system (see Fig. 5.1, outliers marked with ①). When WF-level data are obtained from the aggregation of individual WT measurements, these outliers are highly unlikely to be present (see Fig. 5.2(b), outliers marked with A). Regardless of WT-level or WF-level analysis, some of these outliers may be hidden, e.g., for low WS values;
- High P_{out} —Low WS Outliers, e.g., these marked as ③ in Fig. 5.1: At WT-level, these outliers are typically manifested as the horizontally stacked data

with a narrow range of variations of relatively high P_{out} values (e.g., close to 1.0 p.u.), which are clearly visible when they are above the upper boundary of usual/expected/regular reference WT PC values, i.e., for given WS values, these outliers represent higher than expected and relatively constant P_{out} values. These outliers are denoted as “top-curve stacked outliers” and “stacked data at the top of the curve”, which appear to be one or more horizontal dense data bands, located above and to the left of the regular PC-based WT model (see Fig. 5.1, outliers marked with ③) [156, 166]. These outliers are usually caused by communication errors, or wind speed sensor failures ([156] states that faults of WS sensors happen frequently), e.g., when lower than actual WS values are recorded. When WF-level data are obtained directly, these outliers are similar to WT-level outliers [167, 174], but may have different P_{out} values, based on the number of operational WTs in the WF. When WF-level data are obtained from the aggregation of individual WT measurements, these outliers may or may not be visible, based on the operating point of the WT(s) with faulted measurement/communication system, as well as the (aggregated) operational point of the whole WF;

- Low P_{out_max} Outliers, e.g., these marked as ② in Fig. 5.1 and these marked as B in Fig. 5.2(a) and Fig. 5.2(b): These outliers are again represented by horizontal bands/ranges of relatively constant P_{out} values, which are below the lower boundary of the expected range of PC values (see outliers marked as B in Fig. 5.2(a) and Fig. 5.2(b)). These outliers are denoted as “mid-curve stacked outliers” and “stacked data in the middle of the curve” [156, 166]. Mentioned causes include curtailment, but also down-rating of WTs and data acquisition and communication system errors [11, 175]. In some cases, improper WT operation, for example due to a damage of WT gearbox bearing [154], may also limit the maximum WT P_{out} (e.g., restricted to 60% of the rated maximum power). When WF-level data are obtained directly, these outliers are similar to WT-level outliers [174] (see outliers marked by ② in Fig. 5.1), but may have different P_{out_max} values, based on the number of operational or curtailed WTs in the WF ([176]) and may also include communication and measurement errors of single P_{out} and WS sensors [174]. For aggregated WF-level data,

measurement and communication errors of individual WTs will most likely make these outliers to be hidden, but there may be many horizontal bands, based on the actual number of WTs that are curtailed or down-rated. An important feature of this type of outliers, distinguishing them from the next discussed type, is that there are no big differences in WF power outputs up to the point of curtailment. In terms of true/false outlier analysis, if curtailment (or downrated/derated) operation should be included in the WT/WF model, these outliers are false outliers;

- Shifted PC Outliers, e.g., these marked as C in Fig. 5.2(b): At WF-level, these outliers are not consequence of curtailment or down-rating of individual WTs, but are due to the outages or faults of individual WTs in a WF, reflected in shifting of the PC to the right, i.e., corresponding to the situations in which both maximum power output is reduced and higher WS' are required to produce the P_{out} values close to these when all WTs are in operation (see outliers marked as C in Fig. 5.2(b)). A WT outage is defined as the tripping of n (one or more) WTs in [174], for e.g., protection reasons or unplanned maintenance. Similar outliers may be recorded for an individual WT, when PC is shifted to the right, indicating inaccurate speed sensor readings (wind speed under-reading) or icing on blades, as illustrated in [158];
- Linear P_{out} —WS Outliers, e.g., these marked as C in Fig. 5.2(a): These WT-level outliers appear as linearly related P_{out} -WS recordings, possibly occurring during the data processing phase, when linear interpolation is applied to populate missing recordings (see outliers marked as C in Fig. 5.2(a)). Similar outliers are reported due to malfunction of pitch-control system and dirt deposits on the blades [158]. These outliers may occur at WF-level when data is obtained directly (data recording errors), but they are highly unlikely to be visible when WF-level data are obtained from the aggregation of a larger number of individual WT measurements;
- Scattered Outliers, e.g., these marked as ④ in Fig. 5.1: These outliers are related to irregular or random values around the expected/reference PC range (Fig. 5.1, marked as ④), which may be due to faults and errors, but also due to statistical processing and averaging window. In addition, a strong wind gust

may cause the WT to shut down for a significant amount of time within the averaging window, causing the averaging power output smaller than the value given by the manufacturer. These outliers are called “around-curve outliers” in [156] (in [166], “scattered data around the curve”) and can be caused by random factors, like signal propagation noise, sensor failure and extreme weather conditions. Alternative term is “sparse outliers” (due to random noise) [11]. Also, [164] shows these outliers together with other types of outliers in WF measurements. The reasons for these abnormal data are sensor failure, sensor noise and some uncontrolled random factors [175]. At WT-level and when WF-level data are obtained directly, these outliers have similar causes. When WF-level data are obtained from the aggregation of individual WT measurements, their dominant cause are averaging window and similar statistical processing-based origins (see Section 5.3.4). These outliers are termed as “unnatural data” in [167], where they are analysed separately from other outliers (these sparsely located data can also be seen in Fig. 5.3);

- Constant WS-Variable P_{out} Outliers, e.g., these in Fig. 5.3 marked as E in Fig. 5.2(b), respectively: These outliers are reported at WF level and are manifested as vertical bands of constant WS values (or only with small WS variations) from the meteorological towers recorded over a longer period with synchronously recorded relatively large P_{out} variations, occurring due to errors in data acquisition and data transmission (corrupted/polluted data) [167, 176] (see Fig. 5.3). These constant WS outliers are more strongly pronounced at WT-level and when WF-level data are obtained directly, but may be less pronounced when WF-level data are obtained from the aggregation of individual WT measurements. A variant of the constant outliers is denoted as “slender” in [167], related to approximately vertical band around the cut-out wind speed, explained by the wake effects, i.e., that the WTs within a WF do not cut out together near the cut-out wind speed, because the wind speeds at each WT vary from the measured WS at the meteorological mast [167]. It is emphasised that these data should be categorised as valid data, as they reflect the natural P_{out} fluctuation property of the WF around the cut-out WS, which is important to the system operators. However, it is not clear why these outliers

are not scattered within an oblique band with a negative slope (as shown in Fig. 5.2(b)) for outliers marked with E), but are clustered in a relatively narrow vertical band near the cut-out wind speed (wake effect outliers as discussed in [167]). When WF-level single-sensor measured WS is around the cut-out WS, most likely situation is that some of the WT's in the WF will be operating (their WS is below the cut-out wind speed), while some WT's will be stopped (as their WS is above the cut-out wind speed) and therefore the 10-min average P_{out} values for the whole WF will be between rated power and zero when the single-sensor measured WS's are around cut-out WS. According to [167], it is necessary to distinguish “(invalid) unnatural” outliers with these “valid data”, but this can be a difficult task, since both types of data may be caused by WT cut-out effects. The methods and results of [167] are further discussed and analysed in Sections 5.3.4 and 5.5.

The characteristics and causes of outliers in WT and WF measurements discussed above are summarised in Table 5.2.

Table 5.2 Summary of Characteristics and Causes of Outliers in WT and WF Measurements

Outlier Characteristics	Possible Causes	Ref. (WT-Level)	Ref. (WF-Level)
Low P_{out} -high WS	WT failure or outage, unplanned WT maintenance, or faults of measurement/communication system	[15, 153, 156, 166, 175]	[174]
High P_{out} -low WS	Measurement/communication errors, or wind speed sensor failures	[156, 166]	[167, 174]
Low P_{out_max}	Curtailment, damage of WT gearbox bearing, or measurement/communication system errors	[11, 154, 156, 166, 175]	[174, 176]
Shifted PC	Outages (caused by protection or unplanned maintenance) or faults of individual WT's in a WF	-	[174]
Linear P_{out} -WS	Linear interpolation, malfunction of pitch-control system, or dirt deposits on the WT blades	[158]	-
Scattered	Signal propagation noise, sensor failure, extreme weather conditions, wind gust, averaging window, or uncontrolled random factors	[11, 156, 166, 175]	[164, 167]
Constant WS-variable P_{out}	Errors in data acquisition and data transmission, wake effects, or P_{out} fluctuation around cut-out	-	[167, 176]

5.3.3 Data Rejection Requirements in IEC Standard 61400-12-1 [12]

The IEC Standard [12] does not provide detailed or specific discussions of outliers, but it sets rules for data rejection, i.e., it specifies which data shall be excluded from the analysis, in order to ensure that only data obtained during normal operating conditions of the WT are used in the analysis (and to ensure that used data are not corrupted). Accordingly, [12] states that during the measurement period the WT shall be in normal operation, as prescribed in the WT operations manual. Also, the machine/WT configuration shall not be changed.

According to [12], the data shall be removed under the following circumstances (other applied rejection criteria shall be clearly reported):

- When external conditions, other than wind speed, are out of the specified WT operating range;
- If WT cannot operate because of fault condition;
- When WT is manually shut down, or it is in a test, or in some of maintenance operating modes;
- If there is a failure or degradation (e.g., due to icing) of measurement equipment;
- When WD is outside the measurement sector(s), which generally exclude WDs with significant obstacles and other wind turbines, as seen from both the WT and measurement equipment;
- When WDs are outside the valid (complete) site calibration sectors;
- For any special atmospheric condition filtered during the site calibration, which shall also be filtered during the power curve test.

In particular, [12] emphasises that a large hysteresis loop in the WT cut-out control algorithm may have considerable effects and therefore shall not be included in the obtained WT PC model. It further instructs that all datasets where the WT has stopped generating power due to the cut-out at high wind speed shall be excluded. This is difficult to understand and to acknowledge, as the WTs are typically installed at locations with good wind energy resources, i.e., at locations where it is likely that WS

will reach and go above the cut-out WS, so these operational conditions and their considerable effects should not be excluded from the analysis. However, [12] suggests to collect and present measurements with cut-out behaviour in a special database and further comments that neglecting cut-out effects may lead to overestimations of energy production, especially for cases with higher average WS'. On the other hand, [12] treats cut-in effects differently, saying that power curve model shall capture the effect of hysteresis at the cut-in control algorithm, as well as the effect of parasitic losses below cut-in WS, which is possibly due to a much larger number of measurements around cut-in WS, than around cut-out WS.

Importantly, [12] suggests that all subsets of data collected under special operational conditions (e.g., high blade roughness due to dust, salt, insects and ice deposits, or if grid conditions vary significantly), or atmospheric conditions (e.g., precipitation, wind shear) that occur during the measurement period may be selected as special databases, which basically means that these values are treated as false outliers.

Finally, it should be noted that the rejection of data will result in missing values and these enforced missing data should be distinguished from the missing measurements or recordings, e.g., due to the failures of the data acquisition/measurement systems.

5.3.4 Summary of Common Causes of Outliers in WT and WF Measurements in the Existing Literature

Reported causes and origins of outliers in processing WT and WF measurements are many and various. Unlike Section 5.3.2 which analyses the characteristics and causes of various outliers in the available datasets, this section provides a more general overview based on available literature, where some references provide a more detailed discussion of the origins in terms of the characteristics, manifestations and types, while some references also attempt to classify outliers based on their characteristics and to analyse their causes (these are discussed and further reviewed in Section 5.3.2).

In existing literature on processing WT measurements (WT-level analysis), the outliers may be consequences of:

- WT downtime, offline operation, or outage/fault: [156]

- WT curtailment, down-rating, or derating: [11, 156, 158, 166, 175]
- (Manually set) WT constrained operation: [16, 162]
- Data acquisition failure: [15, 156, 159, 160, 168, 171]
- Inaccurate measurements caused by improper measuring location (usually for WS), or WS under-reading: [16, 158]
- Data transmission (communication system) failure: [11, 151, 156, 160, 168, 171, 173]
- Data processing failure: [151, 171]
- Data management failure: [11, 151, 160]
- Electromagnetic interference: [151, 156]
- WT control system problem, including pitch control malfunction, blade pitch angle error, control program problem, yaw and pitch misalignment, incorrect or unsuitable controller setting and similar: [16, 82, 151, 158, 162, 166, 171, 175]
- WT blade damage, or excessive deposition of dirt, ice or insects: [11, 82, 156, 158, 162, 171]
- Wind shear: [176]
- Shading or wake effects from neighbouring WTs, terrain or physical obstacles: [16]
- Extreme weather conditions, harsh environment: [156, 160, 171, 173, 175]
- Air density fluctuations: [16]
- Issues related to averaging period: [16, 168]
- WT cut-out effects: [168]
- WT data availability issues: [176]
- Other, including WT malfunctions, alarms in WT, low level of gearbox oil, worn-out generator brushes, sensor accuracy, and various errors in sensors: [11, 15, 82, 151, 159, 162, 171, 173, 175]

Regarding the processing of WF measurements (both directly and indirectly), the outliers may be caused by a number of similar sources as in the WT-level analysis. References listed below are explicitly discussing following sources of WF-level outliers (although they refer to problems, issues and conditions of individual WTs):

- WT downtime, offline operation, or outage: [174]
- WT curtailment, down-rating, or derating: [11, 172, 174]
- Data acquisition failure: [146, 167, 172]
- Data transmission (communication system) failure: [11, 167, 172, 174]
- Data management failure: [11]
- WT blade damage, or excessive deposition of dirt, icing or insects: [11]
- Wind shear: [176]
- Extreme weather conditions, harsh environment: [167]
- Wind speed (around the cut-in or cut-out wind speeds): [146]
- WT data availability issues: [176]
- Other, including WT malfunctions, fluctuations in WT performance, multiple non-meteorological factors and various errors in sensors: [11, 146, 147, 167, 172, 174]

Outliers Related to Data Acquisition, Data Transfer and Data Management Errors

The long-term measurements of WT and WF characteristics practically always result in large volumes of data [146]. Handling of these large volumes of data through the related procedures for data capturing, collecting, exchanging, processing, storing and using/managing is a particularly demanding task, which may be a significant source of outliers. Accordingly, incorrect or corrupted data, or data errors, may occur in any of the four following general phases: 1) data acquisition; 2) data pre-processing and data processing (e.g., data compression, averaging, imputation of the missing data, etc.); 3) data transmission (data exporting/importing via communication systems); and 4) data management (e.g., selection and use of various data formats, data storage, data maintenance, etc.) [11, 151, 160, 167, 171, 172].

If the outliers originate solely due to the errors and mishandling during the overall data acquisition, data exchange, data processing and data management procedures, they are defined as wrong measurements, data errors, incorrect data, corrupted data, or polluted/contaminated/dirty data. As they do not correspond to the actual measurement points, these outliers are true outliers, even when they are manifested as hidden outliers, i.e., when they occur within the range or boundary of the pre-specified

regular/reference set(s) of data, or when they seemingly follow related underlying physical principles. As mentioned, their detection in such cases may be difficult, if not impossible, unless the information (e.g., from a separate log of events) is available for the periods of time when there were sensor errors and other problems with measurement instruments, or when measured data were not properly processed, or when data were corrupted during the data transmission and data management procedures. Generally, if the outliers are not caused, or cannot be explained by the errors and mishandling in the data acquisition, transmission, processing and/or management procedures, these outliers will usually require further analysis and some of them are discussed in the further text.

Outliers Determined by the Operational Logs and Event-Logging Systems

Information on the presence of outliers can be simply obtained from the logs of operational events. In such cases, any related measurement, whether correct or incorrect, which is obtained during the period of outage, or testing, or maintenance of an individual WT is a true outlier and should be discarded from the analysis and not included in the operational model of the considered WT, as the conditions under which these measurements are obtained are not related to normal operating conditions (as indicated from the log). With the terminology proposed in this thesis, however, these outliers are true outliers only at the individual WT-level. If the considered WT is part of a larger WF, and if the target model is an operational PC WF model, all outage-related WT-level true outliers are part of the normal operational performance of the WF, what effectively makes them false outliers at the WF-level, i.e., these outliers should be included in the assessment of WF performance. For example, [176] reported that a WF with 274 WTs was operating with no more than 2 WTs out of operation (i.e., with 272, 273 or 274 WTs) for only 61% of time. Similarly, [25] reported that an on-shore WF with 6 WTs and an off-shore WF with 36 WTs were operating with all WTs only 30% and 14% of time, respectively, confirming that a common operational mode of WFs was with one or more WTs out of operation. In terms of the actual fault rates/probabilities of WT components, [184] states that these may vary from around 0.05 failure per year, to about 0.55 failure per year, with the repair times per failure varying from about 1.1 days, to 14.1 days.

Similarly, the information on the WT availability status is in [176] used as a separate log containing binary type of data, where 0 indicates unavailable WT and 1 indicates the WT is available. These input data are sampled and recorded at the start of the 10-min averaging window and [176] confirmed that these data also suffer from data quality issues: if the WT is producing power with an availability status of 0, there is a clear conflict. However, the opposite condition, when there is no power output from the WT, but the availability status is 1 (e.g., when the recorded WS is below/above the cut-in /cut-out WS'), is difficult to validate.

Outliers Due to Applied Period of Averaging (Averaging Window)

Following the recommendations in [12], WS and P_{out} data are typically sampled at 1 Hz, i.e., obtained as second-by-second measurements, but are then averaged and stored as the average 10-min data. It is interesting that [161] mentions that the purpose of using 10-min average values is to reduce the effects of outliers (as well as noise and auto-correlation). As WS typically exhibits strong short-term variations (in the range of seconds) and as (large) WTs usually have inertia time constants from a few to several seconds, this simply means that each pair of measured 10-min average WS and P_{out} values will typically consists of intra-10-min periods with both higher and lower than average recorded WS and P_{out} values. In the most general case, each 10-min period of averaging may include shorter periods of zero/low WS and P_{out} values (no or low wind, resulting in no or low production), when there will be also periods of (much) higher WS and P_{out} values (including wind speeds over cut-out WS, when WT will be stopped). For example, if the wind in the 10-min averaging window has intensive fluctuations, there may be consecutive WS recordings that are higher than the cut-out wind speed, triggering the control system to shut-down the WT, when the WT will re-start generation only if the WS decreases to a certain lower level. In some extreme cases, there may be multiple start-up and shut-down operations within a single 10-min averaging window and [168] explains that these outliers are not indications of faults or anomalous WT operation, basically sharing the same concept of false outliers in this chapter. These outliers are approached merely from a P_{out} -WS relationship in [16], where it is said that P_{out} has a cubic dependency on the WS and therefore strong WS fluctuations in the averaging period will strongly influence measured P_{out} data.

Outliers Due to Cut-Out Effects

The term cut-out effects in this section generally denotes the impact of the WT control system on the measured P_{out} values when the corresponding measured WS values are around and above cut-out wind speed. These measurements can be generally approached as outliers when the measured P_{out} values do not correspond to the values indicated by the Mfr-PC and are directly related to period of averaging and can be regarded as a special case of period averaging outliers from the previous sub-section. According to [168] and previous discussion, during a 10-min averaging window a WT may be stopped by the control system for several reasons, one of which is too high input WS, where the limit is typically denoted as a cut-out WS. If the WS within the averaging 10-min window is constantly (or at least mostly) higher than the cut-out WS, control will shut-down the WT and will keep it in that state, resulting in P_{out} (close) to zero, i.e., there will be no outliers in the recorded data. However, if the WS fluctuates around the cut-out WS, there might be periods within 10-min interval with WS lower than the cut-out WS, when WT will generate some output and when resulting 10-min average P_{out} values are somewhere between the rated power and zero, depending on the proportions of operational and non-operational times in the 10-min window. Essentially, similar averaging effects will also happen near the cut-in WS.

The common causes of outliers in WT and WF measurements are summarised in Table 5.3.

Table 5.3 Summary of Common Causes of Outliers in WT and WF Measurements

Causes	Ref. (WT-Level)	Ref. (WF-Level)
WT downtime, offline operation, or outage	[156]	[174]
WT curtailment, down-rating, or derating	[11, 156, 158, 166, 175]	[11, 172, 174]
(Manually set) WT constrained operation	[16, 162]	-
Data acquisition failure	[15, 156, 159, 160, 168, 171]	[146, 167, 172]
Inaccurate measurements caused by improper measuring location (usually for WS), or WS under-reading	[174]	-
Data transmission (communication system) failure	[11, 151, 156, 160, 168, 171, 173]	[11, 167, 172, 174]
Data processing failure	[151, 171]	-
Data management failure	[11, 151, 160]	[11]
Electromagnetic interference	[151, 156]	-
WT control system problem, including pitch control malfunction, blade pitch angle error, control program problem, yaw and pitch misalignment, incorrect or unsuitable controller setting and similar	[16, 82, 151, 158, 162, 166, 171, 175]	-
WT blade damage, or excessive deposition of dirt, ice or insects	[11, 82, 156, 158, 162, 171]	[11]
Wind shear	[176]	[176]
Shading or wake effects from neighbouring WTs, terrain or physical obstacles	[16]	-
Extreme weather conditions, harsh environment	[156, 160, 171, 173, 175]	[167]
Air density fluctuations	[16]	-
Issues related to averaging period	[16, 168]	-
WT cut-out effects	[168]	-
Wind speed (around the cut-in or cut-out wind speeds)	-	[146]
WT data availability issues	[176]	[176]
Other, including WT malfunctions, alarms in WT, low level of gearbox oil, worn-out generator brushes, sensor accuracy, and various errors in sensors	[11, 15, 82, 151, 159, 162, 171, 173, 175]	-
Other, including WT malfunctions, fluctuations in WT performance, multiple non-meteorological factors and various errors in sensors	-	[11, 146, 147, 167, 172, 174]

5.4 Detection of Outliers in Processing Wind Turbine and Wind Farm Measurements

To the best of the PhD student's knowledge, there is a lack of a comprehensive and systematic research on the detection and treatment of outliers in wind energy applications, [156], as opposed to some other areas, such as biological data, astronomy data, web data, information network data, economics data, medical data, etc., where comprehensive and structured reviews are available (e.g., [185, 186]). Specifically according to [186], approaches for detection of outliers can be divided in three general groups: unsupervised, semi-supervised and supervised, depending on the availability

of data labels, i.e., availability of information on conditions under which the data are obtained. Unlike the literature related to other areas, which usually adopt general statistical approaches, one important contribution of this chapter is that it specifically focuses on the WT and WF applications, providing inputs, illustrations and specific analysis for understanding the causes of outliers, aimed at a better understanding of WT and WF operations through establishing robust methods for detecting true outliers. In terms of WT-level and WF-level measurements, there is usually a lack of documentation and information about important operational conditions, such as wind curtailment [11, 156, 174], WT outage [174] and WF operating states [167]. Similarly, information on the possible causes of outliers, such as data storage mishandling and data communication errors, or WT damage or presence of icing on blades, may also not be documented [11, 174]. As discussed in Section 3.2.2, information from operational logs and event logs may significantly improve detection of outliers [167]. In particular, [156] considers wind curtailment as the source of abnormal data simply because information on wind curtailed operation may not be available. On the other hand, [174] states that planned WT maintenance is not considered as a source of outliers, because the planned maintenance is usually well documented.

The logs with the information on WT downtimes, malfunctions, faults and maintenance periods can be used, but these raw data are also not free from errors and may also require data cleaning, as [152] states. However, this reference focuses on condition monitoring, rather than resolving data quality problems, and fault labelled datasets are actually used for WT abnormality detection. Finally, [172] describes a semi-supervised algorithm for outlier detection, but it uses a limited number of labelled data.

Most of the outlier detection methods in reviewed literature belong to unsupervised approaches, as they did not have, or did not use any further explicit information on the exact causes of outliers. Like in previous discussion, the reported outlier detection approaches reviewed in this section can be also divided in statistical methods, physical constraint-based methods, or combinations of both. Robust statistical approaches proposed in some papers are also reviewed, although this work is not related to outlier detection, but to reducing the effects of outliers.

5.4.1 Statistical Methods for Outlier Detection

Statistical methods for the identification of outliers are most common and can be divided into density-based (aimed at separating sparse and dispersed points from the highly concentrated ones), distance-based (aimed at isolating points located far away from the normal ones, which usually can also be regarded as equivalent to the density-based method), correlation-based (aimed at finding the outliers using joint or conditional correlation analysis), and image-based (aimed at detecting the abnormal data by extracting the principle part from the binary image of P_{out} -WS scatter plot).

Regarding statistical methods, the most popular outlier detection algorithm is the quartile algorithm, which is usually combined with other more complex approaches. In descriptive statistics, the interquartile range (IQR) is defined as $\text{IQR} = Q_3 - Q_1$, where Q_3 and Q_1 are 75th and 25th percentiles, respectively. Quartile algorithms for outlier detection usually demonstrate that any data point outside the range $[Q_1 - k \times \text{IQR}, Q_3 + k \times \text{IQR}]$ may be an outlier, where specifically, $k = 1.5$ indicates outliers and $k = 3$ indicates data that are “far out” [187].

Another frequently used approach for outlier detection is the standard deviation-based method, which usually assumes that the analysed data have normal distributions. However, according to [177] and [188], the normality assumptions usually do not hold, as the P_{out} data in any WS bin cannot pass conventional statistical tests, e.g., Kolmogorov–Smirnov test and Lilliefors test. Like quartile algorithm, this method is usually applied with other more advanced approaches. According to the standard deviation-based method, for a set of sample data $x_1, x_2, \dots, x_n$, any point outside $[\bar{x} - k \times \sigma, \bar{x} + k \times \sigma]$ may be selected as an outlier, where $\bar{x}$ is the mean value of samples, σ is the standard deviation of samples and k is used to set the acceptable range of deviations.

Wind Turbine Measurements

Instead of applying quartile algorithm directly on the measurements, [152] argues that the negative deviations in some WS bins should be prevented and it applies sideband normalisation to regularise the deviations of the graded statistical features before applying the quartile algorithm. In [156], quartile algorithm is combined with a change

detection approach (i.e., change point grouping algorithm) to detect the outliers, where change point is defined as a point where one or more quantities in a series exhibit sudden change, therefore reflecting a qualitative change in the data. In terms of the stacked outliers (see outliers in Fig. 5.1 marked as ①, ②, ③ and outliers marked as A and B in Fig. 5.2(a)), the distribution characteristics of the large amount and highly concentrated abnormal data imply that the mean value, the variance and the change rate of related data sequence change as well [189]. In this reference, WS data are divided into several bins and the variance change rates of the ordered P_{out} values in each bin are calculated, in order to find the change points for distinguishing outliers from the normal data. As the change point grouping algorithm cannot efficiently identify all types of outliers, especially the scattered outliers (e.g., outliers marked as ④ in Fig. 5.1 and marked as D in Fig. 5.2(a)), [156] also adopts quartile algorithm to perform further outlier detection.

Unsupervised clustering algorithms are another widely applied method for analysing outliers in WT measurements, as [11] shows. This reference uses density-based spatial clustering of applications with noise (DBSCAN) algorithm [190], where main idea is that for each point of a regular data cluster, the neighbourhood of a given radius must contain at least a minimum number of points. Accordingly, the approach in [11] first identifies boundary of outliers, with the purpose of eliminating stacked outliers, and then applies the quartile algorithm twice to identify sparse outliers, equivalent to detecting the outliers for both $P(P_{out}|WS \text{ bin})$ and $P(WS|P_{out} \text{ bin})$, which are notations of conditional probabilities: $P(P_{out}|WS \text{ bin})$ is the conditional probability of P_{out} given WS bin, while $P(WS|P_{out} \text{ bin})$ is the conditional probability of WS given P_{out} bin.

Similarly, [160] applies a sliding window technique to divide data into different segments that are considered as different objects, where their attributes consist of time-domain statistical features (e.g., mean, maximum and peak-to-peak values) extracted from each segment. Then, an improved local outlier factor (LOF) algorithm, known as kernel-based LOF (KLOF) algorithm [191], uses these attributes to evaluate whether the segments contain outliers. Traditional LOF algorithm [192] is a density-based unsupervised learning approach, which detects abnormal data by measuring their local deviations from the neighbours, basically sharing some concepts with DBSCAN

algorithm, like core distance and reachability distance. Traditional LOF algorithm has two main shortcomings: 1) it may fail to detect outliers in a complex and large set of data, because it calculates the averages of reachability distance between an object and its neighbours in the local density estimation; and 2) it is very sensitive to the parameter selections. The KLOF algorithm transfers the unsupervised learning into the classic non-parametric regression learning and it tries to overcome the above two shortcomings of LOF by computing the regression estimator of a data point in its neighbours based on Nadaraya-Watson kernel regression.

In addition to DBSCAN, LOF and related improved algorithms, there are other clustering methods used in reviewed literature, as [171] demonstrates. This reference applies *k*-means clustering algorithm based on Squared Euclidean and City-Block distance to categorise the raw P_{out} -WS data into different groups. The detections of abnormal data are performed in the created clusters, where the local Mahalanobis distance of each data point to its cluster centroid is used to detect the outliers based on their score. Similarly, [147] also uses *k*-means clustering algorithm with a conservative approach based on the Mahalanobis distance metric to identify outliers. The main difficulties in traditional *k*-means clustering algorithms are selection of appropriate number of clusters, *k*, and sensitivity *k*-means of obtained results to the initialisations. Therefore, [166] uses an improved version of *k*-means clustering algorithm [193], where residual analysis method is used to automatically obtain both the number of clustering and initial centres from the decision diagram. Furthermore, [173] proposes an unsupervised outlier detection approach combining stacked denoising autoencoder (SDAE) [194] and density-grid-based clustering method [195, 196], where SDAE is utilised to extract data features, followed by the density-grid-based clustering method to detect outliers.

Some of the previous work used classification methods for outlier detection. One example is [162], in which *k*-nearest neighbour (KNN) classification algorithm [197] is used first to predict the desirable power output for the WT operating in normal conditions, which is constructed as a reference WT PC, and then outliers are detected using the residual approach (control theory) and control charts (quality control) [198] on the established PC model. In that way, KNN algorithm is used as a supervised

classification algorithm, which is trained on a selected (training) set of samples in a multi-dimensional feature space, where the labels of the training samples must be known. In the classification phase, for an unlabelled vector (a univariable or high-dimensional sample point), the algorithm will calculate and compare the distances from the k training samples, in order to determine the label to assign. Unfortunately, [162] does not illustrate how the labels for the training set are obtained.

In some references, correlation-based outlier detection models are used, where outliers are detected using Copula theory [51]. More specifically, [168] applies a 2-dimensional Gaussian mixture copula model (GMCM) [199] to fit the P_{out} -WS joint distribution, i.e., $P(P_{\text{out}}, \text{WS})$, where the outliers are excluded by a joint probability contour. Similarly, [169] also uses outlier detection method based on Copula. It uses mixed Archimedes Copula function to calculate the conditional distribution of WT P_{out} for given WS, i.e., $P(P_{\text{out}}|\text{WS})$, from which the confidence boundary of P_{out} is obtained. Acknowledging that the operation characteristics of a WT may change over time, it uses k -cross validation (k -CV) and sliding time window (STW) approaches to determine whether the estimation is stable.

Normal-distribution-based outlier detection approaches are also often reported in previous literatures. Reference [79] first applies exponential smoothing [200] on WT P_{out} series to account for WS fluctuations, where the original P_{out} series is denoted as $x_1, x_2, \dots, x_n$, and the transformed series as $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$. Afterwards, outliers are detected by a standard deviation-based approach, as data points outside the range $\bar{x}_{t-1} - k \times \sigma < x_t < \bar{x}_{t-1} + k \times \sigma$, where the subscript t denotes the data point at the t -th step, σ is the standard deviation, and k is a parameter determined from statistical analysis of low probability events ($k = 3$ is the choice in [79]). Similarly, [15, 82] both assume that distributions of WT P_{out} values in WS intervals/bins follow normal distributions and then use multiple standard deviations around the mean values to determine the outliers. Furthermore, [153] performs recursive data cleaning by utilising the piecewise linear model as a benchmark to remove the outliers, where the main criterion is that for a large number of data samples, there should be at least a certain percentage of normal observations distributed within ρ standard deviations from their associated fitted line segments. Finally, [151] proposes a deterministic PC model, considering

both WS and WD dependencies, i.e., $f(P_{\text{out}}; \text{WS}, \text{WD})$ using least median of squares (LMedS) fitting algorithm. One significant advantage of LMedS, compared to least mean square fitting algorithm, is that it is more robust, as the fitting results are determined by the median values instead of mean values, because the outliers are detected as the points with high residues according to the fitted model $f(P_{\text{out}}; \text{WS}, \text{WD})$, assuming the noise for the inliers has a normal distribution. The standard deviations of the errors from LMedS residuals can be estimated, which determines the thresholds to distinguish outliers from the inliers.

As discussed above, many outlier detection approaches are aimed at identifying the boundaries of possible P_{out} -WS ranges, and [158] develops a unique automatic PC limits calculation algorithm through iterations. The presented algorithm moves the estimated PC left/right or up/down while it searches for the optimal PC limits, trying to include only normal data. In the iteration loop, only the identified normal data from the previous iteration step are selected in each step as the new input data. The iterations stop when the percentage change of identified normal data is below a certain level (this approach is also used in [161]).

Examples of using hypothesis test theory to identify outliers are also found, where e.g., [170] uses a combination of four methods to determine whether a data point is an outlier, representing a cyber-attack state in a system state vector space. Firstly, an F-test [201] is employed, and the null hypothesis is that the data in two vectors come from the same variance for an attacked state vector. Because a small number of outliers will not lead to significant variance change, a majority voting is used if there is no sufficient evidence to reject the null hypothesis. If the results of at least two of three algorithms: fuzzy c-means (FCM) clustering algorithm [202], quartile algorithm and median absolute deviation (MAD) method [203] show that a data point is an outlier, then that data point is recognised as an abnormal data point.

A relatively new technique in the WT measurements outlier detection research is presented in [175], based on WT PC images. After pre-cleaning of some obvious and easy-to-detected outliers (e.g., ① in Fig. 5.1 and A in Fig. 5.2(a)), the obtained WT P_{out} -WS scatter plot is transformed, so it will form a corresponding binary image. Then, the principal part of the binary image, representing the normal data, is extracted

by the mathematical morphology operation (MMO) [204], which is combined with the Hu moment [205].

Wind Farm Measurements

Regarding the WF measurements, [172] uses a limited number of labelled data to supervise a DBSCAN algorithm for the detection of outliers in WF measurements in order to avoid false dismissals and false alarms. The limited number of labelled data is used to form a semi-supervised approach that could lead to a good balance between artificial label-setting of supervised learning and detection precision, despite the fact that DBSCAN itself is an unsupervised technique. Reference [167] utilises LOF algorithm using weighted Euclidean distance as a similarity measurement to detect outliers in WF measurements. In particular, as discussed in Section 5.3.2, [167] states that the vertical band around the cut-out wind speed in the WF P_{out} -WS scatter plot are not outliers, but are caused by wake effects. The results presented there show that LOF algorithm also does not classify these recordings as outliers, which could be simply due to the fact that these data have high densities, as otherwise LOF, a density-based clustering algorithm, will not be able to denote them as the normal data.

The KNN classification based algorithm is also used to detect outliers in WF measurements, with [146] applying approach very similar to [162], as discussed in Section 5.4.1. Correlation-based outlier detection models are applied for WF measurements as well. Copula theory is used to fit the correlation model in [174], where probabilistic WF PC is obtained by the Copula conditional quantile method and where outliers are identified as the points outside its upper and lower boundaries. Finally, [164] first uses a modified moving average algorithm to smooth the raw WF measurements and then applies following censoring rule to remove the abnormal data: if $y_i > a_1(x_i^3)^2$ or $y_i < a_2x_i^3 \cap y_i < a_3$ ($i = 1, 2, \dots, n$), then i -th element is removed, where x is the WS, y is the normalised P_{out} measurement, and a_1 , a_2 , and a_3 are the parameters that should be tuned.

5.4.2 Physical Constraint-Based Methods for Outlier Detection

There is a relatively small number of references dealing with the detection of outliers based on physical constraints. Furthermore, most of the rules applied for setting the constraints are ambiguous. For example, [16, 154] state that inputs and outputs of the WT measurements must be within the expected ranges, which have to be chosen extremely carefully, and that inputs and outputs should be mutually consistent. However, these references do not show explicitly how the ranges and consistencies could be determined, as they focus on condition monitoring, not on data pre-processing for the identification of outliers. Regarding the WF measurements, [176] states that the first step is to remove recordings when less than 272 WTs are available in a WF with total of 274 WTs, adding that this is an arbitrary criterion, so that the remaining data will have less than 1% uncertainty in P_{out} , since $2/274 = 0.73\% < 1\%$. Afterwards, [176] performs general outlier detection checks, removing outliers when: 1) positive P_{out} values occur for $WS < 2$ m/s; and 2) constant WS values are recorded for longer than two consecutive 10-min averaging windows.

5.4.3 Outlier Detection by Combinations of Statistical and Physical Constraints-Based Methods

Reference [177] firstly determines the upper bound of possible WT P_{out} values according to Betz's law, which specifies the maximum power output for given air density, rotor area and input wind speed. Then, it uses method of bins and quartile algorithm to detect the remaining outliers, stating that if the WT P_{out} histogram has multiple modals, only the highest populated interval in the histogram is considered, while P_{out} recordings near the lower mode(s) are excluded.

Table 5.4 summarises possible problems or limitations of outlier detection methods reviewed in Sections 1175.4.1 to 5.4.3.

Table 5.4 Summary of Possible Problems or Limitations of Reviewed Outlier Detection Methods in Practical Applications

Method		Possible Problems or Limitations	Refs.
Type	Name		
Classification	KNN	Plenty of normal data as samples; computational burden	[146, 162]
Clustering	k -means	Usually difficult to determine the parameter k	[147, 171]
	Improved k -means	Does not work well for abundant outliers	[166]
	DBSCAN	Sensitive to manually set parameters; does not work well for abundant outliers	[11]
	(Semi-supervised) DBSCAN	Limited number of labelled data may not have the same statistical properties as the test data	[172]
	LOF	Does not work for stacked outliers; does not work well for abundant outliers	[167]
	KLOF	Does not work well for abundant outliers	[160]
	SDAE and density-grid-based clustering	Does not work well for abundant outliers	[173]
	FCM	Does not work well for abundant outliers	[170]
Correlation	Copula	Stacked outliers could not be effectively filtered	[168, 169, 174]
Distance	Normal distribution	Assumption of normal distribution is not always suitable [169]	[15, 79, 82]
	Quartile algorithm	It is not effective when the proportion of outliers is large	[152, 156, 177]
	MAD	The selection of MAD range may not be applicable in specific situations [203]	[170]
Change point	Change point grouping	Need to manually set parameters; ignores the overall distribution; does not work well for the abundant stacked outliers	[156]
Recursive or iteration data removing		Only works for dataset with large number of recordings, so the algorithms can converge; difficult to determine a stopping criterion	[153, 158, 161]
Statistically robust data fitting		Computational burden [151]	[151]
Data smoothing and censoring		Parameters must be carefully tuned to adapt to different cases [11]	[164]
Computer visual		Cannot distinguish false outliers near the cut-out wind speed	[175]
Hypothesis test	F-test	Assumes that data are normally distributed and that samples are independent [201]	[170]
Physical	Various consecutive WS	Cannot detect outliers caused by P_{out}	[176]
	Expected ranges and mutually consistent	Cannot prove the correctness of selected ranges	[16, 154]
	Betz's law	Can only determine the theoretical upper PC bound	[177]

5.4.4 Robust Statistical Models for Reducing Outlier Effects

Robust statistical models are proposed in some references to reduce the effects of outliers, where, arguably, the main advantage is that an explicit detection of potential outliers is not performed. For example, hidden Markov model (HMM) based on fuzzy scalar quantisation [206] is proposed in [157] for WT fault diagnostics. As few outliers occurring in the HMM training set may lead to a misdiagnosis, fuzzy scalar quantisation assigns to each state in the signal quantisation a discrete probability, rather than a single value, to reduce the effects of such outliers. Furthermore, [159] employs the weighted version of least squares support vector regression (weighted LS-SVR) from [207] to model the relationship between weather conditions and WT P_{out} , stating that a single additional weighted LS-SVR iteration is generally sufficient for excluding the outlier effects by giving minimal weights for the observations with large errors following specific rules [208]. Finally, in [178] WF P_{out} prediction is analysed by combining outlier smooth transition autoregressive (OSTAR) approach and generalised autoregressive conditional heteroskedasticity (GARCH) models, where resulting OSTAR-GARCH model is used to capture the regime switching between the “outlier shocks” and other values in the volatility of WF P_{out} time series, where outlier shocks are defined as the values outside $\pm 2\hat{\sigma}$ interval, with $\hat{\sigma}$ being the estimate of the residuals’ standard deviation.

5.5 Approaches for Testing Success of Outlier Detection and Removal/Treatment Procedures

There is a general lack of benchmark data and labelled, or well-documented data that can be used for testing the accuracy/success of outlier detection and removal/treatment methods. The most widely used method to evaluate the success of outlier detection/treatment approaches is a simple visual check. For example, [145] uses “visual inspection” to ensure that resulting power curve is without outliers, while [153] compares the scatter plots before and after removing outliers concluding: “... it can be seen that the outliers included in the raw data are effectively eliminated in the final dataset”.

In some literature, outlier elimination rate is used as an indicator of the performance of proposed outlier detection/elimination methods. Other widely employed metrics include: confusion matrix, sensitivity (i.e., true positive rate, TP rate), specificity (i.e., true negative rate, TN rate), false positive rate (FP rate), false negative rate (FN rate), and F-score (i.e., F-measure) (e.g., [151, 170, 172-174]). Each row of confusion matrix represents the instance in an actual class, while each column represents the predicted instance, or vice versa. The TP and TN are outcomes where the model correctly predicts the positive and negative classes, respectively. The FP means the model predicts the positive class by mistake, while FN refers to the model mistakenly predicting the negative class. The formulas of TP rate, TN rate, FP rate, FN rate and F-score are listed as following:

$$\text{TP rate} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (5.1)$$

$$\text{TN rate} = \frac{\text{TN}}{\text{TN}+\text{FP}} \quad (5.2)$$

$$\text{FP rate} = \frac{\text{FP}}{\text{FP}+\text{TN}} \quad (5.3)$$

$$\text{FN rate} = \frac{\text{FN}}{\text{FN}+\text{TP}} \quad (5.4)$$

$$\text{F-score} = \frac{\text{TP}}{\text{TP}+0.5 \times (\text{FP}+\text{FN})} \quad (5.5)$$

Furthermore, the comparisons of the statistical properties between the measurements and known datasets without outliers are also considered (e.g., [174]), while comparisons of power curve fitting errors and comparisons of data concentration indices are used to prove the correctness of outlier removing approaches. More specifically, [156] emphasises that the outliers can be effectively identified: “the data deleted in the three turbines are basically outside the normal power output ranges, indicating that this outlier identification algorithm is effective”, as the remaining normal data in the WT P_{out} -WS scatter plot are “close to the ideal-state wind power curve”. However, the definition of normal power curve range is not provided in this reference, which also compares the proposed outlier identification method with other methods based on the outlier elimination rate. Furthermore, [11, 166] both calculate

the outlier elimination rates and compare PC modelling errors before and after removing outliers from the raw measurements, stating that “for a well performed outlier elimination method, both errors of the power curve modelling and the elimination rate should be as small as possible”. Additionally, [167] states: “the simplest way to evaluate the accuracy of the algorithm is by visual inspection” and uses confusion matrix to show the effects of the outlier detection method, but it does not illustrate how the true states in the confusion matrix are obtained. Finally, [169] uses R-squared method to indicate the concentration of data, declaring that a higher concentration means better data cleaning effect.

In terms of a strict analytical or mathematical formulation, visual check/inspection is not an accurate and scientific way to assess the results of outlier detection methods. The use of outlier elimination rate as a performance index is also unreliable, especially when there is a lack of labelled data, or otherwise clearly indicated outlier data. Essentially, obtaining both higher and lower outlier elimination rate is misleading, because neither imply that one method is better or worse than another. Furthermore, comparison of power curve fitting errors, or data concentration indices, may be meaningless, as almost all outlier removing methods lead to a filtered dataset with higher concentration, which will certainly have better fitting results and smaller residuals. However, evaluations through confusion matrix or comparisons of data statistical properties are reasonable, but they inevitably require knowledge of the true outlier labels in the measured data.

In terms of the treatments of the identified outliers, the majority of previous work simply decide to eliminate (i.e., to remove, or delete, or filter-out) outliers after they are detected, without any further processing, e.g., [11, 15, 16, 79, 82, 146, 147, 151, 153, 154, 156, 158, 161, 162, 164, 167, 168, 171, 176, 177]. For example, [168] excludes sparse outliers caused by window-averaging and WT cut-out effects (as discussed in Section 5.3.4), although it states that these values are not due to the faults or otherwise anomalous WT operations. Similarly, [168] keeps the outliers caused by WT curtailments, which is probably because of the used outlier detection method, GMCM, in which these outliers will have higher joint probability densities than the threshold, allowing to identify them as normal data.

Although not explicitly recovering deleted outliers, [11] mentions that these recordings can be corrected by temporal and spatial interpolation methods [209-211], and that the fitted power curve based on normal measurements could also be utilised to infer the P_{out} values for given wind speeds as inputs. Furthermore, [167] states that some imputation approaches (e.g., [212]) can be used to replace (or predict) missing data due to outlier filtering, but it argues that these efforts may not be necessary for two reasons: 1) the outliers are often consistent and lasting for longer periods, and there are no sufficient data to make a smooth imputation; 2) there will be enough remaining data to obtain interesting patterns, as usually a relatively small portion of outliers is removed. Similarly, [154] also stipulates that neglecting outliers instead of approximating them is acceptable, since there is a large number of remaining data available after the outliers are neglected. However, that may not be true in case of e.g., measurements around and above cut-out WS, which are generally with low probability (e.g., in a WT example in Fig. 5.2(a), only 0.5% of the measurements over a 3-year period were recorded for WS' around and above the cut-out wind speed).

A very few references in existing literature try to recover outliers by data imputation. An example is [169], which proposes a bi-directional Markov chain interpolation method to recover the missing values, stating that there will inevitably be a lot of missing data if the stacked outliers caused by wind curtailment are eliminated. Another example is [179], where the outliers are treated as missing values and imputation of these data is performed using multilayer perceptron (MLP) and adaptive neuro-fuzzy inference (ANFIS) networks [213, 214], stating that it is crucial to have an uninterrupted historical dataset, especially for forecasting applications. Finally, [174] exploits the spatial correlation between adjacent WFs to correct the P_{out} outliers in a WF, but outliers in the WS data are not considered.

It should be noted that removing of outliers without recovering or replacing them will result in a new set of additional missing data, which are different from, e.g., missing measurement data due to failures and faults of measurement systems, or empty recordings. Essentially, these genuine missing data are not outliers, as they do not have a value, which could be assessed in terms of their distance from the reference/regular set of data. Current literature (e.g., [183]) usually classifies missing data into three

general types: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR, also known as nonignorable, NI). MCAR means that the reasons for the data being missing are independent of observable and unobservable parameters and missing occurs entirely at random. MAR refers that the probability of missing is independent of the missing values and it may be dependent on the observable values, while MNAR or NI occurs when the missingness is systematically related to the unobserved data. An example of MNAR and NI is that the anemometer may age over time without being noted, which could make its measurements not precise, and the distribution of the measurements thus is distorted. The anemometer may also have different missing probabilities for different wind speeds. In this case, those missing data are believed to be MNAR. If missing data are MCAR, then no more new information can be obtained from the missing data and subsequent analysis of the observable data without considering the missing data will not introduce bias. However, MCAR usually does not happen in practice. For MAR and MNAR, the missingness should be taken into consideration, as the data are missing systematically and without them the analysis will be biased [183]. In existing literature, some of the authors state that their methodologies for filling-in missing data (e.g., multiple imputation) are for MAR only and not for MNAR. Furthermore, it is suggested that in some cases an application-specific approach will significantly outperform more general models (e.g., multiple imputation). Based on the previous discussion in this chapter, the missing data caused by removing outliers due to cut-out effects will result in the nonignorable MNAR-type of missing data (the probability of missing data in the WF fully-operating regime is increasing as the WS approaches the cut-out wind speed, meaning that the values of these data are the reasons why they are missing).

5.6 Discussions and Conclusions

This chapter presents a comprehensive and systematic review of approaches for the detection and treatment of outliers in processing WT and WF measurements. Starting from the analysis of typically available measurement data, it is first found that most of the previous work was focused on the outliers in WT measurements, while less attention was paid to the analysis of outliers at WF-level, particularly when aggregation of individual WT measurements is used to model the whole WF

characteristics. The most widely considered variables are WS and P_{out} . If other recordings are available, the particular attention is given to data normalisation (e.g., power output, wind speed, and turbulence intensity), which is also required by the relevant standard [12]. Afterwards, the chapter gives a review of the definitions, characteristics and causes of outliers in WT and WF measurements reported in existing literature, including discussion of outliers due to faults and failures in data acquisition/measurement systems, but also outliers caused by period averaging, cut-out effects, curtailment, etc., introducing simple distinction between true outliers and false outliers. This part of the analysis also includes an examination of data rejection requirements in the relevant standard [12]. Next, the chapter reviews methods for the detection of outliers, concluding that the statistical methods (density-based, distance-based, correlation-based, or image-based) are the most popular, while the physical-constraint based methods and combined statistical and physical methods are reported in a relatively small number of references. Additionally, the chapter also reviews work aimed at building outlier-insensitive robust statistical models for reducing the effects of outliers without identifying them. Finally, the chapter reviews various approaches for assessing the accuracy/success of outlier detection methods, including visual inspection, data elimination rate, comparison of statistical properties, evaluation of PC fitting errors, and data concentration indices. In terms of outlier treatment procedures, it was found that in most of the previous work outliers are simply removed and that only a small number of references treat the outliers as missing data and perform data imputation. This effectively results in additional sets of missing data, where of particular concern are missing data due to elimination of outliers caused by cut-out effects, representing less-probable, but still plausible and intended operating points of a WT/WF. This is further discussed in Chapter 6, where a new method for the identification of outliers was presented in order to allow full inclusion of WT/WF operating points around the cut-out WS in the formulation of new 6-parameter PC-based models.

The main difficulty reported in the literature on outlier detection and treatment is a general lack of information on the actual causes of outliers, e.g., operational/event logs and similar documentations, which effectively prevent successful, reliable and convincing testing of the proposed outlier detection methods. In different researches,

the dimensions and resolutions of available measurements may differ. The larger number of data could increase the confidence of the detection of some outliers, but that may also introduce other additional or new outliers.

As discussed, the sources of outliers are many and various, of which some may result in outliers with similar, or even the same patterns. Additionally, some of these outliers may be masked and hidden by the normal data. Therefore, it is sometimes almost impossible to determine the exact causes of outliers, even if their presence may be successfully detected. WT-level outliers and WF-level outliers are different, since some outliers typically occur at a WT-level and not at a WF-level (and vice versa), as the WT and WF operating regimes are generally different. For example, data recorded during the outage of an individual WT are true outliers for the analysis at WT-level, but these data generally should not be excluded from the analysis of operational performance of a WF, i.e., these data are false outliers at WF-level (see Chapter 6 and Chapter 9 for further analysis and evaluation of WF operating regimes). This chapter did not find standard or unified approaches for treatments of outliers, as different approaches are highly dependent on their purpose, target analysis, or intended use of the processed measurement data. Generally, this diversity may make different approaches proposed for outlier treatment incomparable. Finally, the occurrences and patterns of some outliers may strongly depend on the specific types of WTs, their locations and variations of their operating, control and environmental conditions. These factors present additional difficulties for both the provision of the general test benchmark values with labelled data and for the specification of commonly accepted methodologies for the confident identification and adequate processing of outliers.

Chapter 6

Detection of Outliers in Wind Farm Measurements and Operational Wind Farm Model

6.1 Introduction

Due to the significant increase of the number of wind-based electricity generation systems, it is important to have accurate information on their operational characteristics, which are typically obtained by processing large amounts of measurements from the individual wind turbines and the whole wind farms. For further processing of these measurements, it is crucial to identify and remove abnormal data, i.e. outliers, as otherwise obtained models may be biased or erroneous. This chapter extends the review and analysis in Chapter 5, and presents results of analysis and research using two datasets of different WFs. The contributions of this chapter can be summarised as the following:

- Present a generic framework that can handle WT-level outliers from both statistical and physical points of view.
- Propose a simulation-based approach that can explain outliers due to WT cut-out effects and applied averaging window.
- Analyse the relationship between WT and WF level modelling data.
- Propose a six-parameter power curve model that can represent both the cut-in and the cut-out regions accurately.
- Demonstrate the relationship among WF operating regimes.
- Propose an equivalent WF model considering various operating regimes. The model shows high accuracy for two general situations: 1) WF data are derived from individual WTs in the same WF; 2) only the whole WF recordings are available.

After the introduction section and description of used datasets, this chapter first identifies outliers in WT recordings using a framework consisting of sliding window and a novel simulation-based random walk Metropolis approaches, and manufacturer power curve with hysteresis control. A six-parameter power curve model is proposed and fitted on the filtered WT measurements, which can simultaneously model the cut-

in and cut-out regions. Then, WF-level analysis is performed for the case of known WT recordings. An equivalent WF model is presented, proving the relationships among various WF operating regimes. Afterwards, another WF-level case study without detailed WT measurements is investigated, where an iterative algorithm is proposed to fit the proposed equivalent WF model. Both case studies show the importance of considering actual operating conditions/regimes in identifying outliers, demonstrating that the proposed model can improve estimations of WF power outputs.

6.2 Description of Available Datasets

The illustrations and case studies in this chapter are based on two available datasets.

Scotland Wind Farm Dataset

This dataset consists of 3-year simultaneously recorded (at WT nacelle) 10-minute average values of wind speed, wind speed standard deviation (WS std.), wind direction and power output for six individual WTs in a WF in Scotland, UK. The WF consists of six Vestas V90-3MW wind turbine units with doubly-fed induction generators [87], sited on relatively flat terrain. The coordinates of the WF are 55°44'01.1"N 4°46'58.1"W. Its layout is depicted in Fig. 6.1, and its parameters are shown in Table 6.2. The WT cut-in, cut-out and restart WSs are 4 m/s, 25 m/s and 20 m/s, respectively. The rated P_{out} is 3 MW, and the rated WSs are between 16 and 19 m/s, depending on the air densities (0.97 kg/m³ to 1.27 kg/m³, according to the manufacturer specification in [87]). Table 6.1 below provides a summary of the WT operating characteristics.

Table 6.1 Operating Characteristics of Vestas V90-3MW [87]

Air Density [kg/m ³]	Cut-in Wind Speed [m/s]	Rated Wind Speed [m/s]	Cut-out Wind Speed [m/s]	Rated Power [kW]
0.97	4	19	25	3000
1	4	19	25	3000
1.03	4	18	25	3000
1.06	4	18	25	3000
1.09	4	18	25	3000
1.12	4	17	25	3000
1.15	4	17	25	3000
1.18	4	17	25	3000
1.21	4	17	25	3000
1.225	4	16	25	3000
1.24	4	16	25	3000
1.27	4	16	25	3000



Fig. 6.1. Dalry wind farm layout.

Table 6.2 Dalry Wind Farm Parameters

WT Name	WT Cut-in WS [m/s]	WT Cut-out WS [m/s]	WT Rated Power [MW]	Number of WTs	WF Location
Vestas V90-3MW	4	25	3	6	55°44'01.1"N 4°46'58.1"W

Every individual WT has synchronous WS, WS std., WD and P_{out} recordings. The total power output of the WF is not directly measured, but there is a wind mast in the WF, which provides simultaneous measurements of WS and WD for the whole WF.

The hourly temperature, relative humidity, and barometric pressure observations in the same area are found in [107], and they are interpolated linearly to align with the time stamps in WT/WF measurements. These additional meteorological recordings are used to calculate the air densities [12].

Zelengrad Wind Farm Dataset

This dataset has 4-year single-sensor simultaneously measured typically at a 2-minute resolution (but also with higher 1-minute resolution) average wind speed and power output values for a WF (with 14 WTs) in Dalmatia region in Croatia. The type of WTs in this dataset is also a Vestas V90-3MW wind turbine [87], which has the same

specifications as in the Scotland Wind Farm dataset. The location of Zelengrad WF is 44°07'30.0"N 15°44'16.8"E, and its parameters are listed in Table 6.3.

Table 6.3 Zelengrad Wind Farm Parameters

WT Name	WT Cut-in WS [m/s]	WT Cut-out WS [m/s]	WT Rated Power [MW]	Number of WTs	WF Location
Vestas V90-3MW	4	25	3	14	44°07'30.0"N 15°44'16.8"E

Fig. 6.2 illustrates the outliers in the measured data using example P_{out} -WS scatter plots based on Scotland Wind Farm Dataset and Zelengrad Wind Farm Dataset, where scatter-points power curves (PCs) are obtained using method of bins [12]. The figures show that the fluctuation of air density would cause significant variety of power output even at the same wind speed, represented by the range between Mfr PC $\rho_{0.97}$ and Mfr PC $\rho_{1.27}$, especially between 10 m/s and 15 m/s. In Fig. 6.2(b), the WF WS and P_{out} are derived from individual WTs measurements, following the method in [142]. More specifically, for synchronous individual WTs recordings where both WS and P_{out} are available, the WF WS is the average of all WTs' WS, while WF P_{out} is the sum of all WTs' P_{out} , which can be per-unitized when divided by the rated power output of the WF.

All three subplots in Fig. 6.2 have recordings outside the normal trends, which may either be highly concentrated on certain regions or sparsely located. Density-based method is very popular in WT and WF measurements outlier detection as discussed in Chapter 5 and Chapter 5, e.g., density-based spatial clustering of applications with noise (DBSCAN) and local outlier factor (LOF) algorithms. The results of directly applying DBSCAN and LOF on the Scotland Wind Farm Dataset and Zelengrad Wind Farm Dataset are also depicted in Fig. 6.2, with settings of DBSCAN and LOF parameters following [11] and [167], respectively. Both DBSCAN PC and LOF PC are almost identical to the PC calculated using method-of-bins [12] based on raw measurements (scatters PC), demonstrating that a single/simple density-based method does not work well in the case of stacked or abundant outliers.

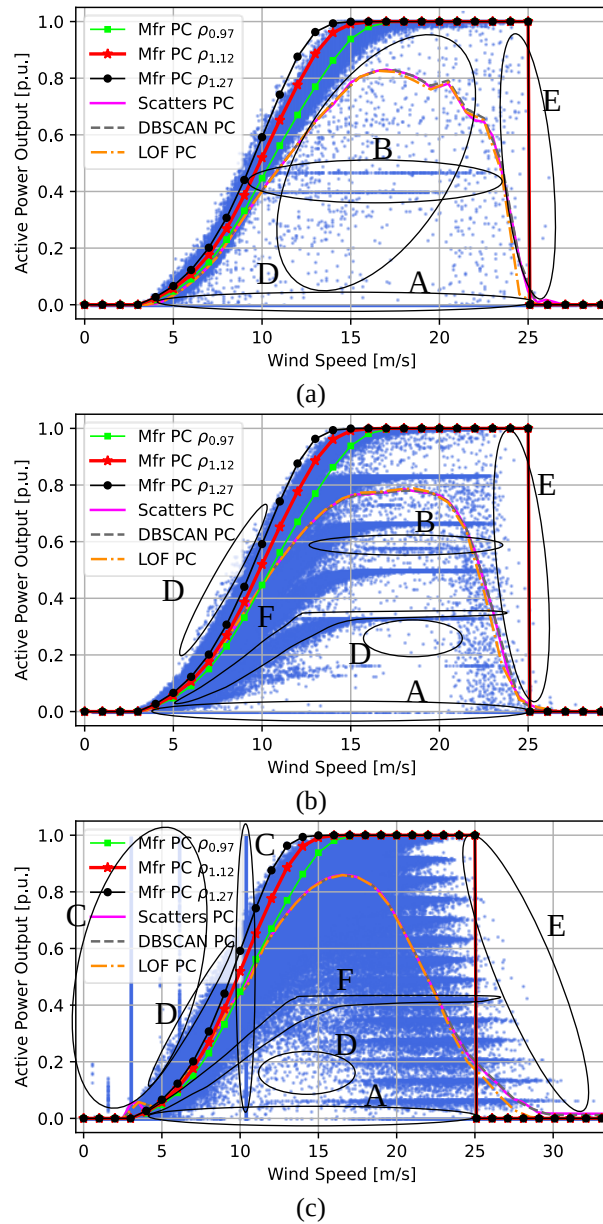


Fig. 6.2. Actual measurements with indicated different types of outliers. $\rho_{0.97}$, $\rho_{1.12}$ and $\rho_{1.27}$ refer to air densities equal to 0.97 kg/m^3 , 1.12 kg/m^3 and 1.27 kg/m^3 , respectively: (a) 3-year recordings for an individual WT in Scotland Wind Farm Dataset (WT1); (b) 3-year recordings for the whole WF in Scotland Wind Farm Dataset (derived from the individual WTs in the same WF [142]); (c) 4-year recordings for a directly measured whole WF in Zelengrad Wind Farm Dataset.

Like the discussions in Chapter 5, the outliers' characteristics and analysis are presented in Table 6.4. Some sources of outliers may result in outliers with similar, or even the same patterns and some of these outliers may be masked and hidden by the normal data. Therefore, it is sometimes almost impossible to determine the exact

causes of outliers, unless additional information (e.g., from a separate log of events) is available (see Chapter 5).

Table 6.4 Summary of Characteristics and Causes of Outliers in WT and WF Measurements

Characteristics	Possible Causes	Category
Low P_{out} -high WS (A in Fig. 6.2)	WT failure or outage, unplanned WT maintenance, faults of measurement/communication system, or WT averaging/cut-out effects	False outlier (normal data) if due to averaging/cut-out, otherwise CAT-III
Low maximum P_{out} (B in Fig. 6.2)	WT curtailment, damage of WT gearbox bearing, measurement/communication system errors, or (large WTs with multi-winding machines) loss of one power electronic converter and still operating at partial power	CAT-I
Linear P_{out} -WS (e.g., constant WS-variable P_{out}) (C in Fig. 6.2)	Linear interpolation, malfunction of pitch-control system, or dirt deposits on the WT blades. Errors in data acquisition and data transmission, wake effects, or P_{out} fluctuation around cut-out	CAT-II
Scattered not near the cut-out WS (D in Fig. 6.2)	Signal propagation noise, sensor failure, extreme weather conditions, averaging window, WT cut-out effects, or uncontrolled random factors	CAT-IV
Scattered near the cut-out WS (E in Fig. 6.2)	Averaging window, or WT cut-out effects	False outlier (normal data)
Shifted PC (F in Fig. 6.2)	Outages (caused by protection or unplanned maintenance), or faults of individual WTs in a WF	False outlier (normal data)

In this thesis, the recordings at WT-level with low maximum P_{out} are denoted as CAT-I outliers (CAT is the abbreviation of category), while linear P_{out} -WS outliers are denoted as CAT-II outliers. Measurements characterised by low P_{out} -high WS, but not due to WT cut-out effects or averaging window, are CAT-III outliers. Similarly, the scattered outliers which cannot be explained by WT cut-out effects or averaging window are defined as CAT-IV outliers. In addition, shifted PC outliers could only be seen at WF-level, which is the WF-level representation of WT-level low P_{out} -high WS outliers. Shifted PC outliers are crucial in identifying WF operating regimes, a core idea behind the equivalent WF model proposed in this chapter.

6.3 Detection of WT-Level Outliers

This section proposes a systematic approach to detect the WT-level outliers. The approach has a “workflow” which first applies a sliding window to iterate over the time series measurements. The discrete difference is calculated for the windowed data,

which can further determine CAT-I and CAT-II outliers. Then, the rest of recordings are compared with the range defined by the lowest and highest air density specified Mfr PCs (Mfr PC $\rho_{0.97}$ and Mfr PC $\rho_{1.27}$), which represent the variations of P_{out} due to air density changes.

To distinguish the measurements outside the range, a novel simulation-based approach is proposed, which utilises Markov chain Monte Carlo (MCMC) to generate a high-resolution wind speed time series from the recorded 10-minute average WS and WS std. data. This novel simulation-based approach also implements WT hysteresis control when calculating P_{out} using air density specified Mfr PC from the simulated 3-second WS. If higher resolutions (e.g., 1 second) are used, the power output calculation from it will not be the synchronous value with the wind speed, because they are impacted by WT inertia. If lower resolutions (e.g., 1 minute) are used, the samples simulated may have already been masked by averaging effects. The resolution of the simulation should be as high as possible but also avoid reaching the region where inertia dominates. There are additional weeks of 1 Hz P_{out} -WS recording from another WF (Aikengall wind farm), which is installed with the same Vestas V90 wind turbines. The analysis shows there is around a 3-second delay (inertia) between input WS and power output, and thus 3-second value is selected in the simulation.

The standard [12] specifies that the wind turbine data (including wind speed) shall be collected continuously at a sampling rate of 1 Hz or higher. And the data acquisition system shall store either sampled data or statistics (including the standard deviation) of data sets. Therefore, the wind speed std. recordings imply the level of turbulence intensity within the 10 minutes at least at a 1-second resolution. The purpose of the simulation is not to recreate the exact high-resolution WS for one 10-minute recording, as it will be impossible. Instead, the simulation is used for uncertainty evaluation. It gives a set of possible high-resolution WS time series that will have the same mean value and standard deviation as the recorded ones, and the simulated data all follow a reasonable distribution (truncated normal distribution in this case). The uncertainty inferred from the simulated samples will be representative if the sample size is large enough, as they will cover sufficient possible scenarios behind the recorded mean WS and its standard deviation.

Afterwards, the boundary rule is applied to distinguish CAT-III and CAT-IV outliers. Finally, this section will summarise the further processing (i.e., to keep or discard) of the classified WT-level outliers in both the WT-level and WF-level analysis.

Importantly, the order of the outlier detection process is non-swappable, since each step will depend on the results of previous steps. The whole WT-level outlier detection process is visualised by a flow chart in Fig. 6.3, where the bold contents are the final classification results.

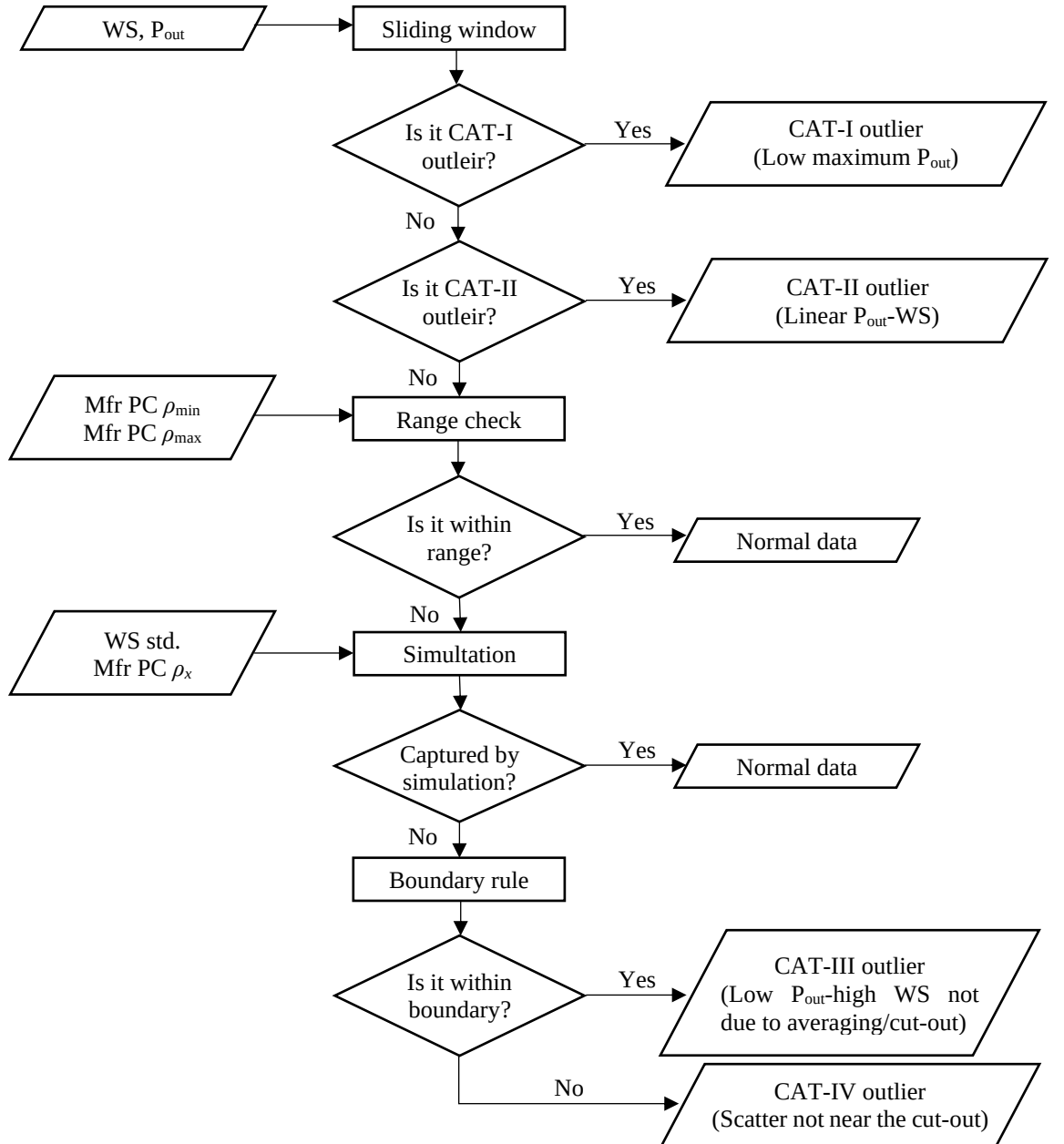


Fig. 6.3. WT-level outlier detection flow chart (Mfr PC ρ_x refers to air density specified manufacturer power curve).

6.3.1 Identification of CAT-I and II Outliers

Let $[\mathbf{T}, \mathbf{V}, \mathbf{P}_{\text{out}}]$ represent a synchronous wind speed and power output dataset, which is a matrix with three features, and $\mathbf{T}$, $\mathbf{V}$ and $\mathbf{P}_{\text{out}}$ are column vectors representing time stamp, WS and P_{out} , respectively. For total n measurements, $\mathbf{T} = [t_0, t_1, \dots, t_{n-1}]^T$, $\mathbf{V} = [v_0, v_1, \dots, v_{n-1}]^T$, and $\mathbf{P}_{\text{out}} = [p_{\text{out};0}, p_{\text{out};1}, \dots, p_{\text{out};n-1}]^T$.

Define a sliding window, $\mathbf{T}_w = [t_i, t_{i+1}, \dots, t_{i+j}]^T$, where $j \geq 1$, $i = 0, 1, \dots, (n - 1 - j)$, and $(t_{i+j} - t_i)$ is the length of the observation window (here, 30-minute window is used). In $\mathbf{P}_{\text{out};w} = [p_{\text{out};i}, p_{\text{out};i+1}, \dots, p_{\text{out};i+j}]^T$, if $\mathbf{P}_{\text{out};w}$ are smaller than 1.0 p.u. but larger than 0 p.u., and the absolute values of 1-st discrete difference of $\mathbf{P}_{\text{out};w}$, i.e., $[|p_{i+1} - p_i|, |p_{i+2} - p_{i+1}|, \dots, |p_{i+j} - p_{i+j-1}|]^T$, are smaller than a very small value (here, 0.0005 p.u. is chosen), then the recordings $[\mathbf{T}_w, \mathbf{V}_w, \mathbf{P}_{\text{out};w}]$ are CAT-I outliers.

Similarly, CAT-II outliers can be detected with the same windowed data. The difference is that CAT-I outliers usually relate to control settings of WT (WT curtailment) and manifest as horizontal bands/ranges of relatively constant P_{out} values below the lower boundary of the expected range, while CAT-II outliers are in more general forms. CAT-II outliers appear as linearly related (not limited to perfectly elastic nor inelastic) P_{out} -WS recordings. The constant WS-variable P_{out} outliers also represent a special case. In general, in a sliding window, excluding CAT-I outliers and recordings with P_{out} equal to 0 p.u. or 1.0 p.u., if the absolute values of 2-nd discrete difference of $\mathbf{V}_w$ are smaller than a very small value (here, set at 0.001 m/s), or the absolute values of 2-nd discrete difference of $\mathbf{P}_{\text{out};w}$ are smaller than a very small value (here, 0.0005 p.u. is used), then $[\mathbf{T}_w, \mathbf{V}_w, \mathbf{P}_{\text{out};w}]$ are CAT-II outliers.

6.3.2 Identification of CAT-III and CAT-IV Outliers

After detecting CAT-I and II outliers, which are formulated as rule-based methods as their causes are explicit, two lowest and highest Mfr PCs (Mfr PC $\rho_{0.97}$ and Mfr PC $\rho_{1.27}$) are used to define a range. The P_{out} variations of recordings within the range can be explained by air density changes, and they are classified as normal data. All the remaining measurements are analysed by a novel simulation approach. This proposed approach can avoid a big limitation of purely statistical methods, which could

successfully detect outliers in statistical aspect, but may also easily classify some physically reasonable (although statically unusual) data to be outliers.

Unlike purely statistical methods, the simulation approach leverages expert knowledge about the WT. It consists of two core techniques: random walk Metropolis (a gradient-free MCMC algorithm [215]), and P_{out} calculation using air density specified Mfr PC with hysteresis control. It can classify the remaining data into two subcategories: 1) recordings caused by averaging window or WT cut-out effects, which are physically possible; and 2) outliers due to other factors.

In the work of [12] and [181], they firstly make a simple assumption that the distribution of wind speed within 10 minutes follows a normal distribution (location parameter is the average WS within 10 minutes, scale parameter is WS std. within 10 minutes). They also assume that at each instant the WT follows a zero turbulence PC. Then, to normalise PC data to a reference turbulence intensity, they present a simulation-based procedure, which draws WS samples from the assumed normal distribution, applies a zero turbulence PC and calculates the expected value of the power outputs.

Instead of using a normal distribution, as an improvement, this section proposes to use the truncated normal distribution to represent high-resolution WS, because WS could not be negative nor infinity. The probability density function, f_{TN} , for $a \leq x \leq b$, is:

$$f_{TN}(x|u, \sigma, a, b) = \frac{1}{\sigma} \frac{\phi\left(\frac{x-u}{\sigma}\right)}{\Phi\left(\frac{b-u}{\sigma}\right) - \Phi\left(\frac{a-u}{\sigma}\right)} \quad (6.1)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and cumulative distribution function (CDF) of standard normal distribution, respectively; μ is location parameter; σ is scale parameter; a and b are the minimum and maximum of x ; $f_{TN}(x) = 0$ for x outside $[a, b]$. In this chapter, a is set to be 0 m/s, b is set to be 70 m/s (based on the highest expected wind gust value in [216]), while the values of μ and σ depend on the 10-minute average WS and WS std. in the actual measurements.

This chapter also prefers autocorrelated high-resolution WS samples in time series. Unlike the independent samples in [12, 181], autocorrelation is desired here, as real

wind speed measurements also have autocorrelation. To generate autocorrelated wind speed samples, random walk Metropolis can be used, and the simulated traces would have 3-second resolution for two reasons: 1) the WT mechanical inertia could not be neglected if higher resolution applied; 2) there are around 4-month of 1 Hz WS recordings in [216], and they are averaged by 3 seconds and used in the fittings of proposal generating functions of random walk Metropolis [215].

It is also crucial to correctly implement hysteresis control. The hysteresis control is a set of rules that start or shut down the WT. The exact hysteresis control is difficult to obtain for the exact type of WT at the given WF. Generally, there are four empirical triggers defined below to test the proposed simulation approach. In a real-world application, if the WF operator has detailed information about WT configuration, then the parameters in the assumed hysteresis triggers can be easily updated accordingly.

- Cut-in trigger: around the cut-in WS region ($WS < 6$ m/s) and when the WT is shut down, if the minimum WS in the past 3 seconds is at least 5 m/s, or the average WS in the past 60 seconds is at least 4 m/s, then WT will start up.
- Cut-in back trigger: around the cut-in WS region ($WS < 6$ m/s) and when the WT is generating, if the average WS in the past 60 seconds is at most(i.e., less than) 3 m/s, then the WT will shut down.
- Cut-out trigger: around the cut-out WS region ($WS > 17$ m/s) and when the WT is generating, if the minimum WS in the past 3 seconds is at least 30 m/s, or the average WS in the past 60 seconds is at least 25 m/s, then the WT will shut down.
- Restart cut-out trigger: around the cut-out WS region ($WS > 17$ m/s) and when the WT is shut down, if the average WS in the past 60 seconds is at most (i.e., less than) 20 m/s, then WT will start up.

The whole process of this novel simulation approach is described below:

Step 1: In synchronous dataset $[\mathbf{V}, \mathbf{P}_{out}, \mathbf{D}]$, where $\mathbf{V}$, $\mathbf{P}_{out}$, $\mathbf{D}$ are column vectors representing WS, P_{out} and WS std., for every recording $[v_k, p_{out;k}, d_k]$, initialise a target truncated normal distribution $f_{TN}(x|\mu=v_k, \sigma=d_k, a=0, b=70)$.

Step 2: For recording $[v_k, p_k, d_k]$, within that 10 minutes, starting from $t = 0$ s, initialize a random variable $x = x_{t0}$ by sampling from $f_{TN}(x)$ obtained in *Step 1*, which is known as the current state of WS. Meanwhile, initialise another variable $s = s_{t0}$ representing WT control signal, which is binary and can either be 1 (on) or 0 (off). If x_{t0} is higher than the cut-out WS, then s_{t0} is initialized to be 0, otherwise as 1.

Step 3: Define a proposal generating function centred at the current state. The proposal generating function implies the stochastic transitions between the current state and the next state by adding perturbation to the current state. Normal distribution, $g_N(y|\mu=x_{t0}, \sigma=h(x_{t0}))$, is used to model the perturbation, whose scale parameter σ is determined by its location parameter $\mu = x_{t0}$ through a function $h(\cdot)$ that takes WS as input and gives WS std. as output. $h(\cdot)$ could be learned from actual measurements in [216] using discrete first-order Markov chain, which estimates the standard deviation of all possible next states transiting from the current state.

Step 4: Sample y_{t0} from $g_N(y)$ in *Step 3* and calculate the acceptance ratio $\alpha_{t0} = f_{TN}(x_{t0} + y_{t0}) / f_{TN}(x_{t0})$.

Step 5: Determine whether to accept or reject the candidate for $t = 3$ s. Generate a uniform random number u between 0 and 1. If $u \leq \alpha_{t0}$, then let x_{t10} be $x_{t0} + y_{t0}$, otherwise $x_{t10} = x_{t0}$.

Step 6: Continue random walk Metropolis for WS for every 3 seconds within 10 minutes, until there are 200 samples of x , i.e., $x_{t0}, x_{t3}, \dots, x_{t597}$, which is defined as one trace/series of 3-second resolution WS'.

Step 7: The WT control signal s is calculated following the four hysteresis rules (triggers) as previously defined.

Step 8: Find the air density data synchronous with v_k in the Scotland Wind Farm Dataset, select one Mfr PC whose air density value is the closest. For the moments with hysteresis control signal s_t equal to 1, apply this Mfr PC and WS x_t and obtain the corresponding power outputs. The P_{out} values are set as 0 for the other moments. This step could yield one trace of 3-second resolution P_{out} values.

Step 9: Take the average value of *Step 8* results and obtain one simulated 10-minute average P_{out} for v_k .

Step 10: Repeat the above process of simulating one P_{out} for v_k 100,000 times (assuming 1 million sets of simulation is sufficient enough to provide robust estimation for the 2.5 – 97.5 percentiles, and also because increasing the simulation size further requires too much memory and computation time). These simulated 10-minute average P_{out} for v_k could define an acceptance range (in this chapter, 2.5 to 97.5 percentiles is used, which covers 95 % scenario, although different choices may be preferred in different cases). If actual recording $p_{\text{out};k}$ is inside the range, recording $[v_k, p_{\text{out};k}, d_k]$ is assumed to be normal. Otherwise, if the recording has v_k between the cut-in and cut-out WS and $p_k \leq 0$, then it is assigned as CAT-III outlier. The recording that does not belong to normal or CAT-III outlier is CAT-IV outlier. It means it can neither be captured by the simulation, nor it manifests WT shutdown pattern.

6.3.3 WT-Level Outlier Detection Results

Following the flowchart in Fig. 6.3 and the detailed discussions above, the corresponding outliers of all WTs in Scotland Wind Farm Dataset could be identified. Fig. 6.4 depicts the example results of WT1 (i.e., WT in Fig. 6.2(a)). Compared to Fig. 6.2(a), all data outside the logical trends have been assigned category labels. In particular, the cut-out region is reasonably represented, showing great P_{out} variation. In the case of extremely high WS std., even at low WS, due to the asymmetric hysteresis control, the WT may eventually have a 10-minute average P_{out} recording deviated from Mfr PC.

Fig. 6.5 shows the averages of percentages of outlier detection results for all six WTs in Scotland Wind Farm Dataset. CAT-IV outliers are the most frequently found outliers, followed by CAT-III outliers, taking together around 23.15 % of all recordings. CAT-I and CAT-II outliers are less than 1 %. Although CAT-I outliers are distinct on the 2D scatter plot, they do not form a high-density cluster, which further confirms that density-based clustering method may not be proper in identifying them.

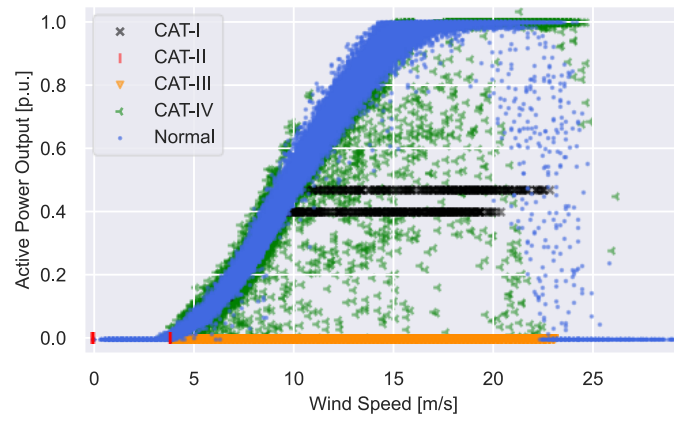


Fig. 6.4. WT1 in Scotland Wind Farm Dataset outlier identification results.

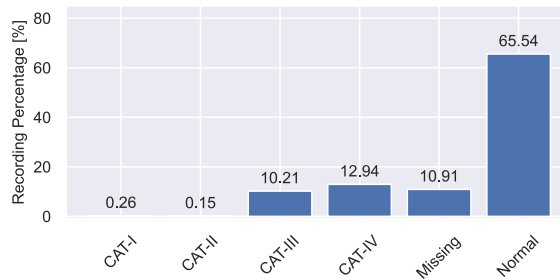


Fig. 6.5. The averages of percentages of outlier detection results for all six WTs in Scotland Wind Farm Dataset.

Table 6.5 presents the suggestions for the treatments of all categories of WT-level outliers in further studies at both WT-level and WF-level. CAT-II and CAT-IV outliers are true outliers at both levels, as they are more likely to be caused by data acquisition, transmission or processing, etc. CAT-I and CAT-III outliers should be removed in the WT-level study, but they are kept at WF-level. The reason is although cases such as WT failure or outage, unplanned WT maintenance, or WT curtailment do not occur very often (around 0.15 % of time according to Fig. 6.5) at the WT-level, the probability of WF operating with at least one WT shutdown or curtailed is very high, which will be discussed in more detail in Section 6.5.

Table 6.5 Suggestions for Treatments of WT-Level Outliers for both WT-Level and WF-Level Studies

WT-Level Outlier Category	WT-Level Study	WF-Level Study
CAT-I	Discard	Keep
CAT-II	Discard	Discard
CAT-III	Discard	Keep
CAT-IV	Discard	Discard

6.4 Six-Parameter Power Curve Model

As reviewed in Section 2.5.2, in the current literature, the five-parameter logistic function, 5PLF, provides very high accuracy in describing the P_{out} -WS relationship since it could address asymmetric data. Despite the high accuracy, one key disadvantage of 5PLF (and also its derivations including single four-parameter and six-parameter logistic functions) is that it could not simultaneously model the nonlinear segments around both the cut-in and cut-out wind speeds correctly, because a single logistic function is monotonic and it is impossible to use one single logistic function to model monotonically increasing and decreasing parts at the same time. To the best of the PhD student's knowledge, the behaviour of WT and WF near the cut-out wind speed is usually neglected in power curve modelling. To solve this problem, this section proposes to combine two 5PLFs sharing the same a and d parameters in Equation (2.8) of Section 2.5.2, since the P_{out} in cut-in and cut-out regions are both limited between 0 p.u. and 1.0 p.u. Also, because a and d control the positions of two asymptotes of 5PLF [83], this chapter fixes a and d to be 1 and 0, respectively. Therefore, there are effectively six parameters in the proposed 6P-PC model, whose equation is shown below:

$$6\text{P-PC}(x) = \min \left(\left(1 + \left(\frac{x}{c_1} \right)^{b_1} \right)^{-g_1}, \left(1 + \left(\frac{x}{c_2} \right)^{b_2} \right)^{-g_2} \right) \quad (6.2)$$

where $b_1, c_1 > 0$ and $g_1 > 0$ are the parameters controlling the cut-in segment, while $b_2, c_2 > 0$ and $g_2 > 0$ model the cut-out section; $\min(\cdot)$ gives the minimum value.

The parameters in 6P-PC are estimated using a genetic algorithm (GA) [217]. To improve the convergence speed, the prior knowledge from properties of 5PLF will be utilised as much as possible [83], which could set the GA initial values and ranges of search space. To illustrate the process, filtered recordings of WT1 in Scotland Wind Farm Dataset, i.e., the normal data in Fig. 6.4, are used as an example of WT power curve modelling.

The GA [217] parameters are set as follows: “population size” is 100, “mutation probability” is 0.1, “elitism ratio” is 0.01, “crossover probability” is 0.5, “parents portion” is 0.3, “crossover type” is uniform, “maximum number of iteration” is 25,000,

and “maximum number of iteration without improvement” is 2,500. The constraints of the parameters in Equation (6.2) are assigned as: $-20 < b_1 < 0$, $10 < c_1 < 20$, $0 < g_1 < 1$, $0 < b_2 < 60$, $20 < c_2 < 30$, $0 < g_2 < 2$, according to experience on the first few trials and also the 5PLF properties [83]. The values of these parameters are initialised as the averages of the boundaries of their constraints. The objective is to minimise the RMSE between 6P-PC outputs and the mean values of power outputs in all wind speed bins in WF fully operating regime. A technique similar to focal loss [218] is used. That is, only the estimations that largely deviate from the recorded values (e.g., larger than 0.005 p.u. in each iteration) will be considered in evaluating RMSE. The final results of 6P-PC parameters estimations on filtered measurements of WT1 in Scotland Wind Farm Dataset are: $b_1 = -8.822$, $c_1 = 12.644$, $g_1 = 0.338$, $b_2 = 57.523$, $c_2 = 21.983$, $g_2 = 0.504$. The fitting results are visualised in Fig. 6.6, where RMSE and MAE between “scatters PC” (obtained after removing outliers), i.e., the power curve calculated using method-of-bins for all scatters data, and “6P-PC” are 0.021 p.u. and 0.014 p.u., respectively.

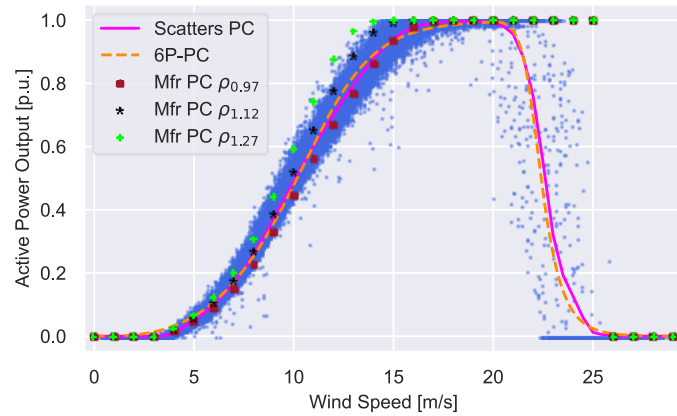


Fig. 6.6. 6P-PC fitting results on normal data of WT1, Scotland Wind Farm Dataset.

6.5 Wind Farm Operating Regimes and Equivalent Wind Farm Model

As a WF is an aggregation of individual WTs, it is therefore essential to understand the operating conditions of WTs while analysing the WF, especially when the WF WS and P_{out} are derived from the WT data [142]. In this case, the WF-level outliers are determined by the WT data. According to the analysis in Section 6.3, since CAT-II

and CAT-IV WT-level outliers are discarded at WF-level, they are also WF-level outliers.

This section will only focus on the case where WF data are derived from the individual WTs in the same WF (i.e., Scotland Wind Farm Dataset) [142] and it will utilise the outlier detection results in Section 6.3 to analyse WF operating regimes, which is a “bottom-to-top” approach. The opposite and more challenging case, “top-to-bottom” analysis, will be presented in Section 6.6, where the Zelengrad Wind Farm Dataset is used. In this section, 6P-PC model proposed from Section 6.4 will be used to fit measurements of WF when it is in fully operating regime and formulate the equivalent wind farm model. An important relationship between WT and WF will be proven, which shows that the mean power output of a partly operating WF can be represented as a linear scaling from the power output of a fully operating WF (plus all existing WT curtailments).

6.5.1 WF Operating Regime

A WT may be normally operating, in curtailment, or in outage, and these operating conditions may show various patterns at WF-level. CAT-I outliers show WT curtailment, while CAT-III outliers happen when the WT is in outage.

There are multiple WTs in a WF and not all of them always have the same operating condition simultaneously. Based on the combinations of different operating conditions of WTs, the WF can be classified into various operating regimes, including one fully operational regime and many other partly operational regimes. For a WF, WF P_{out} can be written as the sum of the synchronous power outputs of all WTs in the WF:

$$\begin{aligned} P_{WF} &= \sum_{i=1}^A P_{WT_i} + \sum_{j=1}^B \min(P_{WT_j}, P_{CTL_j}) + 0 \times \sum_{k=1}^C P_{WT_k} \\ &= \sum_{i=1}^A P_{WT_i} + \sum_{j=1}^B \min(P_{WT_j}, P_{CTL_j}) \end{aligned} \quad (6.3)$$

where: P_{WF} is the total P_{out} of the whole WF, and P_{WT} is P_{out} of an individual WT; A is the number of WTs operating normally, B represents the number of WTs operating under curtailment with maximum power output limited to P_{CTL} , and C indicates the number of WTs in outage. Different combinations of (A, B, C) result in different

operating regimes of the WF. The WS of the WF is calculated following the averaging method in [142], but using only WTs normally generating or in curtailment.

The classification results of Scotland Wind Farm Dataset WF operating regimes are shown in Fig. 6.7, while the percentages of operating regimes are listed in Table 6.6. The notations of the combinations in the form of (A, B, C) represent different WF operating regimes, sharing the identical meanings as in Equation (6.3), and “others” include all operating regimes that occur less than 0.01 % time. The results show that this WF is fully operational (all WTs in normal service and no curtailment, i.e., (A, B, C) = (6, 0, 0)) for around 25.87 % of the time. It suggests that in WF-level studies, the WT-level CAT-I and CAT-III outliers should be correctly represented, as otherwise much larger than actual WF power output will be estimated.

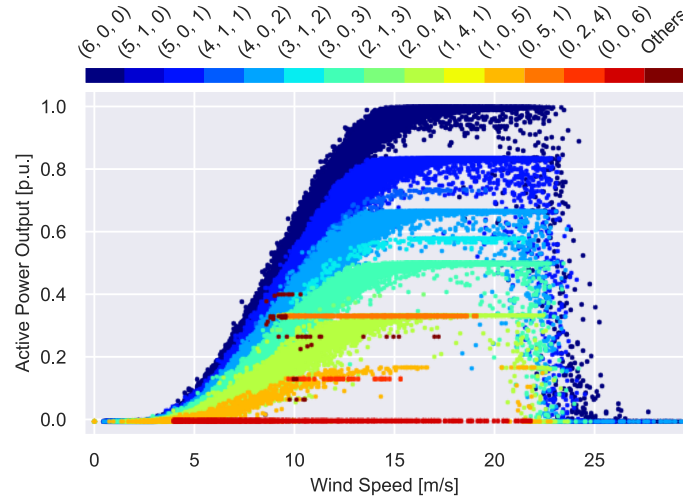


Fig. 6.7. WF operating regime classification results for Scotland Wind Farm Dataset, indicated by different colours and (A, B, C) combinations at the top.

Table 6.6 Percentages of WF Operating Regimes

(A, B, C)	%	(A, B, C)	%	(A, B, C)	%
(6, 0, 0)	25.87	(5, 1, 0)	0.02	(5, 0, 1)	28.74
(4, 1, 1)	0.17	(4, 0, 2)	18.26	(3, 1, 2)	0.17
(3, 0, 3)	11.28	(2, 1, 3)	0.02	(2, 0, 4)	7.84
(1, 4, 1)	0.01	(1, 0, 5)	2.60	(0, 5, 1)	0.20
(0, 2, 4)	0.02	(0, 0, 6)	4.77	Others ¹	0.03

6.5.2 Equivalent Wind Farm Model

For a WF installed with N WTs of the same type, Equation (6.3) shows that total P_{out} of the WF is the arithmetic sum of P_{out} of all normally operating WTs, plus a bias,

which is the generation produced by curtailed WTs. This is denoted as “linear scaling (plus bias) relationship” and can be formulated as Equation (6.4), which is effectively an equivalent wind farm model:

$$P_{WF} = \frac{N_0}{N} P_{WF|FULL} + CLT \quad (6.4)$$

where: N_0 is the number of normally operational WTs, $P_{WF|FULL}$ is the total WF P_{out} if it is fully operating, which can be estimated using 6P-PC and CLT represents the sum of all WTs under curtailment.

For Scotland Wind Farm Dataset, as all WF operating regimes have already been classified in Fig. 6.7, it is straightforward to test the as “linear scaling (plus bias) relationship” by firstly estimating the fully operational regime 6P-PC and applying the “linear scaling (plus bias) relationship”, and then comparing the results with the actual recordings. The results of this equivalent wind farm model are shown in Fig. 6.8. In Fig. 6.8, “filtered data” are the WF data without WF-level outliers, which are identical to the scatters in Fig. 6.7. “Regime PC” represents power curves in all individual WF operating regimes by the method of bins [12]. “6P-PC” is the fitted power curve for the WF fully operating regime, whose parameter values are: $b_1 = -8.302$, $c_1 = 12.188$, $g_1 = 0.326$, $b_2 = 39.234$, $c_2 = 23.180$, $g_2 = 1.082$. “LS 6P-PC” is the linearly scaled 6P-PC. And “LS 6P-PC plus CLT” equals the sum of “LS 6P-PC” and all curtailments. “LS 6P-PC plus CLT” means there is at least one WT being curtailed. Those plots do not occur at all wind speeds, because the WT-level curtailment only takes a small percentage of data, and they only occur at certain wind speeds. Together, “6P-PC”, “LS 6P-PC” and “LS 6P-PC plus CLT” compose the equivalent WF model. This model can relatively accurately estimate the WF P_{out} , but the errors increase dramatically near the cut-out WS, which may be due to the lack of cut-out region data in the WF fully operating regime, causing it difficult to generalise the fitted “6P-PC” for other partly operating regimes through linear scaling. Nevertheless, the results in Table 6.7 still show great improvements compared to using Mfr PC (using the average air density, 1.12 kg/m^3) and the original power curve by the method of bins (“scatters PC” in Fig. 6.2(b)). The errors are calculated based on all WF data without WF-level outliers in Scotland Wind Farm Dataset.

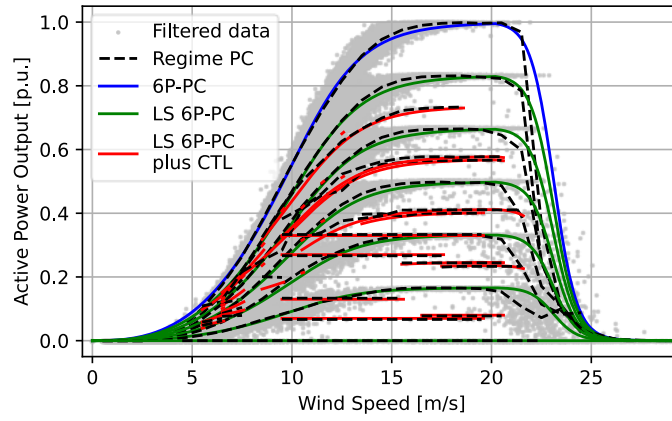


Fig. 6.8. Equivalent wind farm model for Scotland Wind Farm Dataset.

Table 6.7 Comparisons of Power Output Estimations for Scotland Wind Farm Dataset (Derived from WTs)

	RMSE [p.u.]	MAE [p.u.]
Mfr PC $\rho_{1,12}$	0.172	0.093
Original Power Curve by Method of Bins	0.116	0.073
Proposed Equivalent WF Model	0.031	0.017

Fig. 6.9 illustrates the errors respecting all different WF operating regimes. Both RMSE and MAE are calculated between the results of the equivalent wind farm model and the mean values of actual P_{out} recordings in WS bins in each WF operating regime. In all WF operating regimes, the as “linear scaling (plus bias) relationship” gives small errors. The highest RMSE and MAE are around 0.038 p.u. and 0.030 p.u., respectively. The small errors can directly prove the validity of the “linear scaling (plus bias) relationship”.

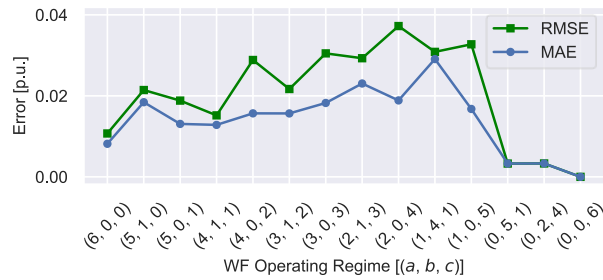


Fig. 6.9. Error evaluation for “linear scaling (plus bias) relationship”.

6.6 Single Sensor WF Measurements

Most of the discussions so far are limited to cases relying on the available individual WT measurements in the same WF. The equivalent wind farm model, Equation (6.4), is a high-level generalisation that does not need recordings of WTs once fitted.

However, to fit Equation (6.4), WF fully operating regime must be identified. As discussed, the method of classifying the WF operating regimes in Section 6.4 is a “bottom-to-top” approach, but in real-world applications, it is common that individual WT’s measurements are not available, thus a “top-to-bottom” approach is desired. This section will propose a recursive algorithm trying to solve this problem using only WF WS and P_{out} data and Mfr PC. This algorithm will simultaneously estimate the equivalent WF model parameters and WF operating regimes. This section uses Zelengrad wind farm dataset as a case study, which is an authentic single sensor measured WF dataset.

In this case, WF-level outliers are identified by applying boundary rules and sliding window method directly on WF WS and P_{out} data. Although it cannot provide accurate insights like the approach that starts from analysing individual WT’s, it is still effective in detecting some obvious true outliers, which may distort the WF modelling (e.g., outliers denoted as “C” in Fig. 6.2(c)).

The WF operating regime is difficult to determine without measurements of WT’s, or logs of WT availability, because the combination of (A, B, C) in Equation (6.3) is unknown. However, based on the “linear scaling (plus bias) relationship”, the maximum WF power output in any certain operating regime can still be estimated. For example, the horizontal constant P_{out} bands in Fig. 6.2(c) happen when the operating WT’s are producing either rated or curtailed power outputs. Therefore, the WF operating regimes of these horizontal P_{out} ranges can be estimated by comparing them with linear scaled Mfr PCs, and the closest linear scaled Mfr PCs indicate the corresponding WF operating regimes.

Similarly, in terms of other WS regions (e.g., the nonlinear parts near the cut-in and cut-out WS), the WF operating regime can be approximated by calculating and assessing the distances between the measurements and a collection of linear scaled PCs, which may be Mfr PC, 6P-PC or any other function mapping WS to P_{out} . This is a distance-based approach, and it is also a hard clustering method. It is not perfect and most accurate approach in classifying WF operating regimes, because different WF operating regimes may intersect with each other due to power output variability. However, it is still important in providing the inputs for the equivalent WF model, and

the following case study will also show how it results in great improvements regarding WF power output estimations, compared to simply using Mfr PC, or original “scatters PC” in Fig. 6.2(c). More specifically, the fitting of the equivalent WF model can be represented using a recursive algorithm below:

Step 1: For a WF (installed with N WT) dataset consisting of n measurements, $[\mathbf{V}, \mathbf{P}_{\text{out}}] = [[v_0, v_1, \dots, v_{n-1}]^T, [p_{\text{out};0}, p_{\text{out};1}, \dots, p_{\text{out};n-1}]^T]$, initialise a 6P-PC whose parameters equal to the averages of the constraints’ boundaries illustrated in Section 6.4.

Step 2: Apply the 6P-PC obtained in *Step 1* and use $\mathbf{V} = [v_0, v_1, \dots, v_{n-1}]^T$ as its inputs, where the calculation results are $\tilde{\mathbf{P}}_{\text{out}} = [\tilde{p}_{\text{out};0}, \tilde{p}_{\text{out};1}, \dots, \tilde{p}_{\text{out};n-1}]^T$.

Step 3: Prepare a collection of linear scaled $\tilde{\mathbf{P}}_{\text{out}}$: $\tilde{\mathbf{P}}_{\text{out}} \times 0/N, \tilde{\mathbf{P}}_{\text{out}} \times 1/N, \dots, \tilde{\mathbf{P}}_{\text{out}} \times N/N$. For each element in $[p_{\text{out};0}, p_{\text{out};1}, \dots, p_{\text{out};n-1}]^T, p_{\text{out};k}$, compare it with $\tilde{p}_{\text{out};k} \times 0/N, \tilde{p}_{\text{out};k} \times 1/N, \dots, \tilde{p}_{\text{out};k} \times N/N$, and calculate the distances: $\tilde{p}_{\text{out};k} \times 0/N - p_{\text{out};k}, \tilde{p}_{\text{out};k} \times 1/N - p_{\text{out};k}, \dots, \tilde{p}_{\text{out};k} \times N/N - p_{\text{out};k}$. Find the one that is the smallest in all positive distances, and define its scaling factor as η_k/N .

Step 4: Make corrections for estimations of η_k/N in the horizontal constant $\mathbf{P}_{\text{out}}$ bands (i.e., when all η_k WT are producing rated 1.0 p.u.). This can be achieved by repeating *Step 2* and *Step 3*, but Mfr PC instead of 6P-PC is used to calculate $\tilde{\mathbf{P}}_{\text{out}}$.

Step 5: Perform elementwise product between $[\tilde{p}_{\text{out};0}, \tilde{p}_{\text{out};1}, \dots, \tilde{p}_{\text{out};n-1}]^T$ and $[\eta_1/N, \eta_2/N, \dots, \eta_{n-1}/N]^T$, and calculate RMSE between the results and $[p_{\text{out};0}, p_{\text{out};1}, \dots, p_{\text{out};n-1}]^T$. Note that since the error evaluation is based on the whole dataset, the tail events (i.e., the low probability recordings occurring at high WS) may be masked by others. Therefore, different weights are assigned to the recordings when assessing the RMSE, which is calculated as the inverse of the number of recordings in the WS bin of v_k .

Step 6: Initialise a GA whose settings are identical to the GA described in Section 6.4, except that the loss function is calculated as above. Repeat *Step 1* to *Step 5*, and use the GA to find the optimised parameters for 6P-PC defined in *Step 1*. When it finishes, repeat *Step 2* to *Step 4* using the currently optimised 6P-PC to find the WF operating regimes.

The results of this algorithm on the Zelengrad Wind Farm Dataset are shown in Fig. 6.10, where the notations and labels have identical meanings as in Fig. 6.7 and Fig. 6.8. The comparisons of the proposed model, Mfr PC (using average air density, 1.12 kg/m^3) and “scatters PC” in Fig. 6.2(c) are shown in Table 6.8, which are based on all recordings in Zelengrad Wind Farm Dataset excluding the outliers. The estimated WF operating regimes have followed the logical trends as the results of “bottom-to-top” approach in Fig. 6.8 and it is noticeable that there are less recordings near the cut-out WS for a larger number of normally operating WTs. It can be also seen that some WTs in this WF may even continue to generate rated power output above the cut-out WS. Nevertheless, as shown in Table 6.8, the proposed model still demonstrates great improvements compared to using Mfr PC and original PC by method of bins.

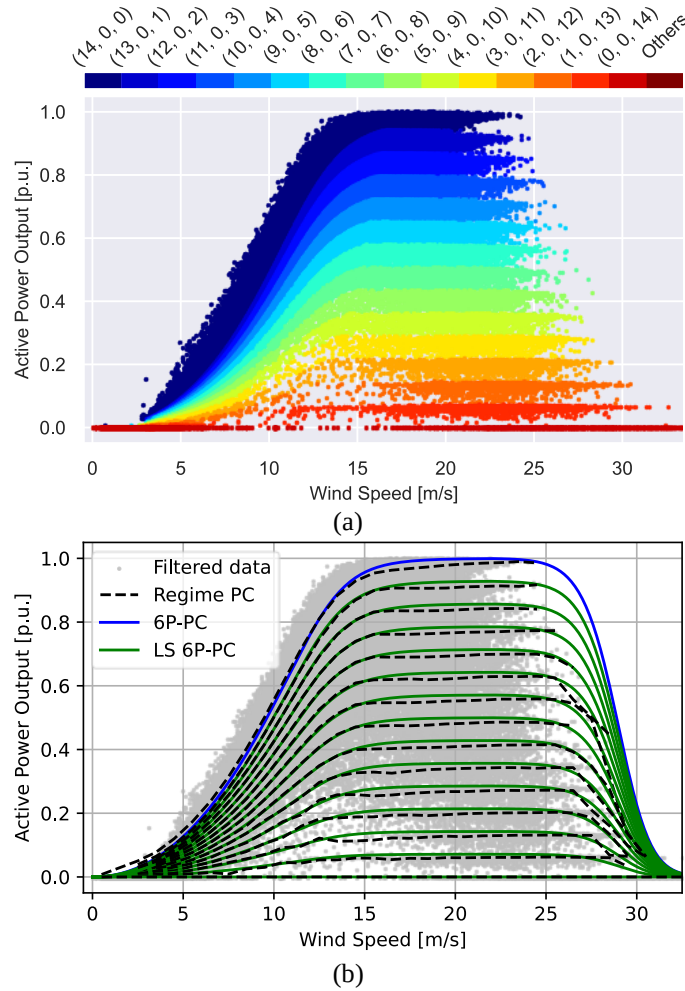


Fig. 6.10. Results of equivalent wind farm model fitting algorithm for single sensor WF measurements (Zelengrad Wind Farm Dataset). (a) WF operating regime classification results indicated by different colours. (b) Equivalent wind farm model.

Table 6.8 Comparisons of Power Output Estimations for Zelengrad Wind Farm Dataset (Single Sensor Measurements)

	RMSE [p.u.]	MAE [p.u.]
Mfr PC $\rho_{1,12}$	0.148	0.076
Original Power Curve by Method of Bins	0.082	0.048
Proposed Equivalent WF Model	0.022	0.013

6.7 Discussions and Conclusions

This chapter extends the analysis of wind turbine and wind farm measurements outliers from Chapter 5. The proposed WT-level outlier detection framework provides a set of comprehensive methods in identifying the outliers from both statistical and physical points of view. In particular, a novel simulation-based approach is proposed to identify false outliers, which should be part of the normal data used for modelling. The results show that the simulation on average rejects 12.94 % of data as scattered outliers (CAT-IV outliers), while the considered case studies have shown that different types of outliers may need to be treated differently in WT-level and WF-level studies.

The presented new 6-parameter power curve model can correctly represent performance of modelled WT or WF for all wind speed regions, including the nonlinear parts near the cut-in and, particularly, cut-out wind speeds. The proposed equivalent WF model considers multiple WF operating regimes, i.e., specific numbers of WTs in normally operating, in curtailment, or in outage states. The presented analysis demonstrate that the mean values of power outputs of the partly operational WF can be estimated as “linear scaling (plus bias)” from the average power outputs of the fully operating WF. The final results of the full equivalent WF models have demonstrated much higher accuracy when compared with manufacturer power curve and the original recordings’ power curve by method of bins.

Chapter 7

Wind Farm 2D Probabilistic Modelling and Conditional Distribution Fittings

7.1 Introduction

As discussed in Chapter 5 and Chapter 6, manufacturers' power curves are typically used if no field (measured) data are available in planning and general design phases of a wind farm. The operational power curves derived from field measurements for both WT and WF are proposed in Chapter 6. The power curves can capture the deterministic relationship between the mean wind speed and the average power output. However, in the actual measurements, the power output may vary even for the specific (e.g., measured, or known) input wind speed. The measurements are usually recorded as 10-minute average values, and the nonlinear power curve (representing nonlinear characteristics of energy conversion) and WT hysteresis control (representing one important control system) lead to P_{out} variations for the same averaged WS. The simulations of WT to identify the outliers in Chapter 6 have also proven this "one (WS) to many (P_{out})" relationship. Therefore, to correctly describe the WF measurements, a probabilistic model is preferred to the deterministic one.

Instead of using deterministic models, this chapter evaluates variations in wind energy resources and WF power generation applying probabilistic distributions and models. The Scotland Wind Farm Dataset, which is introduced in Section 6.2, is again used in this chapter for case studies and related analysis.

7.2 Probabilistic Models for Wind Speed

Wind speed is the most important explanatory variable for WT and WF P_{out} estimation, and distribution of wind speed is essential for determining the performance of wind energy-based generation system. This section considers three analytical distribution functions and compares their accuracy using available wind speed measurements fitting: 1) two-parameter Weibull distribution (2-pW); 2) three-parameter Weibull distribution (3-pW); 3) Generalised Normal Distribution (GND).

7.2.1 Two-Parameter Weibull Distribution

The two-parameter Weibull distribution is analysed first, as it is widely used for modelling wind speed distributions. The probability density function (PDF) and cumulative density function (CDF) of 2-pW distribution are defined as:

$$\text{PDF}_{2\text{-pW}}(x) = \begin{cases} \frac{\kappa}{\lambda} \left(\frac{x}{\lambda}\right)^{\kappa-1} \exp\left(-\left(\frac{x}{\lambda}\right)^{\kappa}\right) & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (7.1)$$

$$\text{CDF}_{2\text{-pW}}(x) = \begin{cases} 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^{\kappa}\right) & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (7.2)$$

where: x is the random variable, $\kappa > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. The maximum likelihood estimation (MLE) for κ and λ is given by [219]:

$$\lambda = \left(\frac{1}{n} \sum_{i=1}^n x_i^{\kappa}\right)^{1/\kappa} \quad (7.3)$$

$$\frac{1}{\kappa} - \frac{\sum_{i=1}^n x_i^{\kappa} \ln x_i}{\sum_{i=1}^n x_i^{\kappa}} + \frac{1}{n} \sum_{i=1}^n \ln x_i = 0 \quad (7.4)$$

Equation (7.4) could be solved using Newton-Raphson method introduced in [220]. It is equivalent to solving $f(\kappa) = \frac{\sum_{i=1}^n x_i^{\kappa} \ln x_i}{\sum_{i=1}^n x_i^{\kappa}} - \frac{1}{\kappa} - \frac{1}{n} \sum_{i=1}^n \ln x_i = 0$, where $\frac{df(\kappa)}{d\kappa} = \frac{(\sum_{i=1}^n x_i^{\kappa} (\ln x_i)^2)(\sum_{i=1}^n x_i^{\kappa}) - (\sum_{i=1}^n x_i^{\kappa} \ln x_i)^2}{(\sum_{i=1}^n x_i^{\kappa})^2} + \frac{1}{\kappa^2}$. According to [221], in Newton-Raphson method, the initial κ is chosen as $\kappa = \left(\frac{\bar{x}}{\text{std}(x)}\right)^{1.086}$, where $\bar{x}$ is the mean value of x_i , and the standard deviation $\text{std}(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$.

Refs [219] and [222] give estimations for κ and λ using method of moments (MOM):

$$\frac{\Gamma(1+\frac{2}{\kappa})}{\Gamma^2(1+\frac{1}{\kappa})} = \frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{\left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2} \quad (7.5)$$

$$\lambda = \frac{\frac{1}{n} \sum_{i=1}^n x_i}{\Gamma(1+\frac{1}{\kappa})} \quad (7.6)$$

Equation (7.5) can also only be solved using the numerical method (Newton-Raphson method). Similarly, it is to solve $f(\kappa) = \frac{\Gamma(1+\frac{2}{\kappa})}{\Gamma^2(1+\frac{1}{\kappa})} - \frac{\frac{1}{n}\sum_{i=1}^n x_i^2}{(\frac{1}{n}\sum_{i=1}^n x_i)^2} = 0$, and $\frac{df(\kappa)}{d\kappa} = \frac{\Gamma(1+\frac{2}{\kappa})\delta(1+\frac{2}{\kappa})(-\frac{2}{\kappa^2}) - 2\Gamma(1+\frac{1}{\kappa})\delta(1+\frac{1}{\kappa})(-\frac{1}{\kappa^2})}{\Gamma^2(1+\frac{1}{\kappa})}$, where $\delta(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is known as the digamma function (whose values are also equal to the first derivative of the logarithm of the gamma function). The chosen initial $\kappa = (\frac{\bar{x}}{\text{std}(x)})^{1.086}$ is like in MLE.

The MLE will produce smaller fitting errors than MOM, as MOM is based on the law of large numbers, which states that for the independent and identically distributed (i.i.d.) random variables, the samples' moments will converge to the population's moments as the sample size increases. Since only the first several orders of moments are used in the implemented fitting algorithm, it usually is not as accurate as MLE. However, MOM is generally easier to calculate analytically than MLE, in particular for distributions with complex forms.

7.2.2 Three-Parameter Weibull Distribution

Compared to the two-parameter Weibull distribution, the three-parameter Weibull distribution has one additional location parameter and it is more flexible and accurate in data fitting. The PDF and CDF of three-parameter Weibull distribution are defined as:

$$\text{PDF}_{3\text{-pW}}(x) = \begin{cases} \frac{\kappa}{\lambda} \left(\frac{x-\theta}{\lambda}\right)^{\kappa-1} \exp\left(-\left(\frac{x-\theta}{\lambda}\right)^{\kappa}\right) & x > \theta \\ 0 & x \leq \theta \end{cases} \quad (7.7)$$

$$\text{CDF}_{3\text{-pW}}(x) = \begin{cases} 1 - \exp\left(-\left(\frac{x-\theta}{\lambda}\right)^{\kappa}\right) & x > \theta \\ 0 & x \leq \theta \end{cases} \quad (7.8)$$

where: θ is the location parameter. The other parameters are the same as in 2-pW. If $\theta = 0$, it reduces to two-parameter Weibull distribution, and θ should be less than $\min(x)$, the minimum value of x , as otherwise, the probability is defined as zero. Ref. [223] proposes MOM to estimate the three-parameter Weibull distribution:

$$\kappa = \frac{\ln(2)}{\ln(\bar{\mu}_1 - \bar{\mu}_2) - \ln(\bar{\mu}_2 - \bar{\mu}_4)} \quad (7.9)$$

$$\theta = \frac{\bar{\mu}_1 \bar{\mu}_4 - \bar{\mu}_2^2}{\bar{\mu}_1 + \bar{\mu}_4 - 2\bar{\mu}_2} \quad (7.10)$$

$$\lambda = \frac{\bar{\mu}_1 - \theta}{\Gamma\left(1 + \frac{1}{\kappa}\right)} \quad (7.11)$$

where $\bar{\mu}_k$ is estimated by: $\bar{\mu}_k = \sum_{r=0}^{n-1} \left(1 - \frac{r}{n}\right)^k (x_{(r+1)} - x_{(r)})$, and $x_{(r)}$ is the sorted random variable x , $x_{(0)} = 0$ and $x_{(r)} \leq x_{(r+1)}$. Many trials of this 3-pW MOM fitting algorithm have been performed on synthetic data sample sets, where the samples are obtained by firstly sampling from 2-pW and then the same bias value is added to represent the effects of the location parameter. It shows that the estimated location parameter θ has a smaller error than the other two parameters κ and λ , but κ and λ can become very unreliable or even unsolvable in some conditions.

A new fitting method is proposed to avoid the mathematically nontrivial problem and invalid solutions. The idea is that if the location parameter θ is set, by giving all sample data an offset equal to $-\theta$, the three-parameter Weibull distribution could be reduced to two-parameter Weibull distribution, as the fitting of 2-pW is much easier and could be estimated using previously developed MLE or MOM techniques.

Therefore, in 3-pW fitting, the candidate estimations for κ , θ and λ are firstly calculated following Equations (7.9), (7.10) and (7.11), which are denoted as κ_0 , θ_0 and λ_0 . Then, 5,000 uniformly distributed samples of θ_{try} are drawn from the range of $[0.95 \times \theta_0, \min(x, 1.05 \times \theta_0)]$. For every θ_{try} , 2-pW MLE estimations are made for the new dataset with offset, $x - \theta_{\text{try}}$, so the corresponding κ and λ can be calculated. Finally, there will be $5,001 = 1 + 5,000$ different groups of estimations for the parameters κ , θ and λ . The group that leads to the minimum root-mean-square error between the fitted CDF and the empirical CDF of the data is selected as the final parameter estimation.

The effectiveness of the fitting methods is firstly tested on a controlled synthetic dataset before applying it to the actual measurement. Fig. 7.1 shows a comparison of 2-pW and 3-pW fittings on a synthetic dataset. The dataset has 10,000 samples, and the samples are randomly taken from a 2-pW with shape parameter and scale

parameter equal to 3 and 5, respectively. All samples are added with an offset of 1,000. It can be seen that the 3-pW provide much better fitting results than 2-pW, and that the additional location parameter can successfully capture the offset of the data samples.

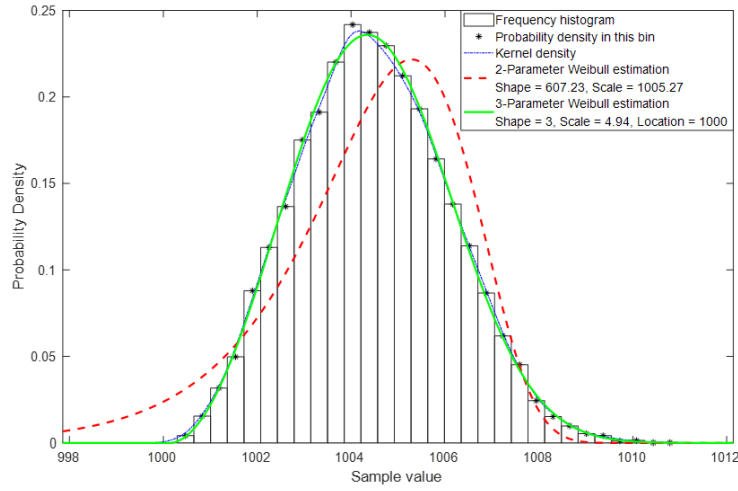


Fig. 7.1. Comparison of 2-pW and 3-pW fittings on a synthetic dataset.

7.2.3 Generalised Normal Distribution

Compared to normal distribution, the generalised normal distribution has three parameters: shape parameter β , scale parameter $\alpha > 0$, and location parameter ζ . The shape parameter introduces skewness to the probability distribution, and when $\beta = 0$, GND becomes standard normal distribution. Fig. 7.2 shows the probability density plots of GNDs with different shape parameters. The larger shape parameter (or smaller, if $\beta < 0$) will lead to more negative skewness (or positive skewness if $\beta < 0$).

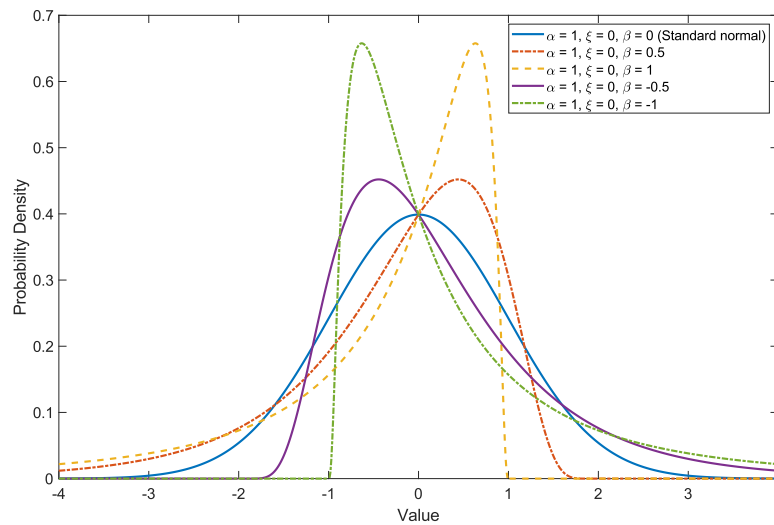


Fig. 7.2. Illustration of GND with different shape parameters.

The PDF and CDF of GND are as follows:

$$\text{PDF}_{\text{GND}}(x) = \frac{\phi(y(x))}{\alpha - \beta(x - \xi)} \quad (7.12)$$

$$\text{CDF}_{\text{GND}}(x) = \Phi(y(x)) \quad (7.13)$$

$$y(x) = \begin{cases} -\frac{1}{\beta} \log \left[1 - \frac{\beta(x - \xi)}{\alpha} \right] & \beta \neq 0 \\ \frac{\beta(x - \xi)}{\alpha} & \beta = 0 \end{cases} \quad (7.14)$$

where: x is the random variable, ξ is the location parameter, α is the scale parameter, and β is the shape parameter. $\phi(\cdot)$ standard normal PDF, and $\Phi(\cdot)$ is the standard normal CDF. The mean, median and variance of GND can be calculated as:

$$\text{mean}_{\text{GND}} = \xi - \frac{\alpha}{\beta} \left(\exp\left(\frac{\beta^2}{2}\right) - 1 \right) \quad (7.15)$$

$$\text{median}_{\text{GND}} = \xi \quad (7.16)$$

$$\text{variance}_{\text{GND}} = \frac{\alpha^2}{\beta^2} \exp(\beta^2) (\exp(\beta^2) - 1) \quad (7.17)$$

Therefore, the parameters of GND can be estimated by solving the following equations:

$$\xi = \text{median}_{\text{GND}} \quad (7.18)$$

$$\frac{\exp(2\beta^2) - \exp(\beta^2)}{(\exp(\frac{\beta^2}{2}) - 1)^2} = \frac{\text{variance}_{\text{GND}}}{(\text{median}_{\text{GND}} - \text{mean}_{\text{GND}})^2} \quad (7.19)$$

$$\alpha = \beta (\text{median}_{\text{GND}} - \text{mean}_{\text{GND}}) / \left(\exp\left(\frac{\beta^2}{2}\right) - 1 \right) \quad (7.20)$$

7.2.4 Wind Speed Fitting Results

Four goodness-of-fit indices are used to evaluate the accuracy of the three analytical function fittings on the available wind speed measurements. The first two are the commonly used ones, RMSE and MAE, as reviewed in Section 2.3.1, and they evaluate the error between the fitted CDF and empirical CDF. Maximum absolute error

among all bins (MAEAAB) and weighted mean absolute error (WMAE) are also proposed and applied in the comparison, which are defined as:

$$\text{MAEAAB} = \max|\hat{P}_j - P_j|, \forall j \in \{1, \dots, N\} \quad (7.21)$$

$$\text{WMAE} = \sum_{j=1}^N \hat{P}_j \times |\hat{P}_j - P_j| \quad (7.22)$$

where: $\hat{P}_j$ and P_j are the measured and fitted probability values for the j -th wind speed bin with a bin width of 1 m/s [12], respectively. The MAEAAB describes the largest estimation error among all WS bins' probability estimations. The WMAE is similar to MAE, but it does not give the same weight to all wind speed bins. The weights are based on the probability of all bins, thus it is more sensitive to errors in the bins with higher frequency.

The estimated parameters for 2-pW, 3-pW and GND, as well as the calculated error metrics, are presented in Table 7.1. Fig. 7.3 and Fig. 7.4 depict the comparisons of PDFs and CDFs of fitted distributions for wind speed.

Table 7.1 Distribution Fitting Results for Wind Speed Measurements

Model	"Best Fit"		Goodness-of-fit indices (evaluated errors)			
	Parameter	Values	RMSE	MAEAAB	MAE	WMAE
2-pW	Shape	2.081	0.00289	0.01043	0.00155	0.00272
	Scale	9.964				
3-pW	Shape	2.053	0.00285	0.00990	0.00151	0.00272
	Scale	9.838				
	Location	0.113				
GND	Shape	-0.267	0.01166	0.01470	0.00283	0.00499
	Scale	4.283				
	Location	8.247				

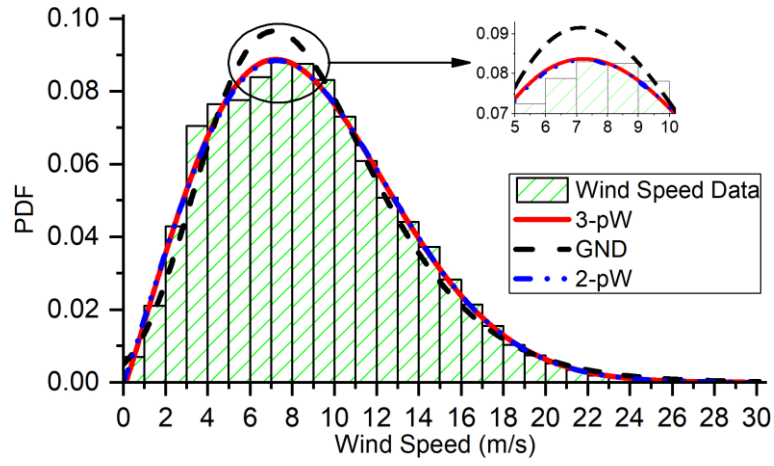


Fig. 7.3. Comparison of PDFs of fitted distributions for wind speed.

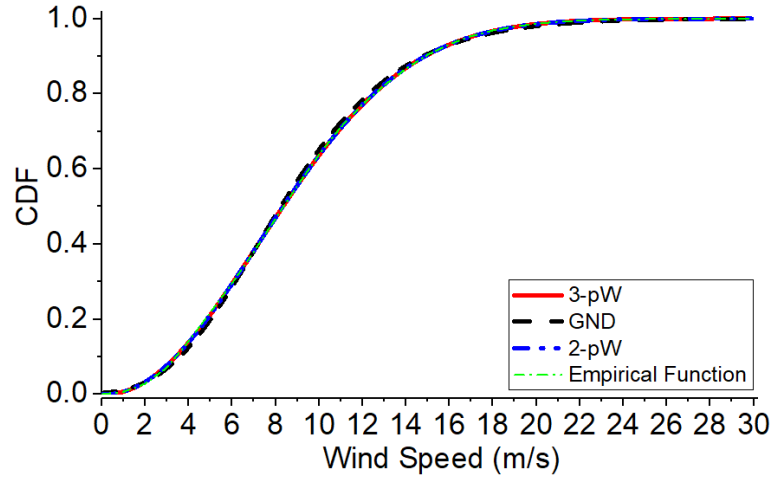


Fig. 7.4. Comparison of CDFs of fitted distributions for wind speed.

The results show that the recorded wind speed data distribution can be more accurately represented with both 2p-W and 3p-W distributions than GND. The error with 2-pW distribution is slightly higher than for 3-pW distribution, indicating that a 3p-W distribution can model the empirical wind speed distribution more accurately.

7.3 Probabilistic Models for Wind Direction

The “method of bins” [12] is firstly applied to analyse the conditional distribution wind direction (and also power output in the next section) regarding wind speed, where the dataset is partitioned according to the bins of WS. Each wind speed bin is analysed independently, and WD data in the WS bin can be fitted using analytical distributions. For example, to model the WD distribution when the WS is between 11 m/s and 12

m/s, all measurements with WS in the range of 11 m/s to 12 m/s are extracted, and the synchronous WD data are therefore conditioned on WS.

7.3.1 Von Mises Distribution and Mixture of Distributions

Von-Mises distribution (VMD) is well-known for modelling the circular random variables. The PDF and CDF of a single von Mises distribution are given as:

$$\text{PDF}_{\text{VMD}}(x) = \frac{1}{2\pi I_0(\tau)} \exp(\tau \cos(x - \rho)) \quad (7.23)$$

$$\text{CDF}_{\text{VMD}}(x) = \int_{x_l}^{x_u} \text{PDF}_{\text{VMD}}(x) dx \quad (7.24)$$

where: x is the circular random variable and $0 \leq x \leq 2\pi$. ρ is the location parameter in the range of 0 to 2π , also equal to the mean, median and mode of VMD; τ is the concentration parameter, and $1/\tau$ is similar to variance; $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero:

$$I_0(\tau) = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} \exp(\tau \cos(x)) dx = \sum_{\tau=0}^{\infty} \frac{1}{(\tau!)^2} \left(\frac{\tau}{2}\right)^{2\tau} \quad (7.25)$$

Wind direction data are usually multimodal, i.e., several peaks of the recorded WD distribution exist for each wind speed bin. Therefore, using a single analytical function is not suitable for modelling WD distributions, since it is only suitable for modelling a single modal. The mixture of distribution is considered to solve this problem. The PDF and CDF of a mixture distribution are described as:

$$\text{PDF}_{\text{Mix}}(x) = \sum_{j=1}^k \omega_j * \text{PDF}_j(x) \quad (7.26)$$

$$\text{CDF}_{\text{Mix}}(x) = \sum_{j=1}^k \omega_j * \text{CDF}_j(x) \quad (7.27)$$

where: ω_j is the weight of the j -th member and is between 0 and 1, and $\sum_{j=1}^k \omega_j = 1$; PDF_j and CDF_j are the j -th member PDF and CDF. The choice of single PDF and CDF is flexible and can be Weibull distribution, generalised normal distribution, or von-Mises distribution. The PDF_{Mix} and CDF_{Mix} are PDF and CDF of the mixture distribution. The thesis analyses three different mixture distributions: 1) mixture of 3p-W (MWB); 2) mixture of GND (MGND); 3) mixture of VMD (MVM).

For each mixture distribution, the values of weighting factors and parameters for its single member distribution in the mixture model are estimated by solving the following constrained optimisation problem:

$$obj. \min (f(\mathbf{\Omega}, \mathbf{v})) \quad (7.28)$$

$$s. t \ 0 \leq \omega_j \leq 1, \text{ for } j = 1, 2, \dots, k \quad (7.29)$$

$$\sum_{j=1}^k \omega_j = 1 \quad (7.30)$$

where: $\mathbf{\Omega} = [\omega_1, \omega_2, \dots, \omega_k]^T$ and $\mathbf{v}$ are the vectors of weight and parameters for individual distributions. The objective function is the minimisation of the RMSE between the estimated CDF by the mixture distribution and the empirical CDF.

7.3.2 Wind Direction Fitting Results

The MWB, MGND and MVM are compared in the fitting of available wind direction data. The above optimisation problem is solved by particle swarm optimisation (PSO) approach [224], which will try to find the minimum number of functions in the mixture model for which $RMSE \leq 0.01$. If RMSE is still greater than 0.01 when a mixture of maximum of five functions is used, PSO algorithm will stop, in order to prevent the use of too complex models. Four examples of wind rose plots and wind direction fitting results are shown for four different WS bins: 1) 6 – 7 m/s; 2) 11 – 12 m/s; 3) 17 – 18 m/s; 4) 24 – 25 m/s. The wind rose plots are shown in Fig. 7.5, and the WD fittings for the same WS bins are depicted in Fig. 7.6.

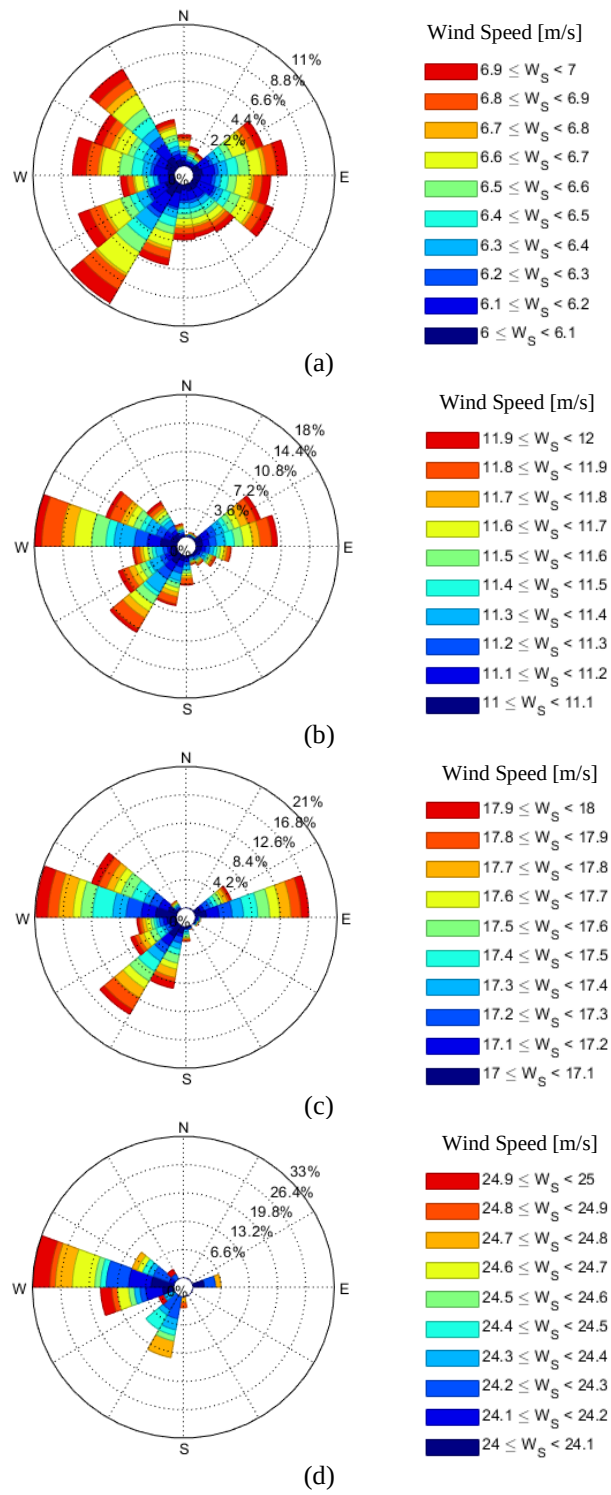


Fig. 7.5. Wind rose plots (N = 0 degree, E = 90 degrees, S = 180 degrees, and W = 270 degrees) for wind direction in different wind speed bins: a) 6 – 7 m/s; b) 11 – 12 m/s; c) 17 – 18 m/s; d) 24 – 25 m/s.

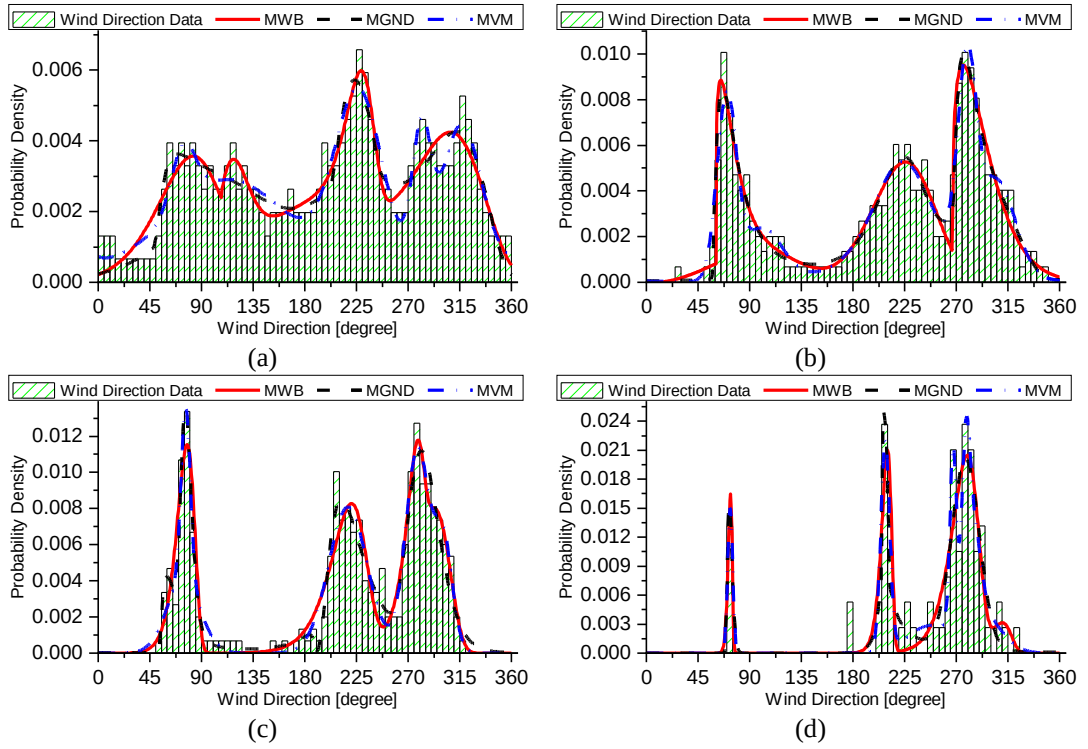


Fig. 7.6. Fitting wind direction in different wind speed bins: a) 6 – 7 m/s; b) 11 – 12 m/s; c) 17 – 18 m/s; d) 24 – 25 m/s.

It can be seen that the WD distributions are different in every WS bin. The mixture distribution can capture all modals, and overall the MVM is the closest to the distribution of the measurements. Fig. 7.7 presents the errors in all wind speed bins.

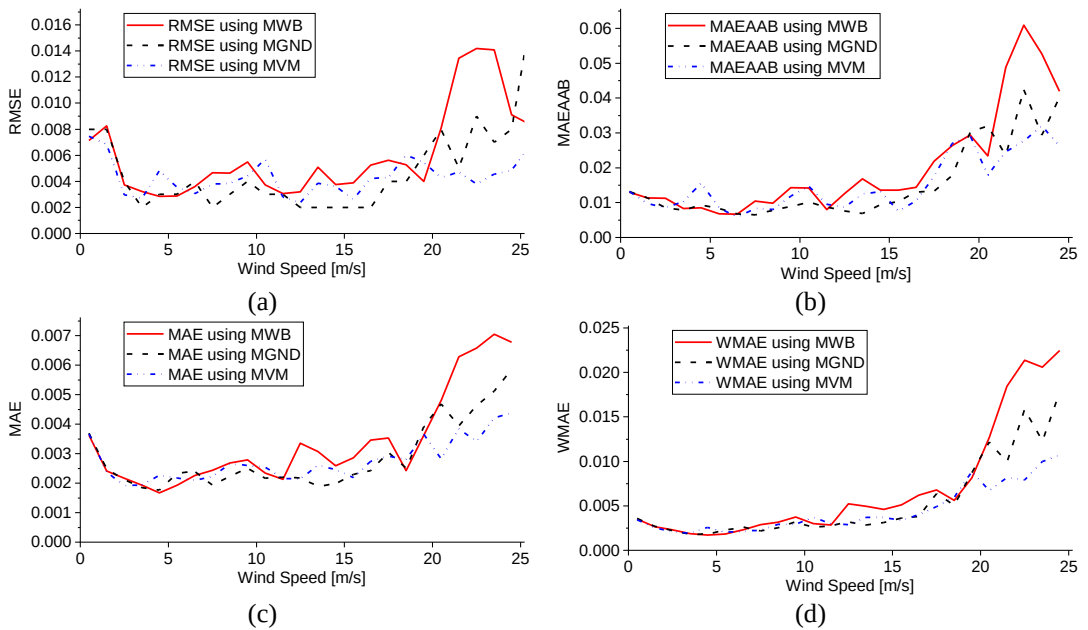


Fig. 7.7. Wind direction fitting error metrics: a) RMSE; b) MAEAAB; c) MAE; d) WMAE.

The errors tend to increase at higher wind speed ($WS > 20$ m/s) for MWB, MGND and MVM, which may be caused by the decrease of the number of available recordings.

7.4 Probabilistic Models for Power Output

Due to a similar issue of possible multimodal distributions, and also to further improve the fitting accuracy, the variation in WF power output is fitted and compared by two mixture distributions: MWB and MGND. MVM is not used here as it is primarily intended for modelling circular variables. The MWB and MGND are fitted using the same PSO algorithm as in WD fitting, and the minimum number of functions in the mixture model is used as long as the RMSE between the estimated CDF and the empirical CDF is below 0.01.

7.4.1 Power Output Fitting Results

Like in WD fitting, four examples of power output fitting results are shown for four different WS bins in Fig. 7.8: 1) 6 – 7 m/s; 2) 11 – 12 m/s; 3) 17 – 18 m/s; 4) 24 – 25 m/s. The mean values of both MWB and MGND are also plotted, while the blue dash-dot lines represent the power output values given by the manufacturer [87].

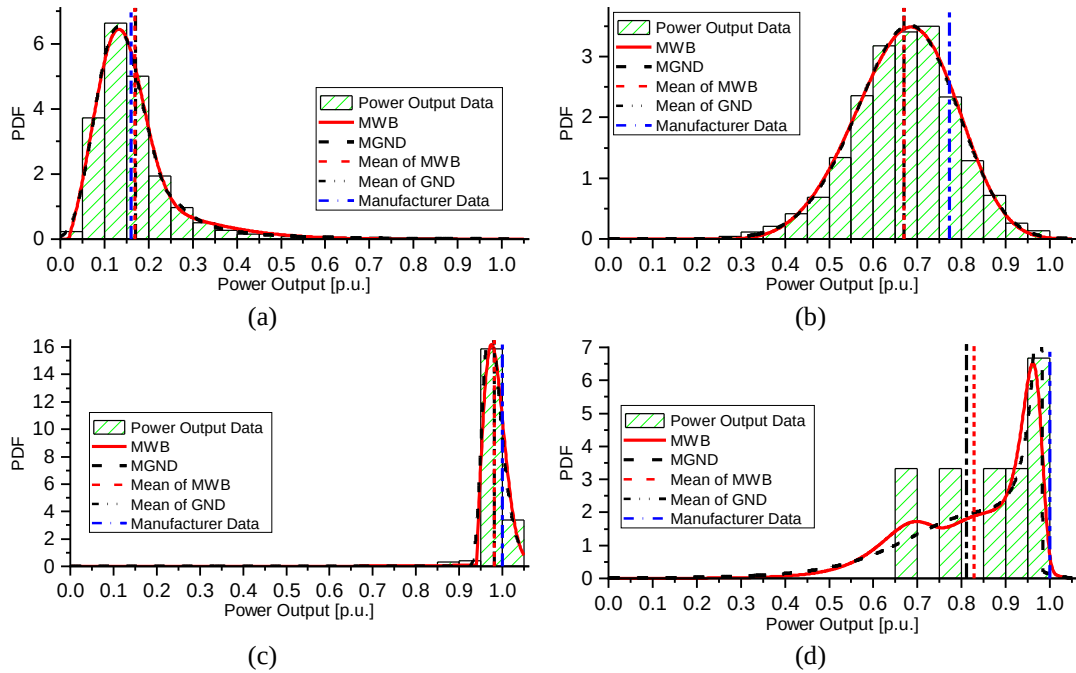


Fig. 7.8. Fitting power output in different wind speed bins: a) 6 – 7 m/s; b) 11 – 12 m/s; c) 17 – 18 m/s; d) 24 – 25 m/s.

In all four wind speed bins, field measurements of power outputs are not deterministic values, as for example provided by the manufacturer. Both MWB and MGND can accurately fit the power output data, and their mean values are different from the manufacturer power outputs. The error metrics of power output fittings for all WS bins are plotted in Fig. 7.9. The results show that, overall, MGND is more competitive than MWB, since it generally has smaller errors, particularly for some higher wind speeds.

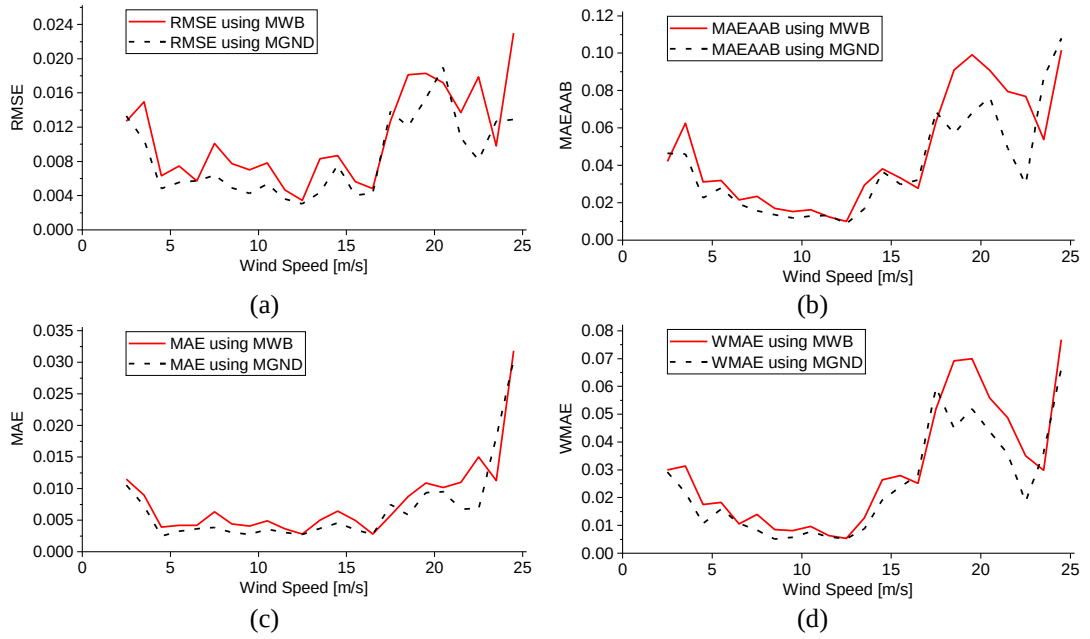


Fig. 7.9. Power output fitting error metrics: a) RMSE; b) MAEAAB; c) MAE; d) WAE.

7.5 Discussions and Conclusions

This chapter firstly analyses three analytical distributions and their parameter estimation methods. Three-parameter Weibull distribution is more precise in wind speed data fitting, but its parameter estimation is more difficult than the widely used two-parameter Weibull distribution. As the name indicates, the generalised normal distribution is a general form of normal distribution with higher flexibility, due to the additional shape parameter. Von-Mises distribution can be used to model the circular data like wind direction, but single von-Mises distribution cannot handle multimodal data.

Mixture distributions are also investigated, including MWB, MGND and MVM. They are more accurate and flexible than using only a single distribution, and they can

successfully capture multiple peaks in the data distribution. Particle Swarm Optimisation is used to find the weights of individual distributions and the parameter values of the mixture distribution.

The proposed fitting method is tested using three-year historically recorded data from a UK wind farm. The wind direction and power output data are conditioned on the wind speed bins using the method of bins, and the measurements in all bins are fitted using the mixture distributions. Unlike the deterministic manufacturer power outputs, the measured power outputs may vary significantly in the same wind speed bin, and the proposed mixture distribution (particularly MGND) can accurately capture the variations. The main reason for formulating analytical best-fit distributions is that they effectively replace a very large number of measured data in a simple PDF and CDF forms and can be then used for the analysis and comparison with other probabilistic models as the raw baseline data which are (very) close to actual empirical distributions.

Chapter 8

3D Vine Copula Model for Evaluation of Wind Turbine Power Outputs with and without Uncertainties in Input Wind Speed and Wind Direction Data

8.1 Introduction

As reviewed in Section 2.5.2, the most common approach in assessing wind energy is to correlate wind turbine (WT) power output with only one predictor variable: wind speed (WS), while wind direction (WD) is usually neglected although it also has an impact on the power output. This chapter analyses importance of including wind direction as an additional explanatory variable with the wind speed for evaluating uncertainty in wind turbine power output.

The conversion of input wind energy into electrical power by a wind turbine is approached as an inherently stochastic process, whose statistical properties can be described and modelled using a suitable set of probabilistic input and output variables. The probabilistic framework in Chapter 7 can provide conditional distribution estimations for $P(P_{\text{out}}|\text{WS})$ and $P(\text{WD}|\text{WS})$, but it has limitations in estimating $P(P_{\text{out}}|\text{WS}, \text{WD})$, because it will lead to fewer samples if more conditions are continuously added. For example, assuming the WD range is defined to be within 0 to 360 degrees, $P(P_{\text{out}}|\text{WS})$ is equivalent to $\int_0^{360} P(P_{\text{out}}|\text{WS}, \theta) d\theta = P(P_{\text{out}}|\text{WS}, 0 \leq \text{WD} \leq 360)$, where the samples in each WS bin are taken regardless of the value of WD. But $P(P_{\text{out}}|\text{WS}, \text{WD})$ means in addition to the conditioning of WS, the WD value (only a subset of the 0 – 360 degrees range, e.g., 0 – 90 degrees) also limits the sample selection, and thus a smaller number of samples are available.

In Chapter 7, its conditions and data partition are achieved by method of bins, which lacks fitting and is similar to an empirical solution. In contrast, Copula method is a more systematic approach and can elegantly fit and capture all correlations and produce estimation for $P(P_{\text{out}}|\text{WS}, \text{WD})$.

This chapter provides a more comprehensive evaluation of uncertainties in WT power outputs, as it takes into account uncertainties in the input wind energy, i.e., uncertainties in the input WS and WD values. For that purpose, 2D and 3D Copula based models are combined with 2D and 3D Markov chain models, and the Scotland Wind Farm Dataset introduced in Section 6.2 is again used in this chapter. The first two years of recorded data are used for training and the third year is used for testing.

For a given averaging window (e.g., 10 minutes), the uncertainty in WT's power output is defined in this chapter as the 5th-95th percentile range of variations of P_{out} values for two general cases: a) for specific or known input WS and WD values (i.e., WS and WD are without uncertainties), or b) for estimated or forecasted input WS and WD values (i.e., WS and WD are both with associated uncertainties). This definition allows for systematic analysis and direct comparison of different causes of uncertainties in WT power output in Case a) and Case b), as well as for a consistent evaluation of uncertainties in both wind-based electricity generation systems hindcasting and forecasting studies, which basically differ only in the way how the input WS and WD data are obtained and used in the developed probabilistic WT model.

The obtained results show that the inclusion of WD will provide noticeable improvement for models with no uncertainties in input WS and WD data, while in the case of forecasted WS and WD data with uncertainties, WS is a much stronger contributor to the total WT P_{out} uncertainty than WD. The conclusion applies to and can be scalable to other WFs, being even stronger for WFs with specific layouts. The case studies in this chapter are performed on a WT, and the same model is extended and scaled to model two WFs (with 16 and 20 wind turbines, respectively) in Chapter 9, which shows the P_{out} estimation from three-dimensional inputs (wind speed, wind direction and air density) has even higher accuracy and smaller uncertainties.

8.2 WT Power Curve Models, Importance of Including WD and Empirical P_{out} Uncertainty

Fig. 8.1(a) compares manufacturer power curve model and average operational power curve (AO-PC) models for one WT (WT3) in Scotland Wind Farm Dataset (the same

as in Chapter 6), whose layout is shown in Fig. 8.1(b) and Fig. 8.1(c). The AO-PC plots are obtained using synchronous WS, WD and P_{out} measurements for winter seasons over the period of three years. As the WF layout and AO-PC results in Fig. 8.1(a) indicate, WT3 has the highest average P_{out} for WDs between 0-90°, corresponding to the minimum turbulence and wake effect conditions. It is also clear that Mfr-PC does not correctly represent the actual operating performance of WT3: the maximum difference from Mfr PC in the WS region between cut-in and rated P_{out} speed is around 10% for WS=13.25 m/s. (Note: WD is measured as a clockwise relative angle between the zero degree North direction, as indicated in Fig. 8.1(b)).

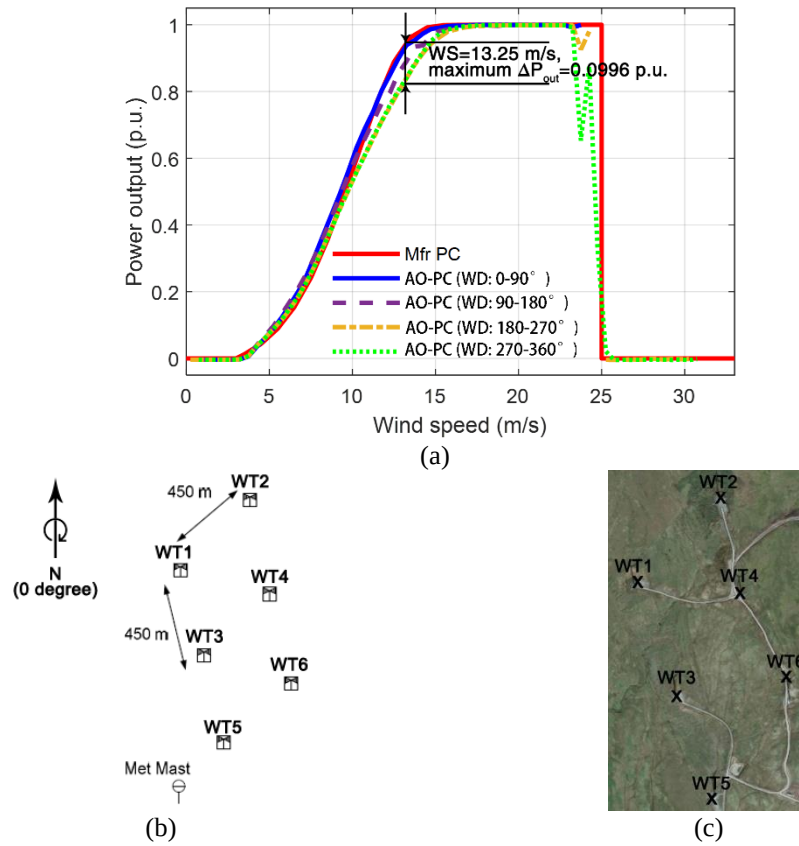


Fig. 8.1. An illustration of the impact of WD on WT power outputs: a) comparison of Mfr-PC and AO-PC models for the four main WD quadrants: b) and c) WF layout, with positions of WTs within the WF.

8.2.1 Manufacturer and Operational WT Power Curves

A typical Mfr-PC WT model specifies the power output generated by the WT for the corresponding input WS using a deterministic segmented function:

$$P_{\text{out}} = \begin{cases} 0 & v < v_i, v \geq v_o \\ f(v) & v_i \leq v < v_r \\ P_{\text{rated}} & v_r \leq v < v_o \end{cases} \quad (8.1)$$

where: v is the wind speed; $P=P_{\text{out}}$ is the corresponding (synchronous) WT power output; P_{rated} is the rated WT power output; v_i , v_r , and v_o are cut-in, rated and cut-out WS values, respectively. The nonlinear function $f(v)$, which represents the WS- P_{out} relationship, is in most Mfr PC models limited to a region between v_i and v_r speeds. However, in actual operating conditions, a range of P_{out} values is measured for a range of specific input WS values, as shown in Chapter 6 and Chapter 7. More specifically, there are 5 regions in the WT P_{out} -WS scatter plot:

- Region I: The WT cannot generate (minimum) power output, until cut-in WS is reached for at least some time within the 10-min averaging window. Therefore, the power outputs in this WS region can be represented as constant zeros and no uncertainty model is needed for WT P_{out} values. (e.g., in Fig. 6.6, when WS is below around 3 m/s)
- Region II: When WS is above the threshold and the WT is operating in non-linear regime, its P_{out} for the same input WS can vary a lot. (e.g., in Fig. 6.6, when WS is above around 3 m/s but below about 18 m/s)
- Region III: The WT power output is constant and equal to the rated value, and again, there is no uncertainty in WT P_{out} values. (e.g., in Fig. 6.6, when WS is above about 18 m/s but below around 20 m/s)
- Region IV: Strong wind or high wind gust within the 10-min averaging window may trigger the WT to shut down temporarily within the cut-out range, and the P_{out} will be a value between 0 and 1.0 p.u. (e.g., in Fig. 6.6, when WS is between around 20 and 25 m/s)
- Region V: The WT no longer produces any power output if WS is greater than a certain value, i.e., WT is fully stopped for the whole duration of 10-min period and there is no uncertainty in WT P_{out} values. (e.g., in Fig. 6.6, when WS is higher than about 25 m/s)

In conclusion, no uncertainty model is required in Regions I, III and V. In Regions II and IV, for the same input WS, P_{out} can vary significantly around the mean value and

its uncertainty is indicated with a 5th-95th percentile range around the mean, and they need to be modelled with two probabilistic models.

8.2.2 Empirical WT P_{out} Uncertainty

For the actual synchronous measurements of WS, WD and P_{out} , “empirical WT power output uncertainty” is defined as the 5th and 95th percentile range of variations around the measured mean P_{out} values for the specific WS value (e.g., for WS in bins of 0.5 m/s) and for the specific WD value (e.g., for WD in bins of 30 °), Fig. 8.2. If there are not enough P_{out} data for particular WS and WD bins, the existing bins are extended over the adjacent bins, so a statistically significant number of data is obtained. This “empirical” WT P_{out} uncertainty is used as a reference for evaluating “modelled” WT P_{out} uncertainty for both specific/known and forecasted WS and WD values (Section 8.5). Fig. 8.2 illustrates both empirical P_{out} uncertainty and possible significant errors if Mfr-PC model is used.

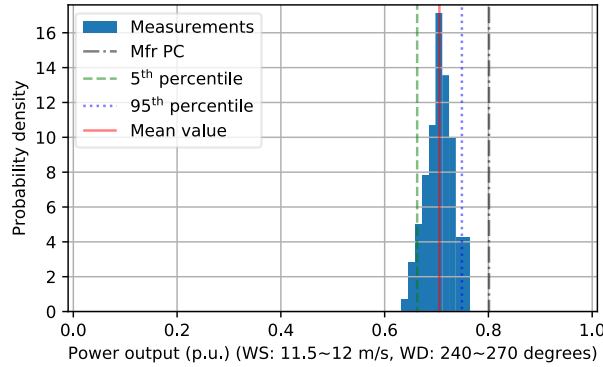


Fig. 8.2. An example of empirical/actual WT P_{out} uncertainty

8.3 Copula-Based Model

The models presented in this section are based on the Copula theory reviewed in Section 2.2.2 and conditional 2D P_{out} -WS and 3D P_{out} -WS-WD probability models. In Copula models with deterministic (i.e., known or measured) input values, the actual measured WS and WD data (without uncertainties) are used as inputs to determine mean P_{out} values and corresponding 5th-95th percentile range (i.e., uncertainty in P_{out}). In Copula models with probabilistic input values, the forecasted WS and WD values with their uncertainty ranges obtained by the MC model (see next section) are used to

calculate conditional probability density of P_{out} , its mean value and 5th-95th percentile uncertainty ranges.

As previously discussed, simpler and more widely used single Copula formulations include elliptical Copulas (e.g., Gaussian Copula and t -Copula) and Archimedean Copulas (e.g., Clayton Copula, Frank Copula and Gumbel Copula). However, they cannot represent multimodal correlations accurately. Gaussian mixture Copula model (GMCM) proposed by [199] has proven to be more flexible and accurate in handling complex correlations, and the Copula distribution and density functions can be written as:

$$C_{1,2,\dots,d}(F_1, F_2, \dots, F_d) = \Psi_{1,2,\dots,d}(\Psi_1^{-1}, \Psi_2^{-1}, \dots, \Psi_d^{-1}) \quad (8.2)$$

$$c_{1,2,\dots,d}(F_1, F_2, \dots, F_d) = \frac{\psi_{1,2,\dots,d}(\Psi_1^{-1}, \Psi_2^{-1}, \dots, \Psi_d^{-1})}{\prod_{i=1}^d \psi_i(\Psi_i^{-1})} \quad (8.3)$$

where: $\psi_{1,2,\dots,d}(\cdot)$ and $\Psi_{1,2,\dots,d}(\cdot)$ are the probability density and distribution functions of Gaussian mixture model (GMM), and ψ_i and Ψ_i^{-1} are marginal density function and inverse distribution function at i -th dimension, $\psi_i(F_i)$ and $\Psi_i^{-1}(F_i)$.

A key advantage of using GMM is that the marginal, conditional and joint distributions of GMM all follow GMM [45]. Consider a GMM of $\mathbf{x} = [\mathbf{x}_a, \mathbf{x}_b]^T$ with n components (with mixture weights vector $\boldsymbol{\Omega} = [\omega_1, \omega_2, \dots, \omega_k]^T$), and k -th component with mean vector $\boldsymbol{\mu}_k = [\boldsymbol{\mu}_{k,a}, \boldsymbol{\mu}_{k,b}]^T$ and covariance matrix $\boldsymbol{\Sigma}_k = \begin{bmatrix} \boldsymbol{\Sigma}_{k,aa} & \boldsymbol{\Sigma}_{k,ab} \\ \boldsymbol{\Sigma}_{k,ba} & \boldsymbol{\Sigma}_{k,bb} \end{bmatrix}$. The conditional distribution $P(\mathbf{x}_a|\mathbf{x}_b)$ is still a GMM, with components weights normalised as $\omega_{k,a|b} = \frac{\omega_k \times N(\mathbf{x}_b|\boldsymbol{\mu}_{k,b}, \boldsymbol{\Sigma}_{k,bb})}{\sum_{l=1}^n \omega_l \times N(\mathbf{x}_b|\boldsymbol{\mu}_{l,b}, \boldsymbol{\Sigma}_{l,bb})}$. The k -th component follows Gaussian distribution, with mean vector $\boldsymbol{\mu}_{k,a|b} = \boldsymbol{\mu}_{k,a} + \boldsymbol{\Sigma}_{k,ab} \times \boldsymbol{\Sigma}_{k,bb}^{-1} \times (\mathbf{x}_b - \boldsymbol{\mu}_{k,b})$, and covariance matrix $\boldsymbol{\Sigma}_{k,a|b} = \boldsymbol{\Sigma}_{k,aa} - \boldsymbol{\Sigma}_{k,ab} \times \boldsymbol{\Sigma}_{k,bb}^{-1} \times \boldsymbol{\Sigma}_{k,ba}$.

8.3.1 P_{out} -WS Model: GMCM Framework

Let x_1 represents P_{out} and x_2 represents WS. As P_{out} is a truncated variable, with minimum typically equal to 0 and maximum typically equal to rated power output, it is transformed into unconstrained linear space as:

$$x_{1;\text{new}} = \log \left(\frac{x_{1;\text{old}} - 0}{P_{\text{rated}} - x_{1;\text{old}}} \right) \quad (8.4)$$

WS is also a truncated variable (it cannot be less than 0) and similar transformation is applied: to build a P_{out} -WS model with GMCMM, first, the GMMs are used as marginal distribution functions, $F_1(\cdot)$ and $F_2(\cdot)$, to transfer x_1 and x_2 into cumulative distribution function domain, i.e., $U(0,1)$. The joint probability density of x_1 and x_2 is then:

$$f_{12}(x_1, x_2) = c_{1,2}(F_1, F_2) f_1(x_1) f_2(x_2) \quad (8.5)$$

For deterministic WS input, the conditional probability density of P_{out} is obtained as:

$$f_{1|2}(x_1|x_2) = \frac{f_{12}(x_1, x_2)}{f_2(x_2)} = c_{1,2}(F_1, F_2) f_1(x_1) \quad (8.6)$$

If the given WS is not deterministic, but features uncertainties, e.g., in a range from $x'_2 \leq x_2 \leq x''_2$, then the theoretical conditional probability density of P_{out} can be derived as:

$$\begin{aligned} & f_{1|2}(x_1|x'_2 \leq x_2 \leq x''_2) \\ &= \frac{f_{12}(x_1, x'_2 \leq x_2 \leq x''_2)}{f_2(x'_2 \leq x_2 \leq x''_2)} \\ &= \frac{\int_{x'_2}^{x''_2} f_{12}(x_1, x) dx}{\int_{x'_2}^{x''_2} f_2(x) dx} \\ &= \frac{1}{F_{2''} - F_{2'}} \frac{F_{12}(x_1, x'_2 \leq x_2 \leq x''_2)}{\partial x_1} \\ &= \frac{1}{F_{2''} - F_{2'}} \left(\frac{\partial C_{1,2''}(F_1, F_{2''})}{\partial x_1} - \frac{\partial C_{1,2'}(F_1, F_{2'})}{\partial x_1} \right) \end{aligned} \quad (8.7)$$

This chapter, however, proposes another numerical method to calculate $f_{1|2}(x_1|x'_2 \leq x_2 \leq x''_2)$, as in the processed data x''_2 and x'_2 may be out of the modelled WS range. The proposed method is “grid search method”, which is based on a mixture distribution combining both deterministic and probabilistic models, which can be formulated as:

$$\begin{aligned}
& f_{1|2}(x_1|x_2' \leq x_2 \leq x_2'') \\
&= \frac{\int_{x_2'}^{x_2''} f_{1,2}(x_1, x) dx}{\int_{x_2'}^{x_2''} f_2(x) dx} \\
&= \frac{\int_{x_2'}^{x_2''} f_2(x) f_{1|2}(x_1|x) dx}{\int_{x_2'}^{x_2''} f_2(x) dx} \\
&= \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n f_2(x_i) f_{1|2}(x_1|x_i)}{\sum_{i=1}^n f_2(x_i)} \tag{8.8}
\end{aligned}$$

where $x_2' \leq x_i \leq x_2''$, so a suitable “data grid” is formed. Equation (8.8) is mathematically equivalent to a mixture of distributions, where the weights are $f_2(x_i)$ and the normalised factor is $\sum_{i=1}^n f_2(x_i)$. From another point of view, Equation (8.8) is also identical to using Monte Carlo to solve the complex integral. In actual calculation, since n cannot be very large due to a heavy computational burden, there may be some biases for small n , even at the risk of overestimating actual P_{out} uncertainty.

8.3.2 P_{out} -WS-WD Model: Vine-GMCM Framework

Let x_3 represent WD, and to build a correlational P_{out} -WS-WD model, as in the P_{out} -WS model, the GMMs are used as marginal distribution functions, $F_1(\cdot)$, $F_2(\cdot)$ and $F_3(\cdot)$, to transfer x_1 , x_2 and x_3 into $U(0,1)$ domain. The joint probability density of x_1 , x_2 and x_3 , $f_{123}(x_1, x_2, x_3)$ can be derived using chain rules: $f_{3|12}(x_3|x_1, x_2) \times f_{2|1}(x_2|x_1) \times f_1(x_1)$, in which $f_{3|12}(x_3|x_1, x_2)$ is equal to Equation (8.9), with $f_{2|1}(x_2|x_1) = c_{12}(F_1, F_2)f_2(x_2)$. Therefore, $f_{123}(x_1, x_2, x_3)$ can be expressed as in Equation (8.10). The key in Equation (8.10) is to estimate conditional Copula density, i.e., $c_{23|1}(F_{2|1}, F_{3|1})$, which is modelled through its arguments, i.e., $F_{2|1}(x_2|x_1)$ and $F_{3|1}(x_3|x_1)$ in Equation (8.11):

$$\begin{aligned}
& f_{3|12}(x_3|x_1, x_2) \\
& = c_{23|1}(F_{2|1}, F_{3|1})f_{3|1}(x_3|x_1) = c_{23|1}(F_{2|1}, F_{3|1})c_{13}(F_1, F_3)f_3(x_3) \quad (8.9)
\end{aligned}$$

$$\begin{aligned}
& f_{123}(x_1, x_2, x_3) \\
& = f_1(x_1)f_2(x_2)f_3(x_3)c_{12}(F_1, F_2)c_{13}(F_1, F_3)c_{23|1}(F_{2|1}, F_{3|1}) \quad (8.10)
\end{aligned}$$

$$\begin{cases} F_{2|1}(x_2|x_1) = \frac{\partial c_{12}(F_1, F_2)}{\partial F_1} \\ F_{3|1}(x_3|x_1) = \frac{\partial c_{13}(F_1, F_3)}{\partial F_1} \end{cases} \quad (8.11)$$

Clearly, the full correlation structure is made from the three Copula pairs: C_{12} , C_{13} , and $C_{23|1}$, which are all fitted using GMCM. To estimate the conditional probability density of power output for given WS and WD values, standard conditional probability formulation and further Copula derivation can be applied, with the final full expression for $f_{1|23}(x_1|x_2, x_3)$ given as:

$$\begin{aligned}
& f_{1|23}(x_1|x_2, x_3) \\
& = \frac{f_{123}(x_1, x_2, x_3)}{f_{23}(x_2, x_3)} \\
& = \frac{f_1(x_1)c_{12}(F_1, F_2)c_{13}(F_1, F_3)c_{23|1}(F_{2|1}, F_{3|1})}{c_{23}(F_2, F_3)} \\
& = \frac{f_1(x_1)c_{12}(F_1, F_2)c_{13}(F_1, F_3)}{c_{23}(F_2, F_3)} c_{23|1}\left(\frac{\partial c_{12}(F_1, F_2)}{\partial F_1}, \frac{\partial c_{13}(F_1, F_3)}{\partial F_1}\right) \quad (8.12)
\end{aligned}$$

The correlation between WS and WD is directly modelled with the additional Copula C_{23} , which is also fitted with a GMCM. If the inputs x_2 and x_3 of the model have been specified by deterministic series of values, the calculation of Equation (8.12) is straightforward. If the input WS and WD are both uncertain and in the forms of ranges, then a similar grid search approach as in Equation (8.8) could be applied:

$$\begin{aligned}
& f_{1|2,3} \left(x_1 \middle| \begin{matrix} x'_2 \leq x_2 \leq x''_2, \\ x'_3 \leq x_3 \leq x''_3 \end{matrix} \right) \\
&= \frac{\int_{x'_3}^{x''_3} \int_{x'_2}^{x''_2} f_{23}(x,y) f_{1|2,3}(x_1|x,y) dx dy}{\int_{x'_3}^{x''_3} \int_{x'_2}^{x''_2} f_{23}(x,y) dx dy} \\
&= \lim_{\substack{n_1 \rightarrow \infty, \\ n_2 \rightarrow \infty}} \frac{\sum_{j=1}^{n_2} \sum_{i=1}^{n_1} f_{23}(x_i, y_j) f_{1|2,3}(x_1|x_i, y_j)}{\sum_{j=1}^{n_2} \sum_{i=1}^{n_1} f_{23}(x_i, y_j)} \quad (8.13)
\end{aligned}$$

where: $x'_2 \leq x_i \leq x''_2$, $x'_3 \leq y_i \leq x''_3$, and the space determined by x'_2 and x''_2 , and x'_3 and x''_3 is divided into a limited number of grids. It is also not required for the input WS and WD to be already correlated with each other, because $f_{23}(x_i, y_j)$, which may be calculated by Copula as $c_{23}(F_2, F_3)f_2(x_2)f_3(x_3)$, will model the correlation between the WS and WD by assigning different weights to different (WS, WD) pairs in the “mixing” process. Another reason why the analytical approach as in Equation (8.7) is not applied is that the Copula distribution function for Vine Copula is unknown and it is assumed to be intractable, therefore requiring numerical integration; also, the grid search method may suffer bias issues, if values of n_1 and/or n_2 are not sufficiently large.

The only requirement during the training process is to fit marginal GMM distributions and the Copula GMCM models, [199], to evaluate Equations (8.6) - (8.8) and Equations (8.12) - (8.13). To have a valid distribution on the full range of x_1 (P_{out} range), array values between x'_1 and x''_1 are created, $\mathbf{x}_1 = [x'_1, x'_1 + step, x'_1 + 2step, \dots, x''_1]$ and conditional distributions, Equations (8.6)-(8.8) and Equations (8.12) - (8.13), generalised as $f(x_1|inputs)$, are calculated for this $\mathbf{x}_1$ array. The results are PDF arrays for $x'_2 \leq x_2 \leq x''_2$, where renormalisation is done to ensure the integration of PDF is equal to 1. Finally, trapezoidal numerical integration is used to calculate the corresponding CDF $F(x_1|inputs)$ from the above PDF and statistical inference in terms of mean value and 5th-95th percentiles for $x_1|inputs$ is obtained directly from obtained PDF $f(x_1|inputs)$ and CDF $F(x_1|inputs)$.

8.4 Markov Chain Model Forecasted Input WS and WD Data with Uncertainties

The MC based models presented in this section as input data at a given instant t_h use the actual measured $WS(t_h)$ and $WD(t_h)$ to forecast the corresponding WS and WD values at the instant $t_h + m\Delta t$ and then to determine the mean P_{out} value, denoted as $E[P_{out}(t_h + m\Delta t)]$. The models also calculate related uncertainty as the 5th-95th percentile range around mean P_{out} for given measured $WS(t_h)$ and $WD(t_h)$.

8.4.1 Discretisation Process

Table 8.1 lists MC model discretisation classes: WS has been discretized in $N_{WS} = 27$ classes, where each class is represented by its central value, ws_i . In order to adapt WS classes to the power curve behaviour, the first and the last classes are chosen so the corresponding P_{out} is null. On the other hand, WD has been discretised in $N_{WD} = 12$ classes, each of 30 ° length and centred around the cardinal points.

Table 8.1 Discretisation Classes for WD (Columns 1-3), and WS (Columns 4-7)

Class	WD [°]	Central Cardinal Point	Class	WS [m/s]	Class	WS [m/s]
1	[345;15]	N	1	[0; 3]	14	[9; 9.5]
2	[15;45]	N-NE	2	[3; 3.5]	15	[9.5; 10]
3	[45;75]	NE-E	3	[3.5; 4]	16	[10; 10.5]
4	[75;105]	E	4	[4; 4.5]	17	[10.5; 11]
5	[105;135]	SE-S	5	[4.5; 5]	18	[11; 11.5]
6	[135;165]	S-SE	6	[5; 5.5]	19	[11.5; 12]
7	[165;195]	S	7	[5.5; 6]	20	[12; 12.5]
8	[195;225]	S-SW	8	[6; 6.5]	21	[12.5; 13]
9	[225;255]	SW-W	9	[6.5; 7]	22	[13; 13.5]
10	[255;285]	W	10	[7; 7.5]	23	[13.5; 14]
11	[285;315]	NW-W	11	[7.5; 8]	24	[14; 14.5]
12	[315;345]	N-NW	12	[8; 8.5]	25	[14.5; 15]
			13	[8.5; 9]	26	[15; 25]
					27	[25; +∞]

8.4.2 Forecasted WS Data

For WS forecasting, the first order MC (FOMC) model from [225] is used, where the state variable is the average WS over time interval $[t_{h-1\Delta t}, t_h]$ and $t_h = h\Delta t$. The FOMC model is based on the construction of a one-step transition probabilities matrix, $\bar{P}_{WS}$, whose generic element p_{ij} represents the transition probability between the class i and class j in one Δt . The dimension of this matrix is $N_{WS} \times N_{WS}$, equal to 27x27 in the considered case.

The method starts from the observed state probability vector at time t_h , whose elements are all zero, except the one corresponding to the measured value, e.g., $WS(t_h)$:

$$\pi(t_h) = [\pi_1(t_h), \dots, \pi_{N_{WS}}(t_h)] \quad (8.14).$$

Afterwards, the method forecasts the state probability vector at m steps ahead (one step-ahead for $m=1$) as:

$$\hat{\pi}(t_{h+m\Delta t}) = \pi(t_h)[\bar{P}_{WS}]^m \quad (8.15)$$

The conditional point predictor of WS is obtained by:

$$WS(t_{h+m\Delta t})|WS(t_h) = \sum_{i=1}^{N_{WS}} ws_i \hat{\pi}_i(t_{h+m\Delta t}) \quad (8.16)$$

Equation (8.16) represents the probability mass function (PMF) of the forecasted WS and can therefore be used to obtain the cumulative mass function (CMF) and the required 5th and 95th percentiles uncertainty range.

8.4.3 Forecasted WS-WD Data

Introducing a new bivariate random variable, $\bar{W} = [WS, WD]$, it is possible to obtain a one-dimensional form of random variable W, characterised by a number of classes $N_{WS}N_{WD}$, each representative of a given pair of WS and WD classes. In this way, a Markov transition probability matrix can be constructed, taking as state variable W. The result is a matrix $\bar{P}_W$ of dimension $(N_{WS} \cdot N_{WD}) \times (N_{WS} \cdot N_{WD})$, i.e., 324x324 in the considered case, whose generic element p_{ij} represents the transition probability between class i and class j of W. The introduction of a one-dimensional random variable W allows to use Equations (8.14)-(8.16) to obtain input data for the 2D and 3D Copula models with WS and WD uncertainties.

8.5 Case Study and Obtained Results

Available measurements are 3-year Scotland Wind Farm Dataset introduced in Section 6.2. The first two years of measurements are used to build the models and then MC model is used to forecast WS and WD values, together with their joint and marginal

probabilities, in the third year. The data from the third year are used to test/validate presented models.

8.5.1 Metrics for Comparison of Uncertainty Ranges

The accuracy of uncertainty ranges estimated by different models is quantified in two ways: one focuses on the errors in capturing two uncertainty interval values, while another attaches more importance to the length of the uncertainty range. The following quantities are calculated by the different models and for each season.

- $\hat{P}_{out,mean}$, $\hat{P}_{out,5\%}$ and $\hat{P}_{out,95\%}$ which are the mean value predicted by the model, the 5th and 95th percentile values, respectively. These values can be assessed using goodness-of-fit indices, such as MAE and RMSE. The corresponding errors are denoted as: δu_{mean} , $\delta u_{5\%}$, and $\delta u_{95\%}$, respectively. In addition, the arithmetic mean value of $\delta u_{5\%}$, and $\delta u_{95\%}$, associated with the interval values estimation are calculated as the errors of estimated uncertainty interval values, using the following expression:

$$\delta u = (\delta u_{5\%} + \delta u_{95\%})/2 \quad (8.17)$$

- If the empirical uncertainty range (see Section 8.2.2) length is $R_e = (P_{out,95\%} - P_{out,5\%})$, and the modelled uncertainty range length is $R_M = (\hat{P}_{out,95\%} - \hat{P}_{out,5\%})$, the normalised uncertainty range estimation error is defined as:

$$\Delta u = (R_M/R_e) - 1 \quad (8.18)$$

Finally, energy-based error metrics introduced in Section 3.4.4 are also applied.

8.5.2 Results of Copula-Based Models without Uncertainties in WS and WD

Comparison of a one-day time series for a day in the winter season of 2D P_{out} -WS GMCM and 3D P_{out} -WS-WD Vine-GMCM Copula model results for mean P_{out} values and corresponding uncertainty ranges when measured WS and WD values are used as

model inputs (without WS and WD uncertainties) is shown in Fig. 8.3, together with the results obtained by using manufacturer power curve (Mfr-PC).

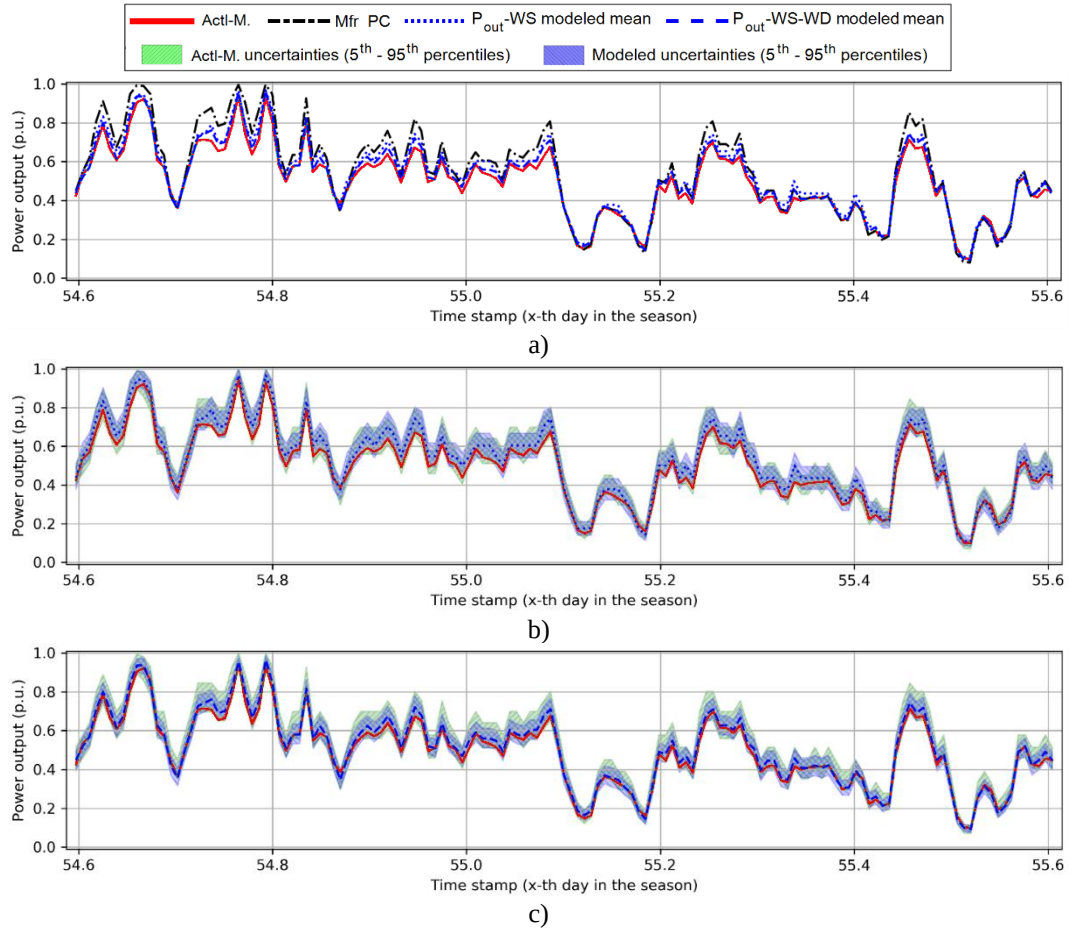


Fig. 8.3. Comparison of results for one whole day in winter for input WS and WD without uncertainties: a) mean values; b) 2D Copula Pout-WS GMCM model with Pout uncertainties; and c) 3D Copula Pout-WS-WD with Pout uncertainties.

Table 8.2 Comparison of P_{out}-WS Model, P_{out}-WS-WD Model, Manufacturer Power Curve and Actual Measurements for the Same Day in Fig. 8.3

Model	MAE (p.u.)	RMSE (p.u.)	δu (MAE)	δu (RMSE)	Δu	E _O (MWh)	E _O (%)	E _U (MWh)	E _U (%)	E _T (MWh)	E _T (%)
Mfr-PC	0.057	0.072	-	-	-	3.901	12.123	0.199	5.068	3.703	10.258
P _{out} -WS	0.023	0.027	0.018	0.021	- 0.122	1.591	4.907	0.067	1.834	1.524	4.221
P _{out} -WS-WD	0.019	0.023	0.023	0.027	- 0.221	1.257	4.120	0.126	2.256	1.130	3.132

Table 8.3 Comparison of P_{out} -WS Model, P_{out} -WS-WD Model, Manufacturer Power Curve and Actual Measurements for Whole Seasons

Model	MAE (p.u.)	RMSE (p.u.)	δu (MAE)	δu (RMSE)	Δu	E _o (MWh)	E _o (%)	E _u (MWh)	E _u (%)	E _T (MWh)	E _T (%)
Spring											
Mfr-PC	0.026	0.041	-	-	-	73.326	6.033	31.505	5.682	41.821	2.363
P_{out} -WS	0.019	0.028	0.014	0.019	- 0.205	55.775	5.178	21.100	3.046	34.675	1.959
P_{out} -WS-WD	0.020	0.030	0.016	0.023	- 0.222	58.167	5.605	22.096	3.018	36.071	2.038
Summer											
Mfr-PC	0.030	0.049	-	-	-	151.525	10.541	37.676	7.682	113.849	5.905
P_{out} -WS	0.017	0.025	0.013	0.018	- 0.311	90.338	6.111	18.964	4.217	71.374	3.702
P_{out} -WS-WD	0.016	0.024	0.014	0.020	- 0.328	83.514	5.758	18.298	3.831	65.216	3.383
Autumn											
Mfr-PC	0.037	0.059	-	-	-	195.288	8.476	36.986	7.800	158.302	5.698
P_{out} -WS	0.018	0.031	0.015	0.021	- 0.352	84.975	4.405	28.466	3.352	56.509	2.034
P_{out} -WS-WD	0.016	0.029	0.019	0.026	- 0.421	75.371	4.073	26.129	2.816	49.242	1.772
Winter											
Mfr-PC	0.029	0.044	-	-	-	92.428	6.594	66.920	8.513	25.508	1.166
P_{out} -WS	0.020	0.032	0.013	0.019	- 0.235	32.164	3.801	76.583	5.708	-44.419	- 2.030
P_{out} -WS-WD	0.018	0.029	0.014	0.020	- 0.275	31.989	3.787	68.939	5.133	-36.950	- 1.689
Whole year											
Mfr-PC	0.031	0.050	-	-	-	512.567	8.061	173.088	7.508	339.479	3.918
P_{out} -WS	0.018	0.029	0.014	0.019	- 0.284	263.252	4.938	145.113	4.354	118.139	1.364
P_{out} -WS-WD	0.017	0.028	0.016	0.022	- 0.322	249.041	4.805	135.462	3.892	113.579	1.311

Table 8.2 quantifies the benefits of the P_{out} -WS-WD model over P_{out} -WS model for the same day in Fig. 8.3, while Table 8.3 compares performance of both models in separate seasons and for the whole year.

The results of 2D and 3D models are numerically close to each other, but 3D P_{out} -WS-WD model is more accurate for mean values and gives smaller uncertainty ranges, compared to 2D P_{out} -WS model, for all seasons (except for spring) and also for the whole year-length datasets. This confirms that introducing WD as an additional explanatory variable can increase the confidence in the evaluation of uncertainties in the WT's P_{out} values. For the spring season results, although P_{out} -WS-WD model still provides smaller uncertainties, its mean outputs are worse than P_{out} -WS model. One possible reason is that the WT3 yawing mechanism was faulty or not operating in the

same way as during the spring seasons in the first two years. The fitted Copula model models the power output of the WT given certain WS and WD. If the yawing mechanism is significantly changed, then the WT may behave differently. In such case, the estimation of $P(P_{out}|WS, WD)$ may even be worse than $P(P_{out}|WS)$ because the input variable, WD, does not make the model response reflect the actual generation process of the WT. However, the WF log of events and maintenance record was not available to confirm that assumption.

8.5.3 Results of Copula-Based Models with Uncertainties in WS and WD

Next set of Copula based model results compares estimated time series for mean P_{out} values and corresponding 5th-95th percentile uncertainty ranges when input WS and WD values are provided with uncertainties (obtained from MC model). Fig. 8.4 and Table 8.4 give results for the same winter day in Fig. 8.3 (there are no results for Mfr-PC, as it is a deterministic model and cannot handle uncertain inputs). The results confirm that the proposed grid search method allows the Copula models with uncertain WS and WD inputs to estimate related uncertainties in P_{out} values. The lag in the P_{out} series, for both mean values and uncertainty ranges, is as expected and due to the inherent feature of the forecasting processes.

In this case, P_{out} uncertainty is evaluated considering both uncertainties in input wind energy resource (i.e., WS and WD) and uncertainties of the conversion of wind energy into electricity by WT. Consequently, the 5th-95th percentile range is much wider than for the results in Fig. 8.3. However, the 3D P_{out} -WS-WD Copula model still provides smaller uncertainty estimation ranges, partly due to the fact that the uncertainty range is an inherent feature of Copula model. The worse estimation about mean values may be because the accuracy of mean values is more determined by the accuracy of input data: 2D P_{out} -WS model considers only WS, and therefore only errors/bias in WS data, while 3D P_{out} -WS-WD model includes errors/bias in both WS and WD data. Also, some bias may be introduced by the proposed grid search method.

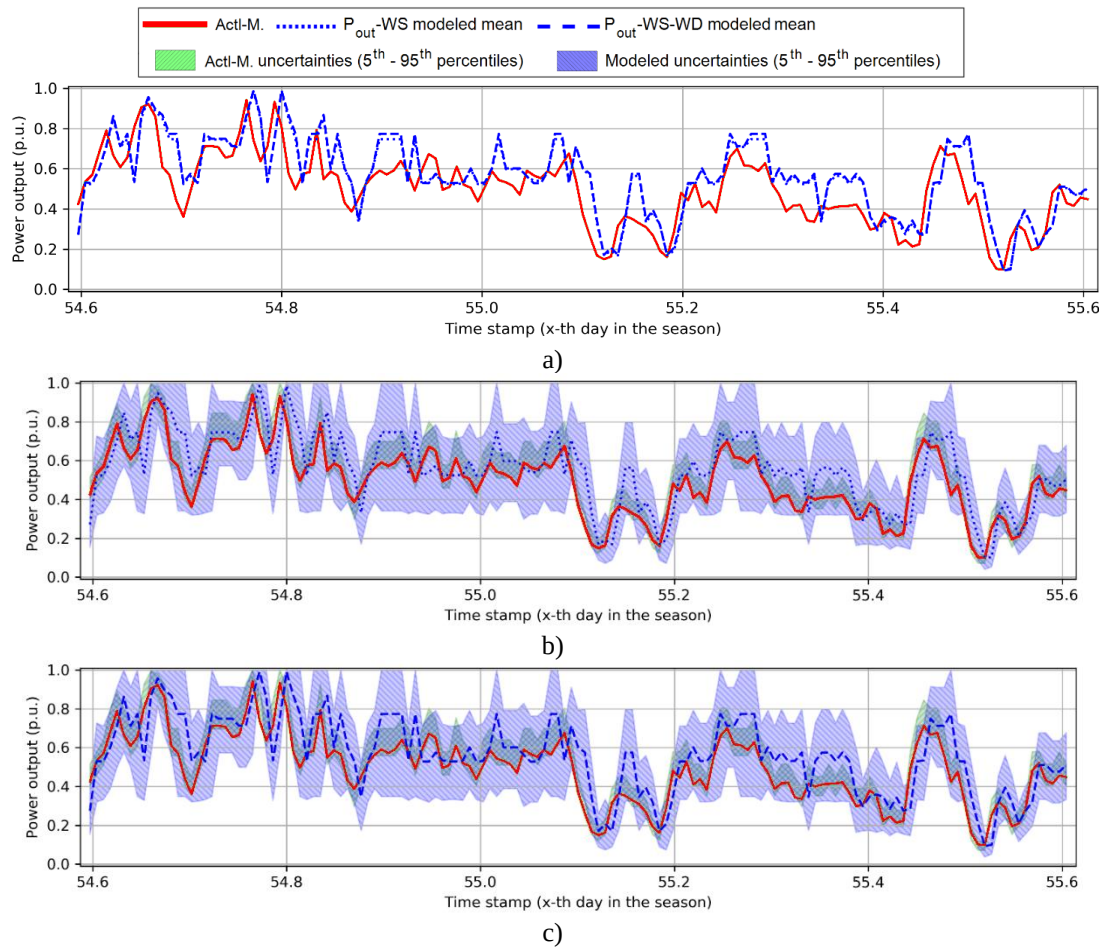


Fig. 8.4. Comparison of results for one whole day in winter (same as in Fig. 8.3) for input WS and WD with uncertainties: a) mean values; b) 2D Copula P_{out} -WS GMCM model with P_{out} uncertainties; and c) 3D Copula P_{out} -WS-WD with P_{out} uncertainties.

Table 8.4 Comparison of P_{out} -WS model and P_{out} -WS-WD model for the same day in Fig. 8.4

Model	MAE (p.u.)	RMSE (p.u.)	δu (MAE)	δu (RMSE)	Δu	E _O (MWh)	E _O (%)	E _U (MWh)	E _U (%)	E _T (MWh)	E _T (%)
P_{out} -WS	0.108	0.133	0.161	0.195	2.512	6.048	24.471	1.812	15.917	4.236	11.735
P_{out} -WS-WD	0.114	0.141	0.160	0.194	2.494	6.606	26.930	1.691	14.615	4.915	13.617

8.6 Discussions and Conclusions

Due to strong and stochastic spatio-temporal variations of wind energy resource, the outputs of wind-based electricity generation systems are difficult to predict and are typically evaluated with associated uncertainty levels. As the wind-based electricity generation systems' contribution to the total system generation capacity increases, the assessment and quantification of these uncertainties are becoming increasingly

important, due to their greater impact on the variations of network power flows and related uncertainties in network operating conditions.

This chapter investigates the importance of including WD as an additional input explanatory variable to the WS for assessing uncertainty in the WT P_{out} . It provides a comprehensive evaluation of uncertainties in WT's P_{out} by comparing a 2D P_{out} -WS model with a 3D P_{out} -WS-WD model for two general cases: a) for the specific or known input WS and WD values (i.e., WS and WD are without uncertainties), and b) for the forecasted or estimated input WS and WD values (i.e., WS and WD are both with associated uncertainties). For that purpose, 2D and 3D Copula based models are combined with 2D and 3D MC models from [225], which are in this chapter used to forecast the mean WS and WD values, together with their 5th-95th percentile ranges of variations, representing uncertainties in input WS and WD data.

The results for an actual WT demonstrate that although input WS has a stronger impact on the WT P_{out} values, the inclusion of WD in the analysis might provide noticeable improvements in models with no uncertainties in input WS and WD data, e.g., in models used for hindcasting studies.

The results for input WS and WD with uncertainties confirm that WS is a much stronger contributor to the total P_{out} uncertainty than WD, which for the considered WT provided at best only a marginal improvement. The main reason is in wide ranges of variations (i.e., large spreads of uncertainties) of WD values associated with all considered WS uncertainty ranges. This part of the presented analysis and developed models allows to obtain useful information for uncertainty importance analysis, as required in e.g., wind energy forecasting studies.

The results for two considered cases (with and without uncertainties in input WS and WD data) are also useful for quantifying the impact of variations of predictor variables of the considered model, while the results of 2D and 3D models can be considered in terms of increased “knowledge” on wind energy resource through the inclusion of WD as an additional explanatory variable. In this context, separation of obtained results into seasons allowed to implicitly include temperature in the analysis, as temperature changes in different seasons will impact variations in air density, which will ultimately

influence changes in WT's P_{out} for the same input WS and WD values in different seasons. Further analysis will be presented in the next chapter.

Although the case studies in this chapter show a WT from a relatively small WF, the conclusions are also useful for other larger WFs. A WF is an aggregation of WTs, and a large WF may see even stronger impacts from wind direction, layout and wake effects. The proposed model is also scalable, which will be used to model a four-dimensional correlation among power output, wind speed, wind direction and air density for two WFs with 16 and 20 WTs in the next chapter.

Chapter 9

Bayesian CNN-BiLSTM and Vine-GMCM Based Probabilistic Forecasting of Hour-Ahead Wind Farm Power Outputs

9.1 Introduction

The importance of the accurate forecasting of power outputs of wind-based generation systems is increasing, as their contributions to the total system generation are rising. However, wind energy resource exhibits strong and stochastic spatio-temporal variations, which further combine with the uncertainties in WF operating regimes (ORs), i.e., numbers of wind turbines in normal operation, under curtailment, or that are faulty/disconnected.

As reviewed in Section 2.5.3, most of the WF P_{out} prediction methods are deterministic, but the probabilistic forecasting is becoming increasingly relevant, if not necessary, with the higher penetration of wind energy. In particular, Bayesian framework [226] and Copula models [227] have been proven to handle and analyse the uncertainty effectively in wind energy forecasting. However, there is a lack of a general work integrating prediction of multi-dimensional weather variables, WF operating regimes, WF P_{out} modelling and uncertainty analysis into one general framework, aimed at a more robust and reliable probabilistic WF P_{out} forecasting.

A possible methodology for building such a general framework is presented in this chapter. It uses actual WS and WD measurements from the WFs in two clusters in close locations in Croatia, while synchronous AD data for these locations are obtained from [116], assuming that AD does not exhibit significant differences within the considered regions. The chapter combines and extends the previous work in Chapter 3, Chapter 6 and Chapter 8, presenting the following novel contributions:

- Use of Bayesian convolutional neural network-bidirectional long short-term memory (Bayesian CNN-BiLSTM) method to obtain multi-dimensional correlated probabilistic predictions for WS, AD and WD, allowing to capture both epistemic and aleatoric uncertainties [37]. Although Bayesian NN models have been

previously reported for wind energy forecasting, most of them are based on the conventional fully connected NN models, e.g., single-layer perceptron [228], multilayer perceptron in [229, 230], while very few utilised different models, e.g., convolutional neural network [231] or convolutional Gated Recurrent Units (GRU) [232]. In time series regression problems, LSTM-based models are widely researched and reported to achieve satisfactory accuracy. However, to the best of the PhD student's knowledge, application of Bayesian CNN-BiLSTM model for wind power forecasting is still missing.

- Forecasting of WF ORs, where a multi-state WF model is used to more efficiently handle the uncertainties propagated by the changes in WF ORs. This chapter shows that WF OR is an important predictor variable, which is almost completely ignored in the previous work on WF forecasting.
- Development of a high-dimensional Vine-Gaussian mixture Copula model (Vine-GMCM) for modelling correlations among WS, AD, WD and P_{out} , which is then combined with Bayesian CNN-BiLSTM to give the final WF P_{out} forecasting. Previous work on Vine Copula (e.g., [233] and on wind energy forecasting) used much simpler and more common single Copula formulations to represent pair correlations, like Archimedean Copulas (e.g., Clayton Copula, Frank Copula and Gumbel Copula), and elliptical Copulas (e.g., Gaussian and t-Copula), which, according to [199], cannot represent multimodal correlation as accurately as GMCM. The previous chapter combines Vine Copula and GMCM, and work presented here extends the original three-dimensional model (joint WS-WD- P_{out} distribution) to a four-dimensional model (WS-AD-WD- P_{out}), confirming importance of AD.

9.2 Description of Modelled WFs, Available Measurements and Used Evaluation Metrics

The analysis in this chapter uses historical 2-minute resolution measurements (aggregated and averaged into 1-hour resolution) for WS, WD and P_{out} available for a period of around two years from two WFs in Croatia, named “Bruska” and “Jelinak”. There are 16 Siemens SWT-93, 2.3 MW WTs in Bruska WF, while Jelinak WF has 20 ACCIONA Windpower AW 82/1500, 1.5 MW WTs. The coordinate of Bruska and

Jelinak WFs are 44°05'34.8"N 15°44'09.6"E and 43°33'14.4"N 16°10'01.2"E, respectively. And the parameters of Bruska and Jelinak WFs are presented in Table 9.1 and Table 9.2, respectively. Most of the WFs in Croatia can be roughly divided into three clusters, based on their geographical locations within the three areas of approximately 50×50 km.

Table 9.1 Bruska Wind Farm Parameters

WT Name	WT Cut-in WS [m/s]	WT Cut-out WS [m/s]	WT Rated Power [MW]	Number of WTs	WF Location
Siemens SWT-93	4	25	2.3	16	44°05'34.8"N 15°44'09.6"E

Table 9.2 Jelinak Wind Farm Parameters

WT Name	WT Cut-in WS [m/s]	WT Cut-out WS [m/s]	WT Rated Power [MW]	Number of WTs	WF Location
Acciona AW 82/1500	3	20	1.5	20	43°33'14.4"N 16°10'01.2"E

The availability of the data varies for different time periods and different WF locations. Therefore, the representative segment was chosen for the two selected WFs. The AD is considered as an additional input parameter, and synchronous AD values for the considered WF locations are taken from [116]. In predicting the weather variables, this paper also adopts the measurements of two other WFs: one in the same cluster as Bruska WF and another for Jelinak WF. The training set lengths are 103 and 99 weeks for Bruska and Jelinak WFs, respectively, while the following consecutive 168 hours (named “test week”, i.e., the 104th and 100th week) and whole following year (named “test year”) are used as the test sets. As there were many outliers and missing values in the test year dataset, only the measurements with good data quality are used: 4,202 and 4,272 hourly samples for Bruska and Jelinak WFs, respectively (both around 25 weeks). All measurements and other data used for the analysis can be accessed in [234]. In the tests, rolling forecasting is applied, where each hour is forecasted individually in the sequence, and once the actual value at the forecasted time stamp has been observed, it is used for forecasting the next ones.

In the presented results, deterministic forecasting errors are assessed by two goodness-of-fit indices, MAE and RMSE (see Section 2.3.1), as well as energy-based error

metrics (see Section 3.4.4). For evaluating probabilistic errors, pinball loss, Winkler score and continuous ranked probability score (CRPS) (see Section 2.3.2) are used. In the calculation of pinball loss, 0.1, 0.2, ..., 99.9 percentiles are used in each calculation. The significance level is set as 0.1 when calculating Winkler score. Performance/accuracy of the model will increase if the calculated error metric values are lower; in the presented tables, mean values of 168 hourly values are calculated for each metric and for each method.

As discussed in Chapter 6, a WF typically consists of multiple individual WT's and at any moment, each WT in a WF may be in one of three general states: a) normally operating, b) curtailed, or c) disconnected due to the planned or unplanned outages. The various WF ORs can be divided into one fully operating regime (all WT's in a WF are operating normally) and a number of other partly operating regimes, in which one or more WT's are in curtailment and/or outage states. For a WF with installed N WT's of the same type, Chapter 6 shows that total P_{out} of the WF is the arithmetic sum of outputs of all normally operating WT's, plus a bias, which is power produced by curtailed WT's (CLT).

In this chapter, curtailment is neglected for simplicity and because the time for which WT's in considered WFs were in curtailment was very short. According to "unavailability plans" data provided by the WF Operators, Bruska and Jelinak WFs had WT-level curtailment for 0.19 % and 0 % of total operational time. By neglecting CLT, WF OR can be simply represented using only the number of WT's in normal operation, a . This is illustrated in Fig. 9.1 for Bruska WF.

The outages of WT's can be: i) planned, e.g., due to a maintenance or servicing, and ii) unplanned, e.g., due to a fault. Planned WT outages are typically known in advance (network operators should be informed on servicing/maintenance) and are available as inputs in the model during the training stage. Unplanned WT outages usually can be identified from the WF event-log, but they were not available for two considered WFs. Therefore, and as illustrated in Fig. 9.1, available measurements are used to formulate simple clustering algorithm for WF P_{out} values with curtailed WT's neglected, allowing to estimate number of WT's in normal operation, i.e., WF OR, which is then considered in the forecasting model. For Bruska WF in Fig. 9.1, OR = (16, 15, and 14), contribute

(15%, 13%, and 11%) to total operating time, respectively, indicating that WF P_{out} estimations would be strongly impacted if WF ORs are not considered.

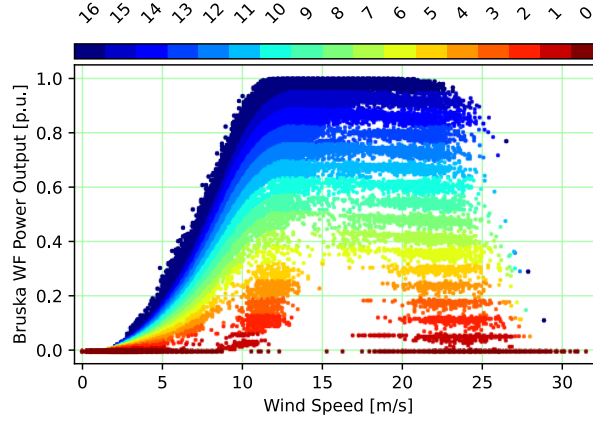


Fig. 9.1. An example of clustered P_{out} values for different ORs of Bruska WF: different colours and numbers at the top of the Fig. represent different numbers of WTs in normal operation.

9.3 Bayesian CNN-BiLSTM for Forecasting Weather Variables and Wind Farm Operating Regimes

Deep learning models are increasingly used in many areas, including energy forecasting [6, 235], and although they have better performance than other approaches in many applications [10], they are often criticised as “black box” models, i.e., as difficult to interpret, and their performance is sensitive to both hyperparameter tuning and quality and quantity of training samples. Bayesian approach extends deterministic deep learning with posterior inference, where model parameters and outputs are assumed to be distributions, not deterministic points. For given data, Bayesian update of distributions is:

$$p(w|D) = \frac{p(D|w)p(w)}{\int p(D|w')p(w') dw'} \quad (9.1)$$

where: $p(\cdot)$ is the probability density, D is the observation data, w is the model parameter, $p(w)$ is the prior distribution of model parameter, $p(D|w)$ is the likelihood of observation data given w , and $p(w|D)$ is the posterior distribution of w for given observation data. The posterior distribution determines the parameter after considering observation data and, therefore, can represent the uncertainty.

9.3.1 Deterministic CNN-BiLSTM Model

The detailed theoretical description about the RNN and CNN can be found in Section 2.2.1. This chapter will utilise the CNN-BiLSTM architecture for forecasting purpose. In the developed CNN-BiLSTM model predictor features are first extracted with the CNN architecture, before entering BiLSTM module. The convolution will reduce input feature variations, making temporal learning of BiLSTM more accurate [133]. Fig. 9.2 shows a general architecture of CNN-BiLSTM model.

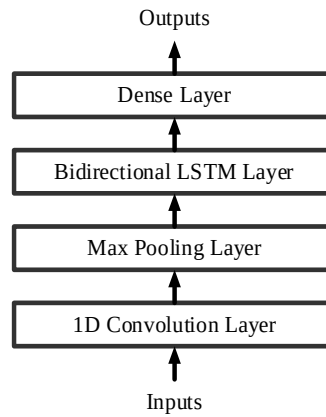


Fig. 9.2. A general CNN-BiLSTM architecture.

The inputs firstly enter the 1D convolution layer, followed by a layer performing the max pooling operation, to extract and embed the features. The third layer is a BiLSTM layer, which learns sequence correlations after the max pooling. The last layer is a dense layer, mapping the hidden deep information of BiLSTM outputs to a smaller number of neurons, whose values are used to parameterise distribution generating final outputs.

9.3.2 Handling of Uncertainties with Bayesian Neural Network

In a Bayesian Neural Network (BNN) model, parameters are distributions, instead of deterministic values, allowing to quantify and reduce uncertainties in the modelled process. Since the parameters are uncertain, the feedforward results of the BNN are also uncertain and can be quantified through the following two steps: 1) sample the parameters with respect to their distributions, and 2) use these sampled parameters to determine the feedforward values, as in a classical NN model. As the outputs of BNN are distributions, output uncertainties can be evaluated after learning distribution

parameters. For example, if Gaussian distribution is assumed for the univariate outputs, BNN will give its location and scale parameters.

In backpropagation, BNN updates parameter distributions for given observation data by Equation (9.1). The likelihood, $p(D|w)$, is easy to compute, but the denominator, $\int p(D|w')p(w') dw'$, which is a normalisation factor and equal to $p(D)$, may be difficult to calculate if number of parameters is high. Ref. [236] proposes a backpropagation scheme using variational Bayes and reparameterisation approach, while there are also other techniques to represent uncertainties, such as “dropout” [237] and “flipout” [238] methods.

The true posterior distribution of a BNN parameter, $p(w|D)$, is often intractable. The main idea of variational Bayes is to find another suitable distribution with explicit functional form to approximate it, which is known as the variational posterior. Supposing the variational posterior is parametrised by θ , i.e., $q(w|\theta)$, the objective is to minimize the Kullback-Leibler (KL) divergence between it and the true posterior distribution [239]:

$$KL(q(w|\theta)|p(w|D)) = \int q(w|\theta) \log \left(\frac{q(w|\theta)}{p(w|D)} \right) dw \quad (9.2)$$

where: $KL(\cdot)$ is KL divergence. However, direct optimisation of Equation (9.2) is intractable, since the true posterior $p(w|D)$ is involved. Therefore, it is further derived applying Bayesian formula:

$$\int q(w|\theta) \log \left(\frac{q(w|\theta)}{p(w|D)} \right) dw = \int q(w|\theta) \log \left(\frac{q(w|\theta)p(D)}{p(D|w)p(w)} \right) dw \quad (9.3)$$

where: $p(D)$ is the “evidence”, assumed to be a constant independent of w . Accordingly, the right-hand side of Equation (9.3) is:

$$\log(p(D)) + KL(q(w|\theta)|p(w)) - \mathbb{E}_{q(w|\theta)}[\log(p(D|w))] \quad (9.4)$$

where: $\mathbb{E}_{q(w|\theta)}[\log(p(D|w))]$ is the expectation of $p(D|w)$ for $w \sim q(w|\theta)$. Minimising Equation (9.4) is equivalent to the minimisation without the first constant evidence term, and the loss function $L(\theta|D)$ can be defined as:

$$L(\theta|D) = KL(q(w|\theta)|p(w)) - \mathbb{E}_{q(w|\theta)}[\log(p(D|w))] \quad (9.5)$$

where: $\mathbb{E}_{q(w|\theta)}[\log(p(D|w))]$ depends on data and is known as the “likelihood cost”, while $KL(q(w|\theta)|p(w))$ is a prior-dependent term representing “complexity cost” [236]. Therefore, the optimisation is trade-off between the expectation of likelihood to the data and the prior and variational posterior difference.

The $L(\theta|D)$ term can be estimated using Monte Carlo:

$$\sum_{i=1}^n \left[\log(q(w^{(i)}|\theta)) - \log(p(D|w^{(i)})) + \log(p(w^{(i)})) \right] \quad (9.6)$$

where: $w^{(i)}$ is the i -th sample model parameter drawn from the variational posterior $q(w|\theta)$. The parameter of interest, θ , should be updated during the backpropagation, but taking derivatives with respect to θ is difficult, since the expectation also depends on θ . The reparameterisation approach presented in [236] can deal with that, proving that the derivative of an expectation can be expressed as the expectation of a derivative in some cases. Assuming $q(w|\theta)$ to be a Gaussian, so that $q(w|\theta) \sim N(\mu, \sigma)$ and $\theta = (\mu, \sigma)$, giving $w^{(i)} = \mu + \sigma \varepsilon^{(i)}$, $\varepsilon \sim N(0, 1)$. Let $f(w^{(i)}|\mu, \sigma) = \log(q(w^{(i)}|\theta)) - \log(p(D|w^{(i)}))$, then $\frac{\partial L(\theta|D)}{\partial \theta}$ can be estimated by $\frac{\partial L(\mu, \sigma|D)}{\partial \mu}$ and $\frac{\partial L(\mu, \sigma|D)}{\partial \sigma}$:

$$\frac{\partial L(\mu, \sigma|D)}{\partial \mu} \approx \sum_{i=1}^n \left[\frac{\partial f(w^{(i)}|\mu, \sigma)}{\partial w^{(i)}} + \frac{\partial f(w^{(i)}|\mu, \sigma)}{\partial \mu} \right] \quad (9.7)$$

$$\frac{\partial L(\mu, \sigma|D)}{\partial \sigma} \approx \sum_{i=1}^n \left[\frac{\partial f(w^{(i)}|\mu, \sigma)}{\partial w^{(i)}} + \frac{\partial f(w^{(i)}|\mu, \sigma)}{\partial \sigma} \right] \quad (9.8)$$

The same term $\frac{\partial f(w^{(i)}|\mu, \sigma)}{\partial w^{(i)}}$ is shared by both the gradients for the mean μ in Equation (9.7) and standard deviation σ in Equation (9.8), which can also be found in the backpropagation algorithm of deterministic neural network. The posterior parameters are iteratively updated with $\mu - \gamma \frac{\partial L(\mu, \sigma|D)}{\partial \mu}$ and $\sigma - \gamma \frac{\partial L(\mu, \sigma|D)}{\partial \sigma}$, respectively, using gradient-based algorithm (e.g., Adam optimiser from [120], which is used in this chapter), where γ is the step size.

9.3.3 Bayesian CNN-BiLSTM Model

To extend the deterministic CNN-BiLSTM model into a BNN, all parameters in 1D Convolution layer and Bidirectional LSTM layer are assigned with Gaussian prior and Gaussian posterior distributions. The max pooling layer does not need updating, since there is no trainable parameter. For the dense layer, the mixture of two Gaussian distributions (one with a small scale value, 1, and another with a large scale value, 10), known as “spike and slab”, are used to set priors and posteriors for improved sampling efficiency [240]. As the mixture requires to infer additional parameters, it is only used in the dense layer, which is directly connected to the final outputs. The selection of suitable priors’ in BNN is subject of active research and explicit meanings of the selections are still not fully clear [10].

The outputs of Bayesian CNN-BiLSTM model are parameterised distributions, whose exact form depends on the specific application. In this chapter, two instances of Bayesian CNN-BiLSTM are implemented in the following case studies.

9.3.4 Example Case Study 9.1: Forecasting of Weather Variables

The fundamental equation of the power output of a WT is:

$$P_{\text{out}} = 1/2 \times AD \times A \times WS^3 \times C_p \quad (9.9)$$

where: AD is air density, A is the WT rotor swept area, WS is wind speed and C_p is the WT power coefficient [241]. While WS is directly included in Equation (9.9), the importance of considering WD is evidenced in Chapter 8, and [16], with [16] emphasising that ambient temperature also has the strongest impact on air density and thus also influences the power output. Refs [15] and [12] show that AD can be determined from temperature, relative humidity and barometric pressure, so this chapter firstly predicts the weather variables (WS, AD and WD), as they will have strong impact on forecasted P_{out} of the WF (made of WTs with specific characteristics, e.g., values of A and C_p).

Before the calculation, WS, AD and WD are transformed using normal quantile transform: the truncated data for WS and AD (as they cannot be negative) and circular data for WD are approximated by simple normal distributions [242]:

$$x_{\text{new}} = \Phi^{-1}(F(x)) \quad (9.10)$$

where: x is the original data, and x_{new} is the transformed data; $F(\cdot)$ is the cumulative distribution function, and in the presented analysis, empirical CDF is used. $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard normal distribution.

To forecast the next hour's weather variables of the considered WF, the WS, AD and WD recordings from the past seven days, i.e., 168 hours, for both the individual considered WF and together with the synchronous weather variables from a nearby WF in the same cluster, are used as the model inputs. The inputs of Bayesian CNN-BiLSTM model can be written as:

$$\left[WS_{-167:0}^{(WF_1)}, WS_{-167:0}^{(WF_2)}, AD_{-167:0}^{(WF_1)}, AD_{-167:0}^{(WF_2)}, WD_{-167:0}^{(WF_1)}, WD_{-167:0}^{(WF_2)} \right] \quad (9.11)$$

where: $WS_{-167:0}^{(WF_1)}$ and $WS_{-167:0}^{(WF_2)}$ are wind speed, WS, recordings for the two WFs in the same cluster in the past 167 hours and in the current moment ($t = 0$), while AD and WD symbols denote similarly air density and wind direction inputs, respectively.

Model outputs are mixture distributions of weather variables in the next hour, consisting of multiple components, which are trivariate normal distributions with a full covariance. The full covariance can capture correlations in weather variables, so forecasted WS, AD, WD are all correlated (mixture distribution captures multiple modals better than single distribution as shown in Chapter 7).

Supposing the Bayesian CNN-BiLSTM has been trained on the dataset, D , which optimises the variational posterior $q(w|\theta)$, its output for given new observation x can be formulated as:

$$p(y|x, D) = p(y|x, \theta) = \int p(y|x, w)q(w|\theta)dw \quad (9.12)$$

where: y is the dependent variable, with distribution $p(y|x, D)$. In the feedforward, the NN parameter w can be first determined by sampling from the learned $q(w|\theta)$, where its uncertainty is denoted as epistemic uncertainty [37]. Afterwards, for the model determined by the known sampled parameter w_0 , uncertainty of dependent variable is characterised only by its distribution, $p(y|x, w_0)$, denoted as aleatoric uncertainty [37].

Fig. 9.3 shows “test week” forecasting results of weather variables for Bruska WF. The shaded areas represent the uncertainty, while “Pred.” (“prediction”) denotes forecasted mean values.

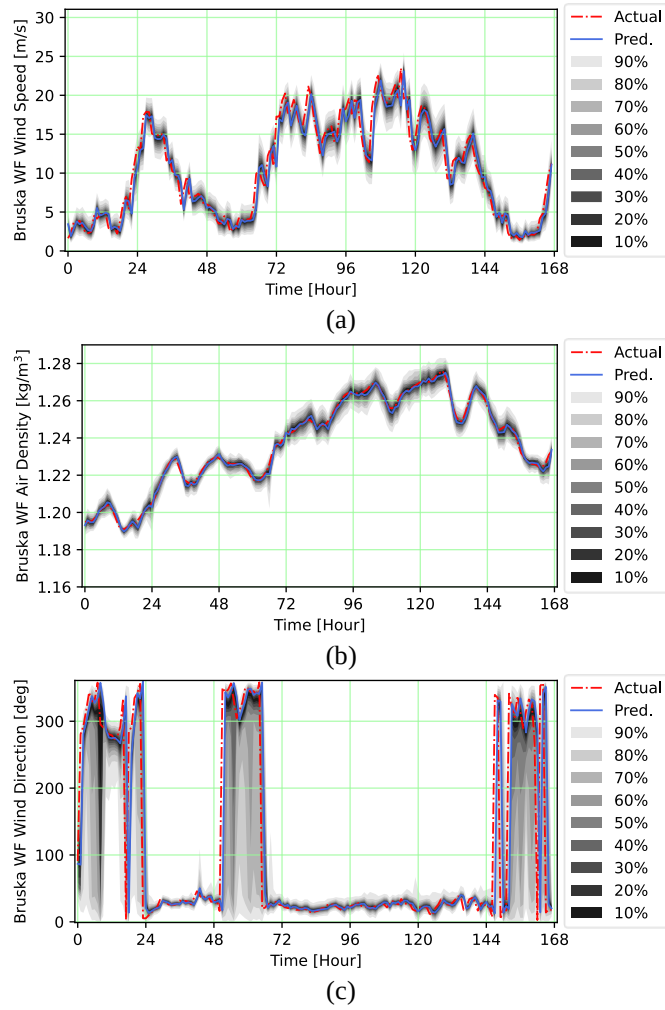


Fig. 9.3. The results of Bruska WF forecasting weather variables in “test week”: (a) WS, (b) AD, (c) WD (zero degree corresponds to a true North direction).

In Fig. 9.3, 3,000 samples of parameter w are taken first for each time stamp. Then, 3,000 samples of y are taken from output distributions of each of the NNs, representing previously discussed epistemic and aleatoric uncertainties. The mean values of

samples of y are plotted as “Pred.” Shaded areas mark various uncertainty ranges and coverage ratios (estimated from the results of samples). In Fig. 9.3(c), the shaded areas (uncertainty ranges) are visually large at some time periods, e.g., between 50 and 70 hours, which is because WD is inherently a circular variable. However, the time series plot can only represent linear variables. Therefore, even though the forecasting has a very small uncertainty, e.g., 355 – 5 degrees (only 10 degrees difference), the plot will visually have a shaded area from 5 to 355 degrees. If the forecasted uncertainty range does not cover 360 degrees (0 degrees), e.g., between 72 and 144 hours, then the plot shows the small uncertainties like being a linear variable. Table 9.3 lists the forecasting errors for both Bruska and Jelinak WFs in “test week”. Both Fig. 9.3 and Table 9.3 show that model accurately captures mean values of weather variables (MAE and RMSE are small in terms of the scale of the data). Uncertainty ranges are reasonably small, except when WD is around 0 (360) degrees, as WD is a circular variable, for which mean and percentile values are calculated following [243].

Table 9.3 Error of Forecasting Weather Variables in “Test Week”

Forecasted WS	MAE [m/s]	RMSE [m/s]	Pinball Loss [m/s]	CRPS [m/s]	Winkler Score [m/s]
Bruska	1.288	1.789	0.470	0.939	7.625
Jelinak	1.274	1.771	0.460	0.920	7.490
Forecasted AD	MAE [kg/m ³]	RMSE [kg/m ³]	Pinball Loss [kg/m ³]	CRPS [kg/m ³]	Winkler Score [kg/m ³]
Bruska	0.001	0.002	0.001	0.001	0.015
Jelinak	0.001	0.001	0.000	0.001	0.006
Forecasted WD ¹	MAE	RMSE	Pinball Loss	CRPS	Winkler Score
Bruska	0.114	0.212	0.043	0.086	0.744
Jelinak	0.167	0.281	0.063	0.125	1.214

¹ The WD errors are calculated by averaging the errors of $\cos(\text{WD})$ and $\sin(\text{WD})$ separately, and are therefore unitless.

9.3.5 Example Case Study 9.2: Forecasting of WF Operating Regimes

The same Bayesian CNN-BiLSTM model is also used for forecasting WF ORs, with only difference in the output distribution. Since the WF OR is a discrete variable, categorical distribution is used as the model output. The inputs of the model are the WS, WD, AD, P_{out} and (known or inferred) OR values of the individual WF in the past 168 hours, while the output and the uncertainty of the next hour’s WF OR are obtained using the same sampling method from trained NN as previously discussed. Fig. 9.4 shows forecasting results, where “Pred.” is the mode of the categorical distribution,

with the darker shaded areas representing higher confidence levels. Fig. 9.4 shows that “Pred.” generally follows the trends of the actual OR. When OR time series is relatively constant, the prediction confidence is higher and “Pred.” is more accurate. However, if OR changes rapidly, more dispersed predictions are assigned to adjacent ORs.

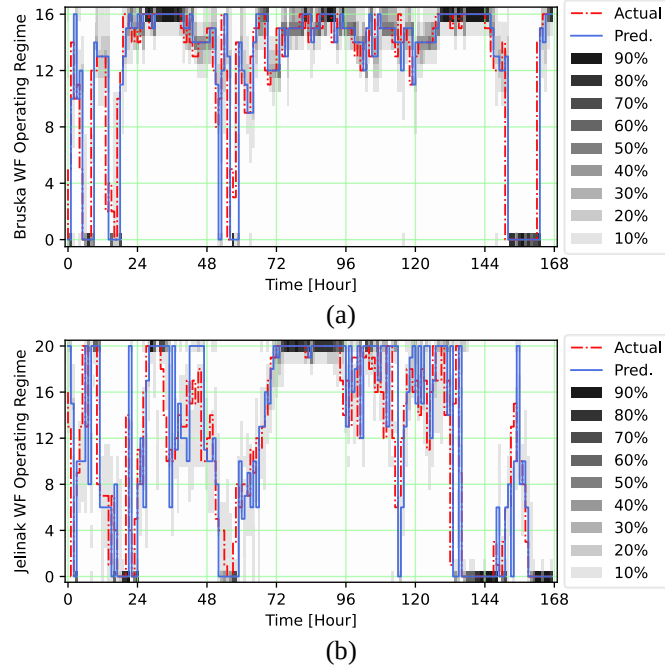


Fig. 9.4. The results of forecasting WF ORs in “test week”: (a) Bruska WF, (b) Jelinak WF (Y-axis represents the number of WTs in normal operation).

For evaluating model accuracy, cross-entropy H is used:

$$H(p, q) = - \sum_{x=0}^{X-1} p(x) \log q(x) \quad (9.13)$$

where for a single test: x represents the class label, which is a discrete integer number from 0 to $X - 1$, for the total of X different classes; $p(x)$ is equal to 1 for the label, if it has been observed, otherwise it is 0; $q(x)$ is the probability predicted by the model for label x . Based on information theory, the cross-entropy measures the difference between p and q distributions by calculating number of bits needed for their identification and encoding: lower values of H denote higher similarity. For the forecasted WF OR results for Bruska and Jelinak WFs, the calculated cross-entropies are 1.59 and 2.07, respectively, while $\pm 20\%$ accuracies (i.e., the percentages of predictions which are within $\pm 20\%$ of actual ORs) are 83 % and 69 %, respectively.

9.4 Vine-GMCM for Correlation-Based Modelling of WF Power Outputs

The correlations between WF's P_{out} , WS, WD and AD are next modelled through a Copula-based method, where the key is to estimate the conditional probability of P_{out} given all other variables, i.e., $p(x_1|x_2, x_3, x_4)$, where x_1 is P_{out} , x_2 is WS, x_3 is AD, x_4 is WD, and $p(\cdot|\cdot)$ is the conditional probability density. After fitting the conditional Copula model, both the mean values and probabilistic percentile values of WF P_{out} could be estimated by sampling from the conditional probability density. A similar Vine-GMCM as in Section 8.3 will be implemented, but it will be extended to consider the additional predictor variable, air density.

9.4.1 Proposed 4-D Correlation Model

In modelling WF P_{out} correlations, the conditional probability density function of P_{out} is $f_{1|2,3,4}(x_1|x_2, x_3, x_4)$, and it is equal to $\frac{f_{1,2,3,4}(x_1, x_2, x_3, x_4)}{f_{2,3,4}(x_2, x_3, x_4)}$. The PDF $f_{1|2,3,4}(x_1|x_2, x_3, x_4)$ can be calculated using the chain rule and Vine Copula theory:

$$\begin{aligned} & f_{1|2,3,4}(x_1, x_2, x_3) \\ &= \frac{f_{1,2,3,4}(x_1, x_2, x_3)}{f_{2,3,4}(x_2, x_3, x_4)} \\ &= f_1(x_1) \frac{\prod_{i=2}^4 [c_{1,i}(F_1, F_i)] \prod_{j=3}^4 [c_{2,j|1}(F_{2|1}, F_{j|1})] c_{3,4|1,2}(F_{3|1,2}, F_{4|1,2})}{\prod_{k=3}^4 [c_{2,k}(F_2, F_k)] c_{3,4|2}(F_{3|2}, F_{4|2})} \end{aligned} \quad (9.14)$$

where: $c_{2,3|1}(F_{2|1}, F_{3|1})$, $c_{2,4|1}(F_{2|1}, F_{4|1})$, $c_{3,4|2}(F_{3|2}, F_{4|2})$ and $c_{3,4|1,2}(F_{3|1,2}, F_{4|1,2})$ are conditional Copula density functions, which can be modelled by their arguments, i.e., $F_{2|1}$, $F_{3|1}$, $F_{4|1}$, $F_{3|2}$, $F_{4|2}$, $F_{3|1,2}$, and $F_{4|1,2}$, respectively. More specifically, these conditional distribution functions can be derived by taking derivatives from the joint distributions, i.e., Copula distribution functions, which are $F_{2|1} = \frac{\partial C_{1,2}(F_1, F_2)}{\partial F_1}$, $F_{3|1} = \frac{\partial C_{1,3}(F_1, F_3)}{\partial F_1}$, $F_{4|1} = \frac{\partial C_{1,4}(F_1, F_4)}{\partial F_1}$, $F_{3|2} = \frac{\partial C_{2,3}(F_2, F_3)}{\partial F_2}$, $F_{4|2} = \frac{\partial C_{2,4}(F_2, F_4)}{\partial F_2}$, $F_{3|1,2} = \frac{\partial C_{2,3|1}(F_{2|1}, F_{3|1})}{\partial F_{2|1}}$, and $F_{4|1,2} = \frac{\partial C_{2,4|1}(F_{2|1}, F_{4|1})}{\partial F_{2|1}}$, where calculations could be implemented recursively. From Equation (9.14), it can be seen that the whole correlation structure is made from the six Copula pairs: $c_{1,2}$, $c_{1,3}$, $c_{1,4}$, $c_{2,3|1}$, $c_{2,4|1}$, and $c_{3,4|1,2}$, which are all fitted using bivariate GMCM following the optimisation algorithm in [199] with

structure depicted by the Vine Copula graph in Fig. 9.5 (the solid, dashed and dotted edges between nodes represent the first, second and layers of correlations, respectively). The combination of GMCM and Vine Copula is the Vine-GMCM model used in this chapter.

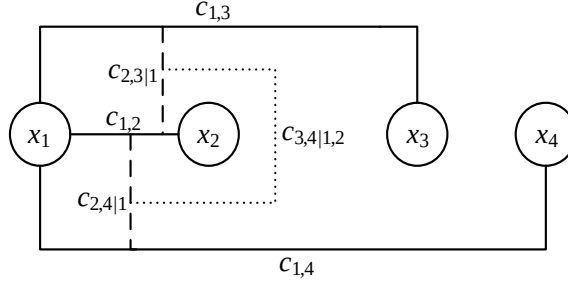


Fig. 9.5. Implemented 4-D Vine Copula structure.

After fitting all GMCM pairs in the Vine Copula structure, $f_{1|2,3,4}(x_1|x_2, x_3, x_4)$ can be calculated straightforwardly for given deterministic inputs of x_2, x_3 and x_4 . In the case of uncertain values of x_2, x_3 and x_4 , e.g., $x_2' \leq x_2 \leq x_2''$, $x_3' \leq x_3 \leq x_3''$ and $x_4' \leq x_4 \leq x_4''$, the conditional probability density is equal to:

$$\frac{\int_{x_4'}^{x_4''} \int_{x_3'}^{x_3''} \int_{x_2'}^{x_2''} f_{2,3,4}(y,z,k) f_{1|2,3,4}(x_1|y,z,k) dy dz dk}{\int_{x_4'}^{x_4''} \int_{x_3'}^{x_3''} \int_{x_2'}^{x_2''} f_{2,3,4}(y,z,k) dy dz dk} = \frac{\mathbb{E}_{f_{2,3,4}(y,z,k)} f_{1|2,3,4}(x_1|y,z,k)}{\mathbb{E}_{f_{2,3,4}(y,z,k)}}; y, z, k \in [x_2', x_2''], [x_3', x_3''], [x_4', x_4''] \quad (9.15)$$

The integrals, i.e., the expectations $\mathbb{E}(\cdot)$, are non-trivial to solve analytically and Monte Carlo integration is used for their numerical estimations in this chapter, similar to Section 8.3.

9.4.2 Example Case Study 9.3: Modelling of WF Power Output Correlations

The presented Vine-GMCM model is also applied to Bruska and Jelinak WFs, to illustrate estimation of the conditional probability of P_{out} on the model inputs (synchronous WS, AD and WD data), while the output is distribution of WF P_{out} values. Initially, measurements for considered WF in the fully operating regime are used to fit the model. After obtaining P_{out} distribution for fully operational regime, it is

transformed in the Vine-GMCM model using linear scaling to obtain the final P_{out} estimation, but now considering WF ORs.

Unlike the proposed Bayesian CNN-BiLSTM model, this Copula model is designed to capture synchronous correlations among all dimensions. Accordingly, model performance is first evaluated assuming that input values of WS, AD, WD and OR are known for the test dataset (actual measured values), with “test week” results presented in Fig. 9.6 and errors in Table 9.4. The use of the forecasted WS, AD, WD and, particularly, impact of using forecasted OR values are discussed in the next section. The results in Fig. 9.6 show that predicted WF P_{out} values are almost the same as the actual recordings, and related uncertainty ranges are also very small. The error metrics in Table 9.4 also confirm very good deterministic and probabilistic performance. The reason is that this is a hindcasting model, with known actual WS, AD, WD and WF OR as inputs, so the Vine-GMCM can correctly give the conditional distribution of WF P_{out} .

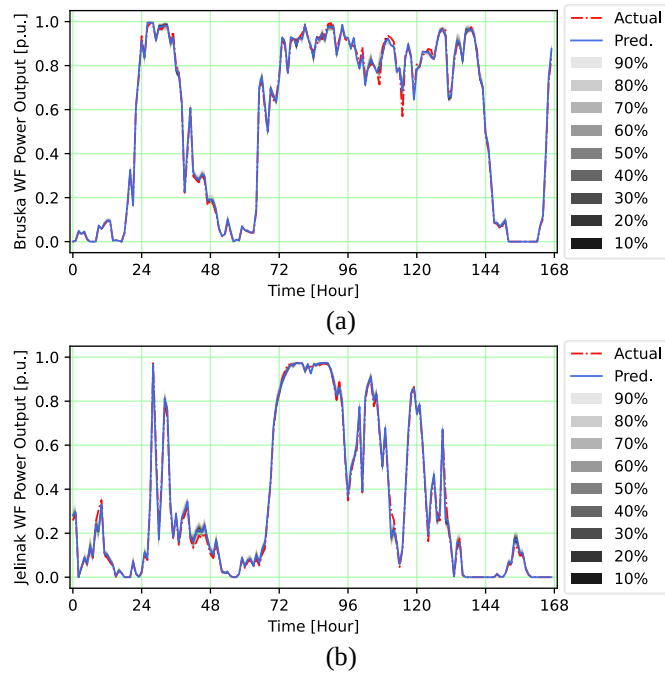


Fig. 9.6. Conditional distribution of WF P_{out} given actual measurements of WS, AD, WD and known ORs in “test week”: (a) Bruska WF, (b) Jelinak WF.

Table 9.4 Errors of WF P_{out} Estimations of WS, AD, WD and Operating Regimes by Vine-GMCM in “Test Week”

Modelling P_{out}	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (MWh)	E_o (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)
Bruska	0.012	0.018	0.004	0.009	0.070	41.721	2.427	30.736	1.991	10.986	0.337
Jelinak	0.014	0.022	0.005	0.011	0.075	47.105	4.290	22.366	4.221	24.738	1.520

Table 9.5 Comparison of Results with and without Considering Wind Farm Operating Regimes in “Test Week”

Forecasting Bruska P_{out}	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (MWh)	E_o (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)
Full OR	0.081	0.119	0.031	0.061	0.532	342.173	16.155	144.877	12.655	197.296	6.047
Known OR	0.059	0.108	0.021	0.042	0.367	135.350	14.161	221.218	9.589	-85.868	-2.632
Forecasted OR	0.070	0.113	0.025	0.049	0.386	156.055	14.227	265.534	12.259	-109.479	-3.355
Forecasting Jelinak P_{out}	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (MWh)	E_o (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)
Full OR	0.116	0.154	0.039	0.078	0.561	481.999	58.026	101.894	12.783	380.105	23.351
Known OR	0.060	0.108	0.021	0.042	0.279	137.829	29.545	164.282	14.147	-26.454	-1.625
Forecasted OR	0.080	0.126	0.028	0.056	0.406	175.947	38.659	226.850	19.345	-50.903	-3.127

9.5 Combining Bayesian CNN-BiLSTM and Vine GMCM for Wind Farm Output Forecasting

The correlational Vine-GMCM model uses samples of hour-ahead forecasted WS, AD and WD values from Bayesian CNN-BiLSTM as the inputs, to output the conditional probability of P_{out} on WS, AD and WD, assuming initially that the WF will be in fully operational state in the next hour. This combined model propagates uncertainties from Bayesian CNN-BiLSTM stage to the Vine-GMCM, which inherently treats them as the parametric variability and then correlates all dimensions for every individual sample. Afterwards, the WF OR forecasting results are used as additional input to scale down the initial P_{out} values (set for fully operational WF). Finally, the mean value and percentiles of the hour-ahead WF P_{out} probabilistic forecasting are estimated from the transformed samples, as in the case of weather variable forecasting. Fig. 9.7 shows flowchart of combined forecasting model, while Fig. 9.8 shows results for Bruska and Jelinak WFs P_{out} forecasting in “test week”. The CNN-BiLSTM can model both epistemic and aleatoric uncertainties, because the model parameters and outputs are distributions, instead of deterministic values (although model inputs can be deterministic or probabilistic). The Vine-GMCM can capture only aleatoric uncertainty, since the model parameters are deterministic values and are optimised in conventional non-Bayesian manner [199], although it can provide probabilistic outputs for both deterministic and probabilistic inputs. Nevertheless, both models are probabilistic models.

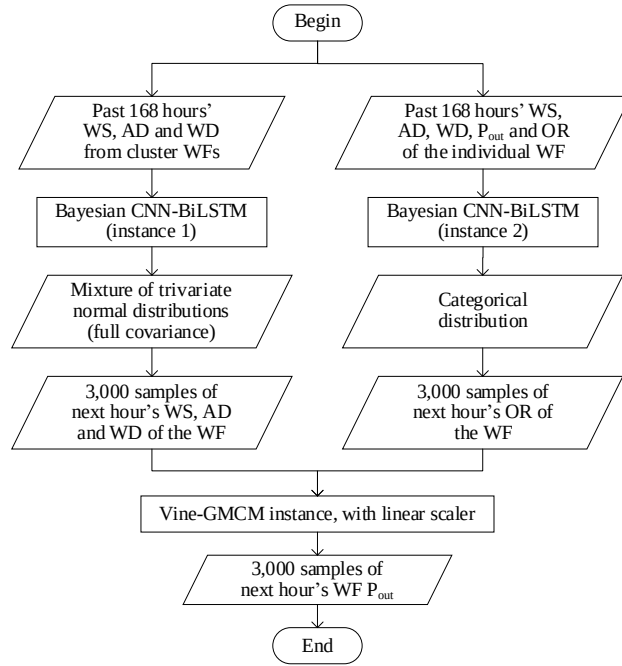


Fig. 9.7. Flowchart of Bayesian CNN-BiLSTM and Vine-GMCM framework.

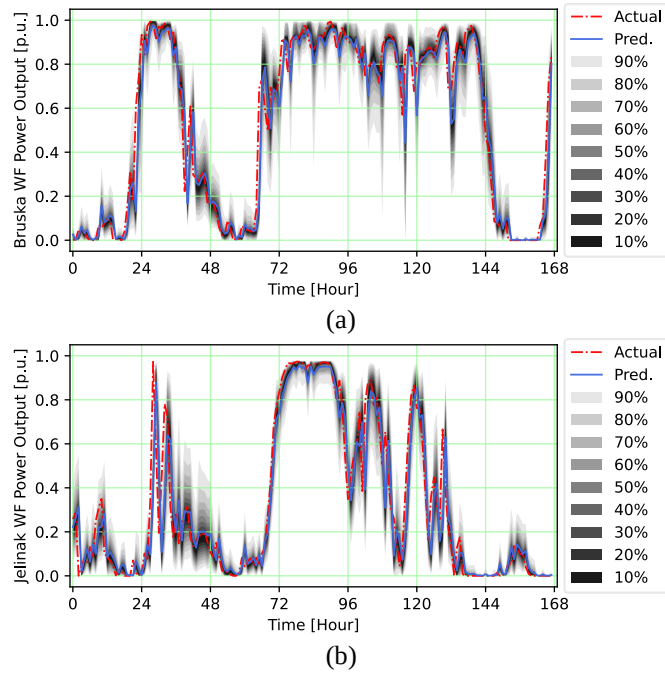


Fig. 9.8. The results of forecasting of WF P_{out} by combined Bayesian CNN-BiLSTM and Vine-GMCM model in “test week”: (a) Bruska WF, (b) Jelinak WF.

In Fig. 9.8, P_{out} forecasting results for both WFs show that the mean values of predictions are close to the actual realisations. The model correctly outputs higher uncertainty ranges when the wind speed is near the cut-out WS. For Bruska WF at $t = 115$ hours, the 90 % coverage ratio of the forecasting has a low boundary near 0 p.u.,

which is because the WF average WS is 23.47 m/s, when individual WTs may cut-out, depending on implemented WT cut-out hysteresis control (e.g., if the instant wind gust, or a rolling averaged WS are higher than related thresholds), resulting in higher uncertainty than for other WS conditions.

9.5.1 Example Case Study 9.4: Incorporating WF ORs in Forecasting

To assess the importance of considering WF ORs, three different cases are compared:

- 1) The P_{out} of WF is forecasted for fully operational WF (all WTs in normal operation, linear scaling in Fig. 9.7 is not applied);
- 2) Information on actual WF ORs is used for forecasting P_{out} of WF with linear scaling stage, i.e., similarly to planned outages, unplanned outages are known in advance;
- 3) The forecasted WF P_{out} values are obtained with predicted WF ORs, as previously discussed.

The errors/differences between the actual and forecasted values for all three cases are listed in

Table 9.5. The unplanned outages are not known in advance, and “Known OR” results are used as a reference dataset for “Forecasted OR” and “Full OR” results. Both “Known OR” and “Forecasted OR” cases provide significantly smaller deterministic and probabilistic error metrics than “Full OR”, confirming benefits of OR-based forecasting: noticeably increased accuracy of mean values and reduced uncertainty. In terms of the energy based metrics, these results confirm that assuming a fully operational state of WF will result in the overestimation of WF outputs, which is more problematic for WF operators and generation dispatching than underestimation: higher than predicted outputs can be easily reduced, while lower than predicted ones have to be provided by a more expensive reserve generation. The used “linear scaling (plus bias) relationship” is simple and easy to integrate with the forecasting model.

Table 9.5 confirms importance of OR predictor variable for improving forecasting results, despite the fact that the WF OR in this chapter is simply represented by

estimating number of WTs in normal operation. It should be highlighted that the TSOs should use the data and information on the operational states of wind turbines in the wind farm due to the observed benefits.

Fig. 9.9 shows the reliability diagram [244], illustrating the relationship between prediction interval nominal confidence (PINC) and average coverage error (ACE) in “test week”. The PINC is defined by boundaries of the forecasted Prediction Intervals (PIs), while ACE gives difference between prediction interval coverage probability (PICP) and PINC. The PICP is calculated as $\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} c_i \times 100$, where N_{test} is the number of test samples, and c_i is equal to 1 if the i -th actual measurement is within the prediction interval, otherwise 0. For both Bruska and Jelinak WFs, if the WF OR is known, the model is precise because the average ACE values are only -2.6 % and -1.5 %, respectively. Assuming full OR will lead to much worse mean values of ACE (-12.1 % for Bruska and -14.5 % for Jelinak). In the case of forecasted OR, the model underestimates PINC for Jelinak WF (ACE mean value of -5.2 %), while it is very precise for Bruska WF (ACE mean value of -1.1 % and all values close to zero).

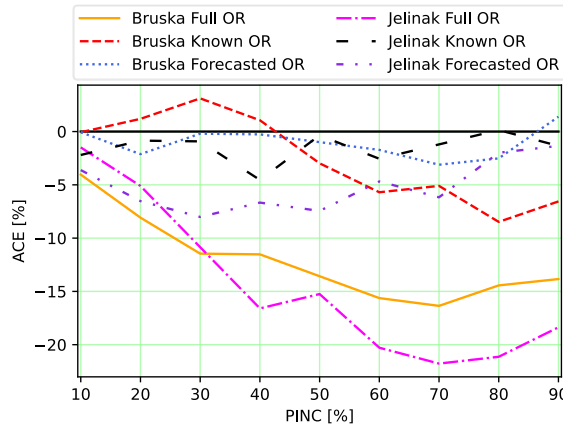


Fig. 9.9. Reliability diagram with and without considering wind farm operating regimes in “test week”.

All calculations are done on a desktop PC running Microsoft Windows 10 21H1 operating system, with AMD Ryzen 1800X CPU, Nvidia GTX 1080ti GPU and 32GB DDR4 RAM. The Bayesian CNN-BiLSTM model is implemented using Python 3.8.5, TensorFlow-probability 0.12.1, and Edward2 0.0.2. The Vine-GMCM model is built by mixture programming with Python 3.8.5, NumPy 1.18.5 and MATLAB R2020b. In the training phase, one Bayesian CNN-BiLSTM requires about 18 – 24 hours to

train, and one Vine-GMCM takes only around 20 – 30 minutes to fit. However, in the test phase, one Bayesian CNN-BiLSTM needs only 1–2 seconds to obtain prediction for one time stamp (drawing 3,000 samples). On average, assuming only one call for MATLAB Engine API from Python in one forecasting week, which is practically slow, one Vine-GMCM requires 9–10 seconds to estimate the final uncertainty range for one time stamp (if the CNN-BiLSTM gives 3,000 samples for each time stamp).

9.5.2 Model Comparison and Benchmarking

Four additional models are used for benchmarking with the combined Bayesian CNN-BiLSTM and Vine-GMCM model, while the next section presents the benchmarking on a day-ahead WF P_{out} forecasting dataset, GEFCom2014 [38], with a reduced modified Bayesian CNN-BiLSTM model from this chapter.

- 1) *Sequence-to-Sequence (Seq2Seq) Model*: This model is based on the original model developed for the Croatian transmission system operator, providing up to 72-hour ahead hourly deterministic forecasting of power outputs of considered WFs. It uses available historical weather forecasts from the weather research and forecasting model (WRF) numerical weather prediction (NWP) model [245] (providing 72-hour ahead forecasted WS and WD for considered locations), the reported available WF capacity (information on WF planned outages) and realised production from WF SCADA monitoring system. Model is based on BiLSTM algorithm, as previously described.

The original model maps weather forecast sequences and reported WF capacity to the sequences of WF power outputs, for the next 72 hours (longer-term) forecasting. However, to ensure direct comparison and benchmarking with the CNN-BiLSTM and Vine-GMCM model, the original model is modified to use an hour-ahead weather forecasts (WS, WD, and AD) and to combine them in the long-term forecasts for training and testing over the designated periods. Afterwards, the WF P_{out} data were refined and averaged to hourly resolution and prepared as overlapping 72-hour sequences with an offset of one hour and finally merged with weather forecasts on the same timestamps. For benchmarking purposes, only the 1st hour of each 72-hour long prediction sequence was used, while the remaining 71 hours were neglected. This was repeated and forecasts were calculated 168

times, consecutively rolling forward for every hour in the specified test period of one week. The WF ORs were added to model as an additional predictor.

- 2) *Markov chain based model*: This model is developed by extending the model from Chapter 8, which considered only two weather variables: WS and WD. When individual WF is considered, a trivariate MC model of weather variables (MCM-WV) as the inputs is developed. The input data, D , belonging to the training period are classified in $N_{WS} = 13$, $N_{AD} = 4$ and $N_{WD} = 4$ equally spaced classes for WS, AD, and WD, respectively. Afterwards, $\overline{WW}^{WF}$ is defined as a trivariate random process associated to a vectorial state variable:

$$\overline{WW}^{WF} = [WS, WD, AD] \quad (9.16)$$

It is possible to obtain a univariate equivalent stochastic process WW^{WF} , discretised into N_{WS} , N_{AD} and N_{WD} states, as in Chapter 8. The one step ahead (hour-ahead) output in the form of probability mass function (PMF) of the weather variables obtained by MCM-WV, $\hat{f}_{W^{WF}}(t_{h+1})$, is then used to obtain an hour-ahead prediction point (mean) and interval forecast (5th-95th percentiles interval) of the total WF output power (P_{out}^{WF}) using the empirical (or experimental) conditional PMF ($P_{out}^{WF}|WW^{WF}$), calculated for the same training set D .

The PMF of forecasted WF output at time t_{h+1} , given WW^{WF} , $\hat{f}_{p_{out_l}^{WF}}(t_{h+1})$ is:

$$\begin{aligned} & \hat{f}_{p_{out_l}^{WF}}(t_{h+1}) \\ &= p[P_{out}^{WF}(t_{h+1}) = p_{out_l}^{WF}|WW^{WF}] \\ &= \sum_{j=1}^N \hat{f}_{ww_j^{WF}}(t_{h+1}) \cdot p\{P_{out}^{WF}(t_{h+1}) = p_{out_l}^{WF}|WW^{WF}(t_{h+1}) = ww_j^{WF}\} \quad (9.17) \end{aligned}$$

From $\hat{f}_{p_{out_l}^{WF}}(t_{h+1})$, the wind power mean/expected values and percentiles are calculated as:

$$\hat{P}_{out}^{WF}(t_{h+1})_{mean} = \sum_{l=1}^N P_{out_l}^{WF} p[P_{out}^{WF}(t_{h+1}) = p_{out_l}^{WF}|WW] \quad (9.18)$$

$$\hat{P}_{out}^{WF}(t_{h+k})_{\alpha} = p_{out_j^{WF}}: \begin{cases} \sum_{l=1}^j \hat{f}_{p_{out_l}^{WF}}(t_{h+1}) > \alpha \\ \sum_{l=1}^{j-1} \hat{f}_{p_{out_l}^{WF}}(t_{h+1}) < \alpha \end{cases} \quad (9.19)$$

where: $\alpha=0.05$ or 0.95 for the 5^{th} and 95^{th} quantiles.

In order to take into account the WF ORs, a different MC model of the variations $\Delta P_{\text{out}}^{\text{WF}}$ of the WF P_{out} value is built, using the available operating WTs in the dataset. Assuming:

$$P_{\text{out regime}}^{\text{WF}} = P_{\text{out}}^{\text{WF}} + \Delta P_{\text{out}}^{\text{WF}} \quad (9.20)$$

and that $P_{\text{out}}^{\text{WF}}$ and $\Delta P_{\text{out}}^{\text{WF}}$ are independent random variables, the hour-ahead PMF of $P_{\text{out regime}}^{\text{WF}}$ is obtained as a convolution of PMFs of $P_{\text{out}}^{\text{WF}}$ and corresponding $\Delta P_{\text{out}}^{\text{WF}}$.

In cases when WS, AD and WD from another WF in the same cluster are also considered, a mixture of the two transition probabilities matrices obtained for the separate sets of data for the two WFs is calculated according to [101]. If WF_1 and WF_2 belong to the same cluster, the PMF of WF_i , at time t_{h+1} , is:

$$\hat{f}_{WW^{\text{WF}_i}}(t_{h+1}) = \lambda \hat{f}_{WW^{\text{WF}_1}}(t_{h+1}) + (1 - \lambda) \hat{f}_{WW^{\text{WF}_2}}(t_{h+1}) \quad (9.21)$$

where: $i = 1, 2$, according to the considered WF and parameter λ is obtained by optimisation, with the constraint $\lambda \geq 0$.

Once the PMFs of the weather predictions of the considered WF are correlated to the weather predictions of another WF from the same cluster, the discussed steps could be repeated to obtain the point and interval predictions of the WF power outputs with the WF OR information included.

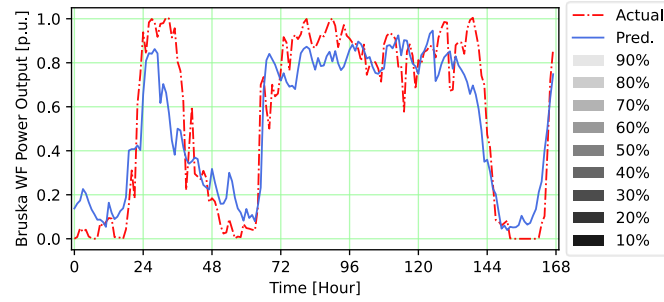
- 3) *Persistence model*: This model assumes the WF P_{out} value at time t_{h+1} will be the same as at time t_h . It does not consider any weather variable or WF OR, and approaches forecasting of power output time series as an auto-regression problem.
- 4) *Bayesian CNN-BiLSTM*: This model uses a single-stage Bayesian CNN-BiLSTM to directly forecast WF P_{out} value, instead of using a two-stage Bayesian CNN-BiLSTM combined with Copula. The inputs of the model remain the same as in the weather variable forecasting in Section 9.3 and output is another mixture distribution with multiple components, where each component is a univariate normal distribution, transformed and normalised by a SoftClip Bijector, so the values are in the range from 0 p.u. to 1.0 p.u.

9.5.3 Benchmarking Results

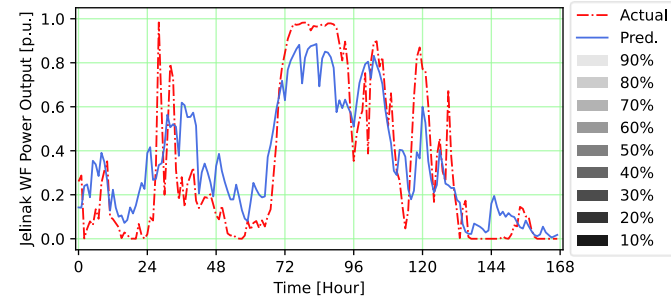
Fig. 9.10 shows the results of all benchmarking models in “test week”, while Table 9.6 to Table 9.11 compare their error metrics in “test week”, the whole “test year” and four seasons in “test year”, respectively. Seq2Seq and persistence models do not provide uncertainty ranges, since their outputs are deterministic. The calculation of pinball loss, CRPS and Winkler score requires the inputs to be quantiles, distributions, or prediction intervals. To compare Seq2Seq and persistence models’ pinball loss, CRPS and Winkler score with other models, they are assumed to follow degenerate distribution, meaning that degenerate random variable ($WF P_{out}$) has a probability of 1 at the value equal to the model output, and 0 elsewhere [246]. Therefore, the quantiles of degenerate random variable are also all equal to the model output value.

Compared to the proposed $WF P_{out}$ forecasting framework based on Bayesian CNN-BiLSTM and Vine-GMCM, the accuracy of Seq2Seq model is lower for the forecasted mean values, which is expected, as this model is used for a longer-term forecasting by the system operator. Although the degenerate distribution has an uncertainty range infinitely close to zero, the probabilistic error metrics are worse as the penalty from the cases when the actually realised $WF P_{out}$ values are away from the predicted ones offsets the benefit of the small range, confirming the advantage of using probabilistic model.

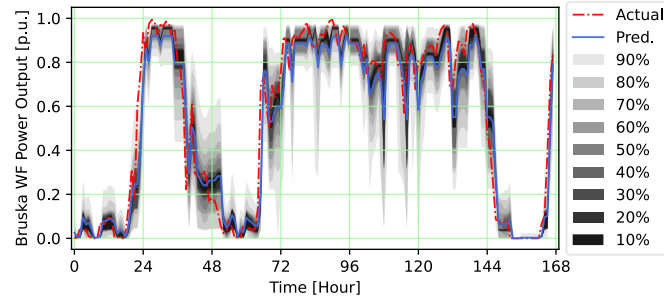
The MC benchmark model achieves better performance for both WFs and can be evaluated in terms of both mean values and forecasted percentile/confidence interval ranges. By comparing results in Fig. 9.8(a) and Fig. 9.8(b) with the results in Fig. 9.10(c) and Fig. 9.10(d), it can be seen that MC model returns wider percentile ranges than the combined Bayesian CNN-BiLSTM and Vine-GMCM model, in particular for Jelinak WF between 72-96 hours. Although the calculated pinball loss and CRPS values do not reflect big differences, Winkler score metric, which is more sensitive to the uncertainty range, more clearly reflects the improvements.



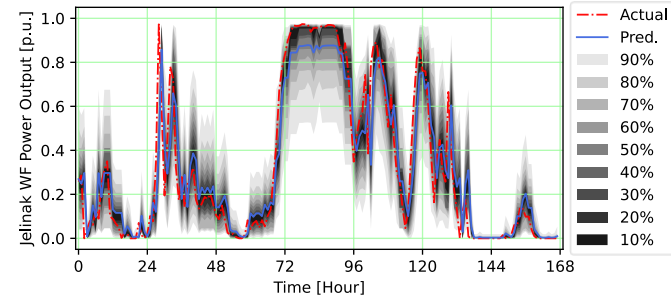
(a)



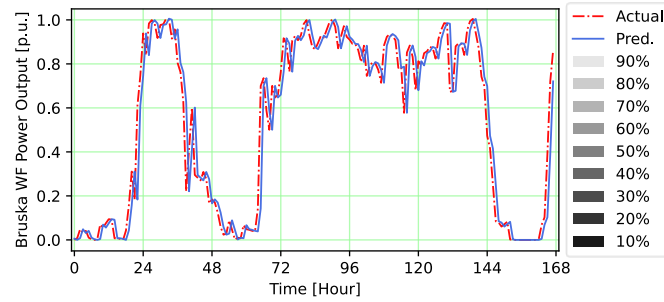
(b)



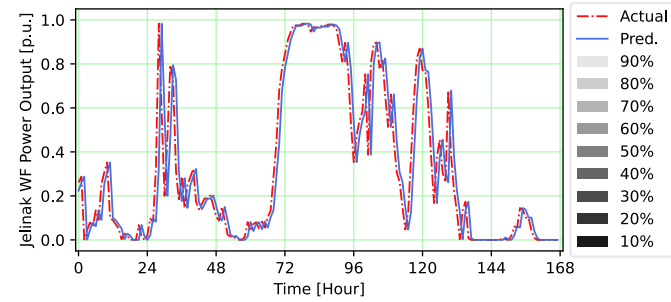
(c)



(d)



(e)



(f)

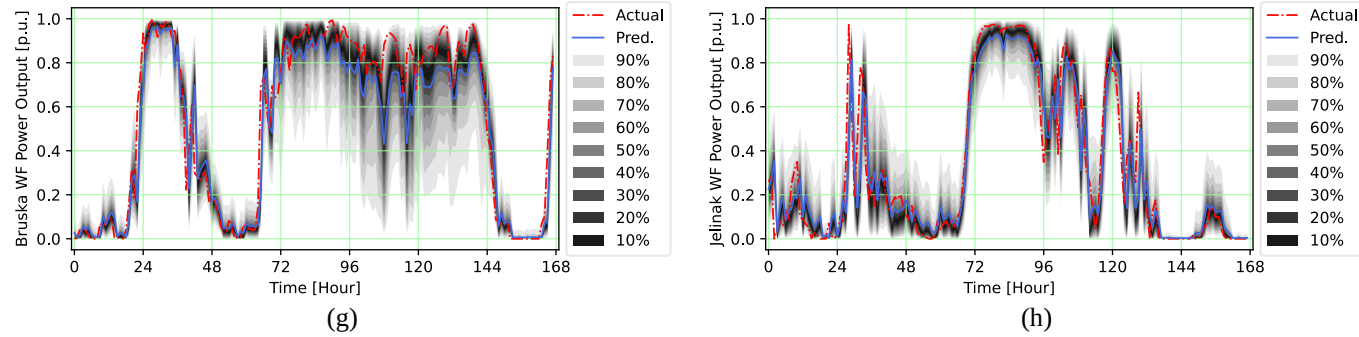


Fig. 9.10. The results of two benchmarking models for forecasting of WF P_{out} in “test week”: (a) Bruska WF by Seq2Seq, (b) Jelinak WF by Seq2Seq, (c) Bruska WF by MC, (d) Jelinak WF by MC, (e) Bruska WF by persistence, (f) Jelinak WF by persistence, (g) Bruska WF by single-stage Bayesian CNN-BiLSTM, (h) Jelinak WF by single-stage Bayesian CNN-BiLSTM.

Table 9.6 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Week”

Forecasted WF: Bruska (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (MWh)	E_o (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)
Seq2Seq	0.129	0.164	0.065	0.130	2.590	316.859	42.817	466.296	18.362	-149.437	-4.557
MC	0.079	0.115	0.026	0.054	0.425	146.792	19.297	326.847	13.147	-180.055	-5.544
Persistence	0.067	0.110	0.034	0.067	1.345	187.904	12.250	218.757	12.532	-30.853	-0.941
Single-stage Bayesian CNN- BiLSTM	0.092	0.133	0.029	0.058	0.527	131.903	22.192	424.459	15.906	-292.556	-8.966
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.070	0.113	0.025	0.049	0.386	156.055	14.227	265.534	12.259	-109.479	-3.355
Forecasted WF: Jelinak (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (MWh)	E_o (%)	E_U (MWh)	E_U (%)	E_T (MWh)	E_T (%)
Seq2Seq	0.148	0.185	0.074	0.148	2.963	439.316	118.884	307.336	24.123	131.98	8.03
MC	0.091	0.130	0.029	0.059	0.473	226.179	57.982	233.227	19.002	-7.048	-0.434
Persistence	0.082	0.134	0.041	0.082	1.633	209.132	31.869	202.431	20.502	6.701	0.408
Single-stage Bayesian CNN- BiLSTM	0.084	0.126	0.029	0.058	0.439	204.130	53.585	217.844	17.472	-13.714	-0.843
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.080	0.126	0.028	0.056	0.406	175.947	38.659	226.850	19.345	-50.903	-3.127

Table 9.7 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Year”

Forecasted WF: Bruska (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.113	0.156	0.056	0.113	2.253	10.059	76.284	6.984	22.658	3.075	6.987
MC	0.079	0.116	0.027	0.094	0.477	6.150	49.736	5.811	18.363	0.339	0.771
Persistence	0.060	0.098	0.030	0.058	1.195	4.583	21.195	4.455	19.898	0.128	0.291
Single-stage Bayesian CNN- BiLSTM	0.070	0.105	0.024	0.048	0.425	4.924	34.774	5.713	19.141	-0.789	-1.793
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.059	0.092	0.022	0.043	0.405	4.444	31.735	4.445	14.812	-0.001	-0.002
Forecasted WF: Jelinak (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.137	0.190	0.069	0.137	2.749	10.918	103.429	6.698	29.326	4.220	12.635
MC	0.077	0.116	0.026	0.102	0.440	5.296	56.554	4.536	18.873	0.760	2.277
Persistence	0.062	0.106	0.031	0.061	1.237	3.960	25.716	3.967	22.043	-0.008	-0.023
Single-stage Bayesian CNN- BiLSTM	0.080	0.124	0.027	0.055	0.454	7.370	48.632	2.875	15.760	4.495	13.460
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.061	0.100	0.022	0.043	0.387	4.200	39.135	3.644	16.075	0.556	1.664

Table 9.8 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Spring Season”

Forecasted WF: Bruska (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.129	0.175	0.064	0.129	2.579	1.861	62.290	1.671	25.622	0.190	1.995
MC	0.094	0.133	0.034	0.090	0.653	1.244	44.042	1.328	19.864	-0.084	-0.878
Persistence	0.067	0.106	0.034	0.066	1.343	0.949	21.091	0.891	17.770	0.058	0.608
Single-stage Bayesian CNN- BiLSTM	0.085	0.123	0.029	0.058	0.493	1.003	42.876	1.332	18.578	-0.330	-3.466
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.067	0.099	0.025	0.049	0.439	0.831	31.477	1.002	14.587	-0.171	-1.799
Forecasted WF: Jelinak (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.158	0.205	0.079	0.158	3.166	1.814	105.070	1.098	26.244	0.716	12.114
MC	0.090	0.128	0.030	0.092	0.506	0.867	53.835	0.797	18.547	0.070	1.177
Persistence	0.074	0.121	0.037	0.074	1.489	0.700	25.635	0.669	21.049	0.032	0.534
Single-stage Bayesian CNN- BiLSTM	0.100	0.149	0.034	0.068	0.566	1.323	59.503	0.510	13.842	0.813	13.756
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.072	0.113	0.026	0.051	0.437	0.659	44.035	0.672	15.225	-0.013	-0.218

Table 9.9 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Summer Season”

Forecasted WF: Bruska (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.121	0.165	0.060	0.121	2.417	3.761	73.400	2.552	21.591	1.209	7.133
MC	0.086	0.122	0.03	0.091	0.486	2.257	48.020	2.241	18.298	0.016	0.093
Persistence	0.072	0.111	0.036	0.071	1.443	1.917	23.019	1.853	21.500	0.064	0.380
Single-stage Bayesian CNN- BiLSTM	0.079	0.114	0.027	0.054	0.452	1.984	34.072	2.155	19.370	-0.171	-1.010
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.069	0.105	0.025	0.050	0.437	1.860	32.471	1.758	15.668	0.102	0.600
Forecasted WF: Jelinak (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.154	0.211	0.077	0.154	3.089	4.544	97.949	3.074	30.536	1.469	9.992
MC	0.085	0.123	0.029	0.096	0.472	2.152	53.406	2.020	18.919	0.132	0.895
Persistence	0.070	0.116	0.035	0.070	1.409	1.756	24.791	1.719	22.550	0.036	0.247
Single-stage Bayesian CNN- BiLSTM	0.087	0.130	0.030	0.060	0.487	3.105	44.945	1.185	15.199	1.920	13.053
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.069	0.108	0.025	0.049	0.431	1.840	36.132	1.573	16.356	0.267	1.817

Table 9.10 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Autumn Season”

Forecasted WF: Bruska (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.095	0.140	0.048	0.095	1.901	1.638	83.755	1.268	23.046	0.370	4.958
MC	0.068	0.103	0.023	0.084	0.407	1.047	54.242	1.020	18.448	0.027	0.365
Persistence	0.054	0.091	0.027	0.052	1.079	0.804	23.470	0.844	20.939	-0.040	-0.536
Single-stage Bayesian CNN- BiLSTM	0.055	0.088	0.020	0.040	0.358	1.021	25.637	0.661	19.044	0.360	4.828
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.051	0.085	0.019	0.038	0.377	0.881	29.043	0.688	15.549	0.193	2.592
Forecasted WF: Jelinak (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.107	0.158	0.053	0.107	2.138	1.540	123.139	1.013	39.456	0.527	13.819
MC	0.061	0.098	0.02	0.11	0.365	0.869	76.043	0.598	22.377	0.271	7.091
Persistence	0.049	0.093	0.025	0.048	0.983	0.579	34.463	0.594	27.820	-0.015	-0.388
Single-stage Bayesian CNN- BiLSTM	0.061	0.097	0.021	0.041	0.357	1.073	48.824	0.382	23.608	0.690	18.089
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.050	0.086	0.018	0.035	0.327	0.678	49.735	0.515	20.973	0.164	4.284

Table 9.11 Comparison (Benchmarking) of Wind Farm Power Output Forecasting Results by Different Models in “Test Winter Season”

Forecasted WF: Bruska (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.104	0.140	0.052	0.104	2.089	2.799	89.744	1.492	21.390	1.307	12.943
MC	0.069	0.105	0.023	0.109	0.399	1.602	55.046	1.222	17.011	0.380	3.763
Persistence	0.043	0.077	0.022	0.040	0.867	0.913	17.001	0.867	18.346	0.046	0.456
Single-stage Bayesian CNN- BiLSTM	0.060	0.093	0.020	0.040	0.395	0.916	45.464	1.565	19.365	-0.648	-6.424
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.045	0.074	0.017	0.035	0.364	0.872	33.517	0.997	13.302	-0.125	-1.236
Forecasted WF: Jelinak (Model)	MAE [p.u.]	RMSE [p.u.]	Pinball Loss [p.u.]	CRPS [p.u.]	Winkler Score [p.u.]	E_o (GWh)	E_o (%)	E_U (GWh)	E_U (%)	E_T (GWh)	E_T (%)
Seq2Seq	0.124	0.171	0.062	0.124	2.480	3.021	102.727	1.514	25.129	1.507	16.811
MC	0.069	0.109	0.023	0.11	0.412	1.409	54.537	1.120	17.550	0.289	3.220
Persistence	0.052	0.094	0.026	0.051	1.044	0.924	23.682	0.985	19.464	-0.061	-0.678
Single-stage Bayesian CNN- BiLSTM	0.073	0.116	0.025	0.050	0.416	1.870	48.864	0.797	15.515	1.072	11.963
Two-stage Bayesian CNN-BiLSTM & Vine-GMCM	0.052	0.088	0.019	0.037	0.342	1.023	36.800	0.885	14.303	0.138	1.539

Regarding the deterministic error metrics, compared to other probabilistic models, the two-stage Bayesian CNN-BiLSTM and Vine-GMCM model has lower values for both MAE and RMSE metrics. In the “test week” results, the persistence model performs surprisingly well in Bruska WF forecasting, achieving smaller MAE, RMSE and better energy-based metrics than all other models. In Jelinak WF, it also has better performance than other benchmarking models regarding the deterministic and energy-based error metrics. Such performance of persistence model is well known for short prediction horizons and there are two main reasons for these seemingly counterintuitive results.

Firstly, the measurements have been aggregated into one-hour resolution, and hour-ahead forecasting is equivalent to a “next-point”, or “point-ahead” forecasting. Ref. [247] illustrates that although the persistence model is a very simple model with very low computational requirements, it continues to perform at a par with, if not to outperform the more advanced models, if the wind energy forecasting horizons are shorter than a few hours. Refs [248] and [249] also confirm that persistence model has a very high accuracy for a short-term wind energy forecasting, from a few seconds to 6-hour ahead, stating that its accuracy will decrease for longer forecasting horizons. Similarly, [250] develops a model for wind farm power output forecasting for the next 0 – 36 hours, also finding that in the first four hours the persistence model performs better, but it produces larger errors when the forecasting horizon is longer. The results in [251] show that in the prediction of icing-related wind power production loss, the conceptually simple persistence model also achieves smaller MAEs than the proposed model (random forest), if the forecasting horizons are only a few hours. Ref. [252] researches the day-ahead PV output forecasting at the 15-minute interval, again finding that the persistence model can outperform the proposed model (autoencoder-LSTM) for certain weather conditions (e.g., continuous clear days). In [253], it was also found that in the point-ahead solar irradiance forecasting (15-minute data resolution after aggregation), the persistence model has the lowest deterministic errors.

Secondly, the two-stage Bayesian CNN-BiLSTM and Vine-GMCM s not optimised for minimizing the deterministic errors. In stage one, the loss function in Equation (9.5) has two parts, which consist of “likelihood cost”, the expectation of

$p(D|w)$ for $w \sim q(w|\theta)$, and “complexity cost”, prior-dependent term $KL(q(w|\theta)|p(w))$. In stage two, the optimisation for fitting GMCM is also maximization of a non-convex log-likelihood function subject to linear constraints (more details could be found in [199]). Therefore, instead of achieving the lowest deterministic errors (e.g., MAE or RMSE), the presented models are built and trained to minimise probabilistic errors.

In almost all seasons, the proposed model (Two-Stage Bayesian CNN-BiLSTM & Vine-GMCM) has the best probabilistic forecasting performance for both Bruska and Jelinak WFs regarding pinball loss, CRPS, and Winkler score. The only exception is Autumn season at Bruska WF, where Single-Stage Bayesian CNN-BiLSTM model achieves a 5 % lower Winkler score than the proposed model, while both pinball loss and CRPS still indicate better performance of the two-stage model.

The most likely reason for a higher Winkler score of the two-stage model is that Winkler score is a metric focusing on a single prediction interval, while pinball loss uses a set of quantiles and CRPS considers the whole distributions. An example plot of a Winkler score calculation results is shown in Fig. 9.11. More specifically, Winkler score gives constant values for all observations within the prediction interval (equal to the width of the interval) and it will add a penalty proportional to a distance of the observation from a prediction interval. Winkler score effectively assumes uniform distribution for a given prediction interval and it does not consider more realistic distributions, nor the location/probability of the observation within the interval.

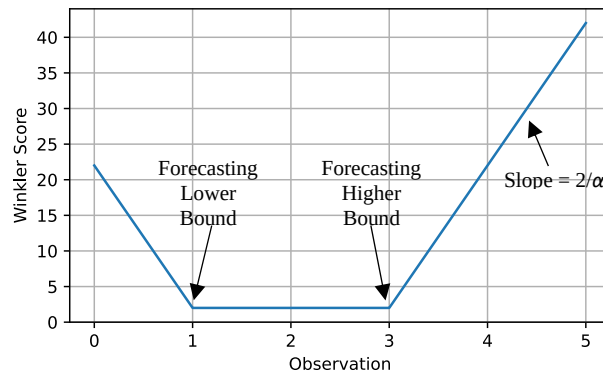


Fig. 9.11. Visualisation of Winkler score (α is the significance level, see Equation (2.16)).

Therefore, Winkler score is a single interval-based indicator, while CRPS is more comprehensive, as it assesses the whole distribution. To provide a more detailed

information, it is common to use a more complex variant of Winkler score with multiple intervals, which will eventually approximate to CRPS [253]. The proposed two-stage model always has lower pinball loss and CRPS values (and therefore better performance) than a single-stage model. After checking the cases where two-stage model had higher Winkler score than single-stage model, it was found that the prediction intervals in both models covered the observations, but interval lengths were slightly larger for two-stage model, resulting in a slightly higher Winkler score. These samples occurred more often in the autumn season in Bruska WF, where average Winkler score for two-stage method is higher, but its pinball loss and a more comprehensive CRPS metric are consistently lower. An example when the observation is 0.471 p.u., the prediction interval by the Single-Stage Bayesian CNN-BiLSTM model is [0.450, 0.842] p.u., the prediction interval by the proposed two-stage model is [0.352, 0.803] p.u., with the plots of their distributions is illustrated in Fig. 9.12.

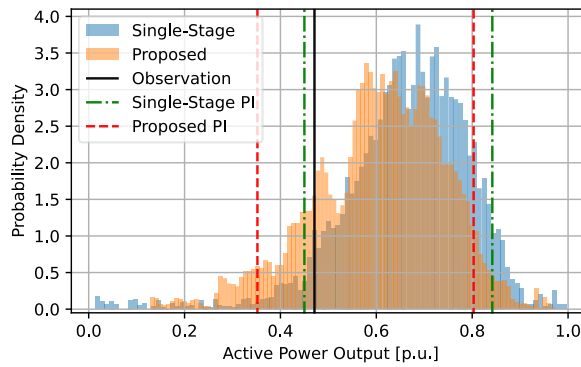


Fig. 9.12. Example of prediction distributions.

It can be seen that in this case both models have different distributions and that the proposed two-stage model gives wider prediction intervals, as it more accurately reflects the possible variations and uncertainties in the predicted outputs (e.g., the observation occurs near one of the modals of the two-stage model distribution in Fig. 9.12). However, it is clear that although the observation is closer to the lower interval boundary given by the Single-Stage Bayesian CNN-BiLSTM model, implying worse model performance due to a lower probability, the Winkler score neglects that fact. The calculated pinball loss, Winkler score and CRPS are 0.069, 0.392 and 0.139 p.u. for the Single-Stage Bayesian CNN-BiLSTM model and 0.044, 0.451 and 0.089 p.u. for the proposed two-stage model, respectively.

The proposed model usually also outperforms all other benchmarking models regarding the deterministic errors, with two exceptions: MAE in Autumn at Jelinak WF and Winter at Bruska WF. The possible reasons have been discussed before. Overall, the presented results demonstrate the consistent satisfactory performance of the proposed model.

Persistence model has a satisfactory accuracy if uncertainty is not considered. However, persistence model has much larger probabilistic errors compared to other probabilistic models (MC, single-stage Bayesian CNN-BiLSTM, and two-stage Bayesian CNN-BiLSTM & Vine-GMCM) for similar reasons as in the case of Seq2Seq model.

The results obtained by direct WF P_{out} forecasting of single-stage Bayesian CNN-BiLSTM are less accurate than the proposed forecasting framework. For Bruska WF forecasting in “test week”, its performance is only better than Seq2Seq model. In the “test year”, the obtained results also show the consistent improvement of the two-stage approach, which effectively confirms benefits of the proposed combined Bayesian CNN-BiLSTM and Vine-GMCM model, which provides the most accurate forecasting results.

9.6 Extended Example Case Study 9.5: Day-Ahead Forecasting with NWP

The results presented here are an initial investigation of the capability of the presented Bayesian CNN-BiLSTM model (single-stage from a proposed two-stage modelling framework) to be applied for a day-ahead forecasting using NWP. This model is compared and benchmarked against some of the models participating in the GEFCom2014 competition [38].

The GEFCom2014 competition was quite different forecasting problem from the one in this chapter, which also used significantly different dataset: its purpose was to estimate WF power outputs for provided day-ahead numerical weather prediction (NWP) forecasted directional wind speeds. The competition was active on CrowdANALYTIX platform [254] for four months (8 Aug, 2014 – 15 Dec, 2014) and consisted of 15 tasks for 10 WFs. In each task, day-ahead WF power outputs should

be forecasted on a rolling basis for the period of one month [38]. The host platform [254] shows 16 tasks instead of 15 in [38] and 250 solvers, where Leaderboard in [254] gives detailed results for the top solvers in every task.

For the purposes of direct comparison and benchmarking against the GEFCom2014 dataset, the model presented in this chapter is modified, where the key differences were:

- 1) Although removing the outliers may have strong impact on the final results (Chapter 5), mainly the raw data from the GEFCom2014 dataset are used (only missing data in short periods, < 6 hours, are imputed by linear interpolation.
- 2) No feature engineering (including correlation and clustering analysis) is performed and model hyperparameters remain the same for all tasks and all WFs. No WF OR analysis is performed.
- 3) In the models modified for this benchmarking, the same predictor variables are used, including NWP wind speeds and wind directions at 10 m and 100 m heights in the past 72 hours and the next 24 hours from all 10 WFs. Since there were no air density (or temperature) data that could be used to represent seasonality, “date and time” were explicitly included as the additional predictor variables (using cos and sin components) to inform the model about the hour of the day, day of the month and season of the year, as in [255-257]. The outputs of the modified model are day-ahead distributions of hourly power output for each WF, which are mixture distribution of multiple univariate normal components, transformed and renormalised using a SoftClip Bijector, to be in range 0 p.u. – 1.0 p.u.
- 4) In terms of the further modifications of the model, only Bayesian CNN-BiLSTM is used, while Copula model is not considered. In the Bayesian CNN-BiLSTM, Max Pooling layer is removed, since it degrades model performance in the initial trials. The uncertainty for two BiLSTM stacked layers is calculated following [237], instead of [236], since this forecasting is more difficult to converge than the hour-ahead forecasting problem

Table 9.12 below shows the comparison of aggregated monthly errors with [256] and [258] (the best model in this competition). There are sets of three numbers in each cell (x, y, z), where x is the pinball loss (or RMSE) provided in the corresponding reference,

y is the pinball loss (or RMSE) of the reduced model variant, and z is the percentage difference, i.e., $(y - x) / x \times 100$.

Ref. [259] uses neural network-based approach, but it calculates pinball loss metric for individual WFs as the averaged values in year 2013 (Jan 13 to Dec 13). In comparisons, the calculated pinball loss values for all WFs from Dec 12 to Nov 13, with Table 9.13 listing average errors (the same format as in Table 9.12).

Table 9.12 Comparison and Benchmarked Numerical Values of the Reduced Model with the Results from [256, 258]

Month	RMSE [p.u.] (versus [256])	Pinball Loss [p.u.] (versus[256])	Pinball Loss [p.u.] (versus [258])
Dec 12	(0.15796, 0.16162, 2)	(0.03998, 0.04250, 6)	(0.039, 0.043, 9)
Jan 13	(0.14651, 0.15612, 7)	(0.03792, 0.04183, 10)	(0.036, 0.042, 18)
Feb 13	(0.15689, 0.16021, 2)	(0.04080, 0.04333, 6)	(0.040, 0.043, 9)
Mar 13	(0.15743, 0.16535, 5)	(0.04014, 0.04364, 9)	(0.039, 0.044, 12)
Apr 13	(0.13284, 0.14351, 8)	(0.03273, 0.03589, 10)	(0.031, 0.036, 17)
May 13	(0.15809, 0.17122, 8)	(0.04018, 0.04541, 13)	(0.039, 0.045, 16)
Jun 13	(0.13447, 0.13970, 4)	(0.03350, 0.03610, 8)	(0.032, 0.036, 13)
Jul 13	(0.15330, 0.16395, 7)	(0.03824, 0.04213, 10)	(0.036, 0.042, 16)
Aug 13	(0.15499, 0.17174, 11)	(0.03829, 0.04469, 17)	(0.037, 0.045, 21)
Sep 13	(0.16584, 0.18463, 11)	(0.04014, 0.04696, 17)	(0.039, 0.047, 21)
Oct 13	(0.17514, 0.18185, 4)	(0.04360, 0.04764, 9)	(0.042, 0.048, 14)
Nov 13	(0.16185, 0.16708, 3)	(0.04176, 0.04596, 10)	(0.039, 0.046, 18)
Average	(0.15461, 0.16392, 6)	(0.03894, 0.04301, 10)	(0.037, 0.043, 15)

Table 9.13 Comparison and Benchmarked Numerical Values of the Reduced Model with the Results from [259]

WF	Pinball Loss [p.u.]	WF	Pinball Loss [p.u.]
WF1	(0.04900, 0.03608, -26)	WF2	(0.04300, 0.04264, -1)
WF3	(0.04200, 0.04480, 7)	WF4	(0.04500, 0.04252, -6)
WF5	(0.04700, 0.04335, -8)	WF6	(0.04900, 0.04520, -8)
WF7	(0.03600, 0.03743, 4)	WF8	(0.04300, 0.04546, 6)
WF9	(0.04000, 0.04333, 8)	WF10	(0.05400, 0.04927, -9)
Average	(0.04480, 0.04301, -4)		

Fig. 9.13 below depicts the probabilistic forecasting results for the first 100 hours in Jan 2013 for WF1. The results from the same time and for the same WF can also be found in “Fig. 1 of [259]”, while “Fig. 3 of [256]” shows the first three days of this forecasting period in Jan 2013 for WF1.

The results in Table 9.12 and Table 9.13 and Fig. 9.13 confirm that the reduced variant of the model is a reasonably accurate day-ahead wind power forecasting model, fully comparable to other models used for benchmarking.

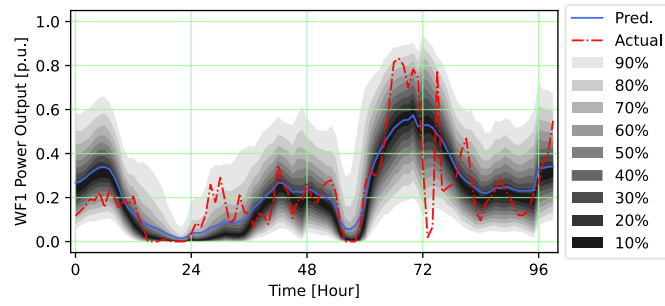


Fig. 9.13. Forecasting results for the first 100 hours in Jan 2013 for WF1.

9.7 Discussions and Conclusions

This chapter presents a novel generalised framework for probabilistic forecasting of WF power outputs, which combines Bayesian CNN-BiLSTM and Vine-GMCM methods into a single model.

The Bayesian CNN-BiLSTM stage is used to analyse and handle high dimensionality of input meteorological variable (WS, AD and WD), where measurements from the nearby WFs (in the same WF clusters) are used as the additional model inputs. The chapter also shows how presented Bayesian CNN-BiLSTM model can include WF OR as another important predictor variable, where the unplanned outages (e.g., due to faults) are forecasted as a categorical OR variable in the model. The results from this stage clearly show importance of considering WF ORs for wind energy forecasting and include both the epistemic and aleatoric uncertainties in the output results.

Afterwards, Vine-GMCM model is introduced, which models and captures cross-correlations among WS, AD, WD and WF P_{out} , while also considering WF ORs. At the final stage, Bayesian CNN-BiLSTM and Vine-GMCM models are combined in the proposed general framework, which can both improve the accuracy of the WF P_{out} predictions and reduce the uncertainties around the forecasted mean values.

The performance of the combined model is compared to four other models, of which one was used by the Croatian system operator, while two other models are commonly applied for wind energy forecasting, confirming the overall robustness and accuracy, i.e., higher forecasting confidence levels of the presented combined model.

The fourth model, which is a reduced (single-stage) variant of the presented Bayesian CNN-BiLSTM model, is also applied for a day-ahead forecasting and compared with GEFCom2014. The reduced day-ahead model built within the time available for the PhD may not be the best model for this GEFCom2014 dataset, in particular when compared to [258]. However, this “benchmarking exercise” is very useful and, based on the presented analysis, the results in Table 9.12 and Table 9.13 and Fig. 9.13 confirm the this reduced model is a reasonably accurate day-ahead wind power forecasting model, fully comparable to other sophisticated models used in the presented benchmarking analysis.

10.1 Thesis Summary

This thesis presents a number of novel and improved data-driven models for analysing the measurements of both demand and supply sides. The proposed methods have shown satisfactory performance in several research topics, including process modelling, time series forecasting, and uncertainty analysis.

Chapter 2 introduces the main concepts and literature review related to two important modelling and analysis tools: neural network-based models and Copula correlation models. It also presents and compares the commonly used error metrics for assessing the deterministic and probabilistic models. Afterwards, a literature review of modelling and forecasting of load, and wind generation at wind turbine and wind farm levels is presented. Finally, it reviews the definitions and causes of different types of uncertainties and discusses the potential approaches for their representation and resolution.

Chapter 3 firstly compares two methods in assessing the variations of demands, and it finds that both of them produce evaluation of ranges of variations of the load profiles. The three datasets from different aggregation levels are analysed: household, substation and city. The load forecasting was first analysed with the comparison of day-ahead load forecasting with MLP, GBRT and SB-LSTM methods, where all the demand data, time/calendar variables and weather parameter recordings are considered. The most promising model, SB-LSTM, is further extended to provide week-ahead load forecasting and multi-dimensional total demand and disaggregated heating load forecasting.

Chapter 4 proposes a novel Fourier series regression method to extract selected frequency components from time series of total active power demands, temperature, and solar irradiance. The results show that the original time series can be reasonably accurately represented using these selected frequency components. Then, it presents the analysis of the correlations between different frequency components, confirming

that specific combinations of demand components have stronger negative correlations with temperature and solar irradiance than the original non-disaggregated time series. Afterwards, the stronger negatively correlated frequency components are used as inputs in a sophisticated load disaggregation model built with CNN-BiLSTM, for which the final disaggregation results are more accurate. In both cases, the hyperparameters of CNN-BiLSTM instances are selected through Bayesian optimisation, which can reduce the inherent effects of uncertainty of hyperparameter values.

Chapter 5 presents a detailed review of WT and WF measurement outliers, starting with the analysis of typically available WT and WF measurement data. Then, it gives a review of the definitions, characteristics and causes of outliers in WT and WF measurements reported in existing literature. Afterwards, it reviews methods for the detection of outliers, where statistical methods, physical methods, the combination of statistical and physical methods, and robust statistical models are all reviewed. Finally, it reviews various approaches for assessing the accuracy/success of outlier detection methods.

Chapter 6 extends the review and analysis of wind turbine and wind farm measurements from Chapter 5. It firstly proposes a WT-level outlier detection framework to identify outliers from both statistical and physical points of view. In particular, a novel simulation-based approach is proposed to identify the normal data and explain the “false outliers” around the cut-out wind speed, which are due to the WT hysteresis control. Then, it proposes a 6-parameter power curve model, which can accurately model the whole range of operational wind speeds, including region around cut-out wind speed. Finally, it also presents an iterative approach to cluster different WF operating regimes if there are only single-sensor WF measurements available.

Chapter 7 firstly analyses three analytical distributions and their parameter estimation methods.: two/three-parameter Weibull distribution, generalised normal distribution and von-Mises distribution. Then it applies particle swarm optimisation to find the weights of individual distributions and parameter values of mixture distribution. Afterwards, it fits empirical distributions with three analytical distribution models, in

order to accurately and efficiently represent the variations of wind speed, power output and wind direction using the historical data from a UK wind farm.

Chapter 8 investigates the importance of including WD as an additional input explanatory variable to the WS for assessing the uncertainties in the WT P_{out} . It compares the 2D P_{out} -WS model from Chapter 7 with a 3D P_{out} -WS-WD model for two general cases of inputting either deterministic or uncertain/range of WS and WD values. The 3D model is implemented using Vine-GMCM, which can accurately model the pair dependencies and the cross-correlation structures. It shows that the inclusion of WD in the analysis might provide noticeable improvements in models with no uncertainties in input WS and WD data, and if the input WS and WD are ranges of values instead of deterministic values, WS is a much stronger contributor to the total P_{out} uncertainty than WD.

Chapter 9 presents a novel generalised framework for probabilistic forecasting of WF power outputs, which combines Bayesian CNN-BiLSTM and Vine-GMCM methods into a single model. It firstly uses Bayesian CNN-BiLSTM to provide hour-ahead forecasting for the meteorological variable (WS, AD and WD) and WF operating regime. Then, Vine-GMCM model is introduced in the model to capture cross-correlations among WS, AD, WD and WF P_{out} , while also considering WF ORs. Afterwards, the models are combined in a general framework, which can both improve the accuracy of the WF P_{out} predictions and reduce the uncertainties around the forecasted mean values. The proposed model is also benchmarked against four other models showing a very good performance in terms of both deterministic and probabilistic error metrics. Finally, a reduced variant of the proposed model is applied for a day-ahead forecasting and compared with other models based on available datasets for GEFCom2014 competition.

10.2 Research Implications

The research presented by this thesis firstly demonstrates the performance of SB-LSTM model in forecasting day-ahead and week-ahead total load and heating load, where the calendar and meteorological explanatory variables are also taken into account. It demonstrate that the SB-LSTM model is capable of processing long

sequence data and has a very good performance for modelling the relationships among multiple dimensions.

The proposed novel Fourier series regression can effectively identify the periodical changes in the total demands (e.g., daily, weekly, and seasonal variations), and the reconstructed load profiles can also achieve reasonable accuracy. The following correlation analysis between load time series and meteorological variables (e.g., temperature and solar irradiance) helps the CNN-BiLSTM load disaggregation model reach more satisfactory performance, because the heating load and lighting load are negatively correlated with temperature and solar irradiance. The Bayesian optimisation method can help select the hyperparameters of the model systematically and reduce the negative effects of suboptimal hyperparameter settings.

The WT and WF measurements outlier review firstly shows that the previous work was focused on the outliers in WT measurements, while less attention was paid to the analysis of outliers at WF-level. A comprehensive review of definitions, characteristics and causes of outliers in WT and WF measurements reported in the existing literature is also presented. In terms of outlier detection, it was found that the statistical methods (density-based, distance-based, correlation-based, or image-based) are the most popular. Regarding outlier treatment procedures, it was found that in most of the previous work outliers are simply removed and that only a small number of references treated the outliers as missing data and performed data imputation.

The proposed generic WT-level outlier detection framework handles WT-level outliers from both statistical and physical points of view and the simulation-based approach is able to explain outliers due to WT cut-out effects and applied averaging window. Unlike most of the power curve models in existing literature, the novel six-parameter power curve model can represent both the cut-in and the cut-out regions accurately. The simple operational wind farm model implies the relationship between WT and WF level data, and also the relationship among different WF operating regimes. This simple relationship, as applied and shown in the WF P_{out} forecasting work, further improves the overall prediction accuracy.

The thesis also proposes and compares two types of probability models for modelling uncertainty of wind energy resources, including P_{out} , WS, WD and AD. The first category of method is a conditional probability model, which fits a single or a mixture of distributions to the measurements located within a WS bin (i.e., method of bins). The mixture of distributions is found to be more accurate, as it can capture multiple modals existing in the data. In particular, von-Mises distribution can correctly model the circular data like wind direction.

The second type of probability model is a joint probability model through Copula theory. More specifically, the firstly proposed Vine-GMCM model has shown great accuracy when modelling high-dimensional correlation with nonlinear pair dependencies. The thesis has implemented both 3D P_{out} -WS-WD model and 4D P_{out} -WS-AD-WD model using Vine-GMCM. Through chain rules and Vine theory, the conditional probability can also be derived, where the most interesting ones are $p(P_{out}|WS, WD)$ and $p(P_{out}|WS, AD, WD)$. The results show that additional explanatory variables can reduce the uncertainty of the estimations of P_{out} , and both models are able to take ranges of predictor variables as inputs, which then successfully consider the uncertainty due to parametric variability.

In the WF P_{out} probability forecasting research, the proposed two-stage framework concatenating Bayesian CNN-BiLSTM and Vine-GMCM has shown satisfactory performance according to deterministic and probabilistic error metrics. For the first time, Bayesian CNN-BiLSTM is used in wind energy forecasting. In this framework, Bayesian CNN-BiLSTM can accurately provide probabilistic forecasting for the weather variables and WF operating regime, and Vine-GMCM, as discussed above, then takes the ranges of the inputs and provide the conditional probability estimation for P_{out} . More importantly, the WF operating regime relationship is also applied to scale the forecasting P_{out} , and the improved results confirm the benefits of integrating operating regime with the forecaster.

10.3 Limitations of the Research and Future Work

The SB-LSTM model for load forecasting outperforms all other benchmarking models, but the limitation of current work is that a simple deterministic SB-LSTM

model is unable to provide uncertainty estimation. Since Bayesian optimisation is not applied, and there is still uncertainty due to empirical selection of the hyperparameters. In the load disaggregation work, although CNN-BiLSTM model reduces the hyperparameter uncertainty and side effects after applying Bayesian optimisation, the neural network itself is still a deterministic model. Therefore, in future research, a Bayesian neural network may be applied to obtain additional information on the uncertainty of the results. In addition, the neural network is essentially a “black box” model, and in future research, one way to significantly improve the current work is to concentrate research on increasing interpretability, because being able to understand which predictor variable contributes the most to the error, and what the sensitivities are, is crucial in decision making.

In the presented load disaggregation research, only two types of load are analysed: heating and lighting. Due to the rapid development and penetration of many other types of loads, e.g., smart devices and electric vehicles, it is also becoming increasingly important to be able to estimate them from the aggregate total load. The smart devices’ datasheet and unique operating characteristics may also be researched to help to improve the disaggregation model accuracy. Since some types of loads may be impacted by the electric price, the price data could be included in the model, which is expected to improve the model’s performance further.

From the outlier review work, and as reported in the literature on outlier detection and treatment, there is a general lack of information on the actual causes of outliers, e.g., operational/event logs and similar documentations, which effectively prevent successful, reliable and convincing testing of the proposed outlier detection methods. The sources of outliers are many and various, of which some may result in outliers with similar, or even the same patterns. This diversity may make different approaches proposed for outlier treatment incomparable. In future research, it is highly desired to build and publish benchmark data and labelled, or well-documented data that can be used for testing the accuracy/success of outlier detection and removal/treatment methods.

In the outlier identification research and WT/WF power curve modelling work, for the WT simulation method, the biggest challenge is the hysteresis control algorithm

parameter setting. The actual setting for the exact wind turbine at the specific location and for given time period is very difficult to obtain, but they may have significant impacts on the final results. The second limitation of the algorithm is neglecting of the presence of wind direction and the reorientation of the WT. A full simulation should also consider the wake effects and turbulence, and a dynamic model is desired. In clustering WF operating regime for single sensor WF measurements, the current algorithm takes only the distance between the observed P_{out} value and deterministic power curve as a criterion, and it is a naïve approach without appreciating that different operating regimes also have various probabilities. In future research, a Bayesian probability framework and posterior estimation may be utilised, where the prior is the probability of operating regimes and the likelihood can be estimated using another probability model (e.g., Copula).

In this thesis, to fit the mixture of distributions to the data within a wind speed bin, the distribution parameters are estimated using particle swarm optimisation method. The limitation is that the optimisation may be very slow in some cases, and to prevent local optimum results, the same optimisation algorithm may be required to rerun several times until converging properly. In future research, it is suggested that other more efficient methods could be tried and checked, for example, the expectation–maximisation (EM) algorithm. In addition, it is desired to set the number of single distributions used in a mixture distribution properly, because too few distributions may cause under-fitting and too many distributions may lead to over-fitting (and also computation burden). To achieve this, certain error metrics that consider both the model complexity and accuracy can be used, for example, Bayesian information criterion (BIC).

Although Vine-GMCM has shown great accuracy in correlation modelling, there are still scalability limitations. If there are more variables to be included in the joint probability model, in theory, all of them should be able to be included in the model similarly following the chain rule and Vine theory. However, in practice, because GMCM is used to model every pair correlation individually, its parameter estimation and optimisation algorithm take a long time and are again prone to finding only local optimum. In addition, since GMCM is essentially a mixture model, the optimal number

of elements are difficult to determine, which is similar to the case for the mixture of distributions discussed in the last paragraph. Therefore, in future research, the new optimisation algorithm for parameter estimation of GMCM could be improved and the new error metrics (e.g., BIC) may also be used to assess the model. To improve the fitting efficiency for each pair dependence in scaled problems, other simpler Copulas may also be developed and tested, which may be used to replace GMCM considering the tradeoff between computation time and accuracy.

The two-stage WF P_{out} probability forecasting framework consisting of Bayesian CNN-BiLSTM and Vine-GMCM is accurate, but it still can be improved. Firstly, the parameters of Vine-GMCM are obtained as deterministic values through the optimisation algorithm, while in future research, the Bayesian method could be applied to include the uncertainties from model parameters. Secondly, the current two stages are eventually two separate models. That has advantages of looser coupling and simplicity, as it is easier and more convenient to optimise two models separately using different algorithms at the same time, but it may also introduce additional errors. In future research, it would be interesting to see and compare the case if the two models are trained at the same time using a single optimisation algorithm and with a loss function that only assesses the difference between the predicted P_{out} and the actual P_{out} .

The thesis focuses on wind energy generation, where the actual measurements from a number of wind farms are used to train and test the proposed models, and the presented analysis and equations show that those models can be extended to higher dimensions and scaled to larger wind farms. The proposed models from Chapter 6 to Chapter 9 can also be applied to analyse other renewable based generation systems, such the PV, because those methods are data-driven models, which means they are not constrained by the physics of wind turbine and wind farm, and they are built and trained to describe the complex relationships between an array of input numbers of output values. In future work, a more complex real-world energy system can be used to further test and optimise the proposed models. Importantly, in such a system, in addition to wind energy, there may exist other renewable generations (e.g., PV) and other types of load (e.g., transportation and heating). And both electric market and policy play important roles in the system as well, thus it is essential to have comprehensive solutions that

benefit various stakeholders, which means the proposed model should also be extended in such a way that it can interact/be integrated with other modules/players in the system.

Bibliography

- [1] J. A. P. Lopes *et al.*, "The future of power systems: Challenges, trends, and upcoming paradigms," *WIREs Energy and Environment*, vol. 9, no. 3, p. e368, 2020.
- [2] "World Energy Outlook 2021," International Energy Agency (IEA), 2021, Available: <https://www.iea.org/reports/world-energy-outlook-2021>.
- [3] "UK enshrines new target in law to slash emissions by 78% by 2035," UK Government, 2020, Available: <https://www.gov.uk/government/news/uk-enshrines-new-target-in-law-to-slash-emissions-by-78-by-2035>.
- [4] "Transitioning to a net zero energy system: smart systems and flexibility plan 2021," UK Government, 2021, Available: <https://www.gov.uk/government/publications/transitioning-to-a-net-zero-energy-system-smart-systems-and-flexibility-plan-2021>.
- [5] "Wind energy in the UK: June 2021," Office for National Statistics, UK, 2021, Available: <https://www.ons.gov.uk/economy/environmentalaccounts/articles/windenergyintheuk/june2021>.
- [6] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, Jan. 2019.
- [7] G. Papaefthymiou, B. Hasche, and C. Nabe, "Potential of Heat Pumps for Demand Side Management and Wind Power Integration in the German Electricity Market," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 636-642, Oct. 2012.
- [8] E. Yao, P. Samadi, V. W. S. Wong, and R. Schober, "Residential Demand Side Management Under High Penetration of Rooftop Photovoltaic Units," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1597-1608, May 2016.
- [9] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis, "Context Aware Energy Disaggregation Using Adaptive Bidirectional LSTM Models," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3054-3067, Jul. 2020.
- [10] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 188-201, 2020.
- [11] Y. Zhao, L. Ye, W. Wang, H. Sun, Y. Ju, and Y. Tang, "Data-Driven Correction Approach to Refine Power Curve of Wind Farm Under Wind Curtailment," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, pp. 95-105, Jan. 2018.
- [12] *Wind Energy Generation Systems: Power Performance Measurements of Electricity Producing Wind Turbines*, BS EN 61400-12-1, BSI Standards Publication, 2017.
- [13] H. Kim, C. Singh, and A. Sprintson, "Simulation and Estimation of Reliability in a Wind Farm Considering the Wake Effect," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 274-282, Apr. 2012.
- [14] J. Tongdan and T. Zhigang, "Uncertainty Analysis for Wind Energy Production with Dynamic Power Curves," in *2010 IEEE 11th International Conference on*

- Probabilistic Methods Applied to Power Systems (PMAPS)*, Singapore, Jun. 2010, pp. 745-750.
- [15] D. Villanueva and A. Feijóo, "Normal-Based Model for True Power Curves of Wind Turbines," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1005-1011, Jul. 2016.
 - [16] M. Schlechtingen, I. F. Santos, and S. Achiche, "Using Data-Mining Approaches for Wind Turbine Power Curve Monitoring: A Comparative Study," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 3, pp. 671-679, Jul. 2013.
 - [17] P. Pinson, L. E. A. Christensen, H. Madsen, P. E. Sørensen, M. H. Donovan, and L. E. Jensen, "Regime-switching modelling of the fluctuations of offshore wind generation," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2327-2347, 2008.
 - [18] Y. Zhao and Y. Wang, "An Equivalent Aggregation Model of Wind Farm Considering Wind Curtailment," in *2021 IEEE 4th International Electrical and Energy Conference (CIEEC)*, 2021, pp. 1-7.
 - [19] D. Fang, M. Zou, G. Coletta, A. Vaccaro, and S. Z. Djokic, "Handling Uncertainties with Affine Arithmetic and Probabilistic OPF for Increased Utilisation of Overhead Transmission Lines," *Electric Power Systems Research*, vol. 170, pp. 364-377, May 2019.
 - [20] M. Zou, D. Fang, S. Djokic, V. Di Giorgio, R. Langella, and A. Testa, "Evaluation of Wind Turbine Power Outputs with and without Uncertainties in Input Wind Speed and Wind Direction Data," *IET Renewable Power Generation*, vol. 14, no. 15, pp. 2801-2809, Nov. 2020.
 - [21] M. Zou and S. Z. Djokic, "A Review of Approaches for the Detection and Treatment of Outliers in Processing Wind Turbine and Wind Farm Measurements," *Energies*, vol. 13, no. 16, p. 4228, Aug. 2020.
 - [22] M. Zou, S. Zhu, J. Gu, L. M. Korunovic, and S. Z. Djokic, "Heating and Lighting Load Disaggregation Using Frequency Components and Convolutional Bidirectional Long Short-Term Memory Method," *Energies*, vol. 14, no. 16, p. 4831, Aug. 2021.
 - [23] M. Zou *et al.*, "Bayesian CNN-BiLSTM and Vine-GMCM Based Probabilistic Forecasting of Hour-Ahead Wind Farm Power Outputs," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1169-1187, Apr. 2022.
 - [24] D. Fang, M. Zou, and S. Djokic, "Probabilistic OPF Incorporating Uncertainties in Wind Power Outputs and Line Thermal Ratings," in *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Boise, United States, Jun. 2018.
 - [25] M. Zou, D. Fang, S. Djokic, and S. Hawkins, "Assessment of wind energy resources and identification of outliers in on-shore and off-shore wind farm measurements," presented at the 3rd International Conference on Offshore Renewable Energy (CORE), Glasgow, United Kingdom, Aug. 2018.
 - [26] D. Fang, J. Gunda, M. Zou, G. Harrison, S. Z. Djokic, and A. Vaccaro, "Dynamic Thermal Rating for Efficient Management of Post-Contingency Congestions," in *2019 IEEE Milan PowerTech*, Milan, Italy, Jun. 2019, pp. 1-6.
 - [27] S. Z. Djokic, M. Zou, D. Fang, V. D. Giorgio, R. Langella, and A. Testa, "On the Importance of Correlating Wind Speed and Wind Direction for Evaluating

- Uncertainty in Wind Turbine Power Output," in *2019 International Conference on Clean Electrical Power (ICCEP)*, Otranto, Italy, Jul. 2019, pp. 36-45.
- [28] M. Zou, D. Fang, S. Djokic, and S. Hawkins, "Equivalent power curves and short-term forecasting of power outputs of an off-shore wind farm based on a multi-state operational model," presented at the 4th International Conference on Offshore Renewable Energy (CORE), Glasgow, UK, Aug. 2019.
 - [29] M. Zou, D. Fang, G. Harrison, and S. Djokic, "Weather Based Day-Ahead and Week-Ahead Load Forecasting using Deep Recurrent Neural Network," in *5th IEEE Int. Forum on Research and Technology for Society and Industry (RTSI)*, Firenze, Italy, Sep. 2019, pp. 341-346.
 - [30] M. Zou *et al.*, "Comparison of Three Methods for a Weather Based Day-Ahead Load Forecasting," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Bucharest, Romania, Nov. 2019, pp. 1-5.
 - [31] D. Fang, M. Zou, J. Gunda, G. Harrison, S. Z. Djokic, and A. Vaccaro, "Multi-Stage Congestion Management Considering Maximum Lead Time and Voltage-Dependent Load Models," in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Oct. 2020, pp. 1000-1004.
 - [32] V. D. Giorgio, R. Langella, A. Testa, S. Z. Djokic, and M. Zou, "First Order Non-homogeneous Markov Chain Model for Generation of Wind Speed and Direction Synthetic Time Series," in *2020 International Conference Probabilistic Methods Applied to Power Systems (PMAPS)*, Liege, Belgium, Aug. 2020.
 - [33] D. Fang *et al.*, "Deterministic and Probabilistic Assessment of Distribution Network Hosting Capacity for Wind-Based Renewable Generation," in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Liege, Belgium, Aug. 2020, pp. 1-6.
 - [34] A. Zakaria, D. Fang, M. Zou, G. Harrison, and S. Z. Djokic, "Comparison of Deterministic and Probabilistic Approaches for Hosting Capacity Allocation of Wind and PV Generation in Distribution Networks," in *2021 IEEE Madrid PowerTech*, Madrid, Spain, Jun. 2021, pp. 1-6.
 - [35] Shuyang Zhu, Mingzhe Zou, and S. Djokic, "Evaluation of Uncertainty and Error of LSTM-Based Day-Ahead Load Forecasting Models," in *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Espoo, Finland, Oct. 2021.
 - [36] V. D. Giorgio, A. Testa, M. Zou, R. Langella, and S. Z. Djokic, "Simplified Markov Chain Models for Generation of Synthetic Time Series of Wind Speed and Direction," in *2022 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2022, accepted.
 - [37] A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?," presented at the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, USA, Dec. 2017.
 - [38] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896-913, 2016.

- [39] E. Ceperic, V. Ceperic, and A. Baric, "A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4356-4364, 2013.
- [40] H. Guo, "Accelerated Continuous Conditional Random Fields For Load Forecasting," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2023-2033, 2015.
- [41] G. Xie, X. Chen, and Y. Weng, "An Integrated Gaussian Process Modeling Framework for Residential Load Prediction," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7238-7248, 2018.
- [42] K. Y. Lee, Y. T. Cha, and J. H. Park, "Short-term load forecasting using an artificial neural network," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 124-132, 1992.
- [43] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv*, 2012. [Online]. Available: <https://arxiv.org/pdf/1207.0580.pdf>.
- [44] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *ICML '04 Proceedings of the twenty-first international conference on Machine learning*, 2004.
- [45] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, 2006.
- [46] G. Chen, "A Gentle Tutorial of Recurrent Neural Network with Error Backpropagation," *arXiv*, 2016. [Online]. Available: <https://arxiv.org/pdf/1610.02583.pdf>.
- [47] M. C. Mozer, "A focused backpropagation algorithm for temporal pattern recognition," in *Backpropagation: Theory, architectures, and applications*: L. Erlbaum Associates Inc., 1995, pp. 137-169.
- [48] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [49] M. Schuster and K. K. Paliwal, "Bidirectional Recurrent Neural Networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673-2681, 1997.
- [50] M. Hermans and B. Schrauwen, "Training and analyzing deep recurrent neural networks," presented at the Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, 2013.
- [51] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges [N-dimensional joint and marginal distribution functions]*. Paris, France: Université Paris 8, 1959, pp. 229-231.
- [52] M. Sun, I. Konstantelos, and G. Strbac, "C-Vine Copula Mixture Model for Clustering of Residential Electrical Load Pattern Data," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2382-2393, 2017.
- [53] Z. Wang, W. Wang, C. Liu, Z. Wang, and Y. Hou, "Probabilistic Forecast for Multiple Wind Farms Based on Regular Vine Copulas," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 578-589, 2018.
- [54] M. Cui, V. Krishnan, B. Hodge, and J. Zhang, "A Copula-Based Conditional Probabilistic Forecast Model for Wind Power Ramps," *IEEE Transactions on Smart Grid*, pp. 1-1, 2018.

- [55] 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.
- [56] G. Papaefthymiou and D. Kurowicka, "Using Copulas for Modeling Stochastic Dependence in Power System Uncertainty Analysis," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 40-49, 2009.
- [57] Q. Lu, W. Hu, Y. Min, F. Yuan, and Z. Gao, "Wind power uncertainty modeling considering spatial dependence based on Pair-copula theory," in *2014 IEEE PES General Meeting | Conference & Exposition*, Washington, DC, United States, 2014, pp. 1-5.
- [58] Y. Wang, N. Zhang, C. Kang, M. Miao, R. Shi, and Q. Xia, "An Efficient Approach to Power System Uncertainty Analysis With High-Dimensional Dependencies," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2984-2994, 2018.
- [59] Z. Q. Xie, T. Y. Ji, M. S. Li, and Q. H. Wu, "Quasi-Monte Carlo Based Probabilistic Optimal Power Flow Considering the Correlation of Wind Speeds Using Copula Function," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2239-2247, 2018.
- [60] H. Joe, *Multivariate Models and Multivariate Dependence Concepts*. Taylor & Francis, 1997.
- [61] Y. Lin, M. Yang, C. Wan, J. Wang, and Y. Song, "A Multi-Model Combination Approach for Probabilistic Wind Power Forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 226-237, 2019.
- [62] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. The Wiley Finance Series, 2006.
- [63] S. Kyung-Bin, B. Young-Sik, H. Dug Hun, and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 96-101, 2005.
- [64] Y. T. Quek, W. L. Woo, and T. Logenthiran, "Load Disaggregation Using One-Directional Convolutional Stacked Long Short-Term Memory Recurrent Neural Network," *IEEE Systems Journal*, vol. 14, no. 1, pp. 1395-1404, 2020.
- [65] J. D. Kelly, "Disaggregation of Domestic Smart Meter Energy Data," PhD thesis, University of London, 2017.
- [66] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-Intrusive Load Disaggregation Using Graph Signal Processing," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1739-1747, 2018.
- [67] V. S. Hana Altrabalsi, J. Liao, L. Stankovic, "Low-complexity energy disaggregation using appliance load modelling," *AIMS Energy*, vol. 4, no. 1, pp. 884-905, 2016.
- [68] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2014, pp. 535-540.
- [69] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870-1891, 1992.
- [70] J. Z. Kolter and T. Jaakkola, "Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation," presented at the

Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, 2012.

- [71] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models," *Applied Energy*, vol. 208, pp. 1590-1607, 2017.
- [72] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Transactions on Power Delivery*, vol. 10, no. 3, pp. 1200-1210, 1995.
- [73] H.-H. Chang, "Non-Intrusive Demand Monitoring and Load Identification for Energy Management Systems Based on Transient Feature Analyses," vol. 5, no. 11, pp. 4569-4589, 2012.
- [74] J. Kim, T.-T.-H. Le, and H. Kim, "Nonintrusive Load Monitoring Based on Advanced Deep Learning and Novel Signature," *Computational Intelligence and Neuroscience*, vol. 2017, p. 4216281, 2017.
- [75] J. Kelly and W. Knottenbelt, "Neural NILM: Deep Neural Networks Applied to Energy Disaggregation," presented at the Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, Seoul, South Korea, 2015.
- [76] N. Sadeghianpourhamami, J. Ruysinck, D. Deschrijver, T. Dhaene, and C. Develder, "Comprehensive feature selection for appliance classification in NILM," *Energy and Buildings*, vol. 151, pp. 98-106, 2017.
- [77] P. A. Schirmer and I. Mporas, "Statistical and Electrical Features Evaluation for Electrical Appliances Energy Disaggregation," vol. 11, no. 11, p. 3222, 2019.
- [78] D. Villanueva and A. Feijóo, "Comparison of logistic functions for modeling wind turbine power curves," *Electric Power Systems Research*, vol. 155, pp. 281-288, 2018.
- [79] T. Ouyang, A. Kusiak, and Y. He, "Modeling wind-turbine power curve: A data partitioning and mining approach," *Renewable Energy*, vol. 102, pp. 1-8, 2017.
- [80] M. Lydia, A. I. Selvakumar, S. S. Kumar, and G. E. P. Kumar, "Advanced Algorithms for Wind Turbine Power Curve Modeling," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 3, pp. 827-835, 2013.
- [81] V. Sohoni, S. C. Gupta, and R. K. Nema, "A Critical Review on Wind Turbine Power Curve Modelling Techniques and Their Applications in Wind Based Energy Systems," *Journal of Energy*, vol. 2016, 2016.
- [82] E. Taslimi-Renani, M. Modiri-Delshad, M. F. M. Elias, and N. A. Rahim, "Development of an enhanced parametric model for wind turbine power curve," *Applied Energy*, vol. 177, pp. 544-552, 2016.
- [83] P. G. Gottschalk and J. R. Dunn, "The five-parameter logistic: A characterization and comparison with the four-parameter logistic," *Analytical Biochemistry*, vol. 343, no. 1, pp. 54-65, 2005.
- [84] Y. Wang, Q. Hu, L. Li, A. M. Foley, and D. Srinivasan, "Approaches to wind power curve modeling: A review and discussion," *Renewable and Sustainable Energy Reviews*, vol. 116, p. 109422, 2019.

- [85] V. Thapar, G. Agnihotri, and V. K. Sethi, "Critical analysis of methods for mathematical modelling of wind turbines," *Renewable Energy*, vol. 36, no. 11, pp. 3166-3177, 2011.
- [86] K. H. Hohenemser and A. H. P. Swift, "Wind Turbine Speed Control by Automatic Yawing," *Journal of Energy*, vol. 7, no. 3, pp. 237-242, 1983.
- [87] Vestas., "Vestas Wind Systems A/S, Report 950010.R1: General Specification V90–3.0 MW," Vestas, Aarhus, Denmark, 2013.
- [88] C. Gallego, P. Pinson, H. Madsen, A. Costa, and A. Cuerva, "Influence of local wind speed and direction on wind power dynamics – Application to offshore very short-term forecasting," *Applied Energy*, vol. 88, no. 11, pp. 4087-4096, 2011.
- [89] D. Yuyu, Z. Hai, T. Zhiping, C. Ying, and D. Jie, "The influence of wind direction on short-term wind power prediction: A case study in north China," in *IEEE PES Innovative Smart Grid Technologies*, 2012, pp. 1-5.
- [90] Y. Dvorkin, M. Lubin, S. Backhaus, and M. Chertkov, "Uncertainty Sets for Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3326-3327, 2016.
- [91] L. Han, R. Zhang, X. Wang, A. Bao, and H. Jing, "Multi-step wind power forecast based on VMD-LSTM," vol. 13, no. 10, pp. 1690-1700, 2019.
- [92] J. L. Wu, T. Y. Ji, M. S. Li, P. Z. Wu, and Q. H. Wu, "Multistep Wind Power Forecast Using Mean Trend Detector and Mathematical Morphology-Based Local Predictor," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1216-1223, 2015.
- [93] J. Wang, M. Shahidehpour, and Z. Li, "Security-Constrained Unit Commitment With Volatile Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1319-1327, 2008.
- [94] M. A. Matos and R. J. Bessa, "Setting the Operating Reserve Using Probabilistic Wind Power Forecasts," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 594-603, 2011.
- [95] D. Lee, H. Shin, and R. Baldick, "Bivariate Probabilistic Wind Power and Real-Time Price Forecasting and Their Applications to Wind Power Bidding Strategy Development," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6087-6097, 2018.
- [96] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1033-1044, 2014.
- [97] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 61, no. 4, pp. 555-576, 2012.
- [98] Y. Yu, X. Han, M. Yang, and J. Yang, "Probabilistic Prediction of Regional Wind Power Based on Spatiotemporal Quantile Regression," *IEEE Transactions on Industry Applications*, vol. 56, no. 6, pp. 6117-6127, 2020.
- [99] A. U. Haque, M. H. Nehrir, and P. Mandal, "A Hybrid Intelligent Model for Deterministic and Quantile Regression Approach for Probabilistic Wind Power Forecasting," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1663-1672, 2014.

- [100] L. Yang, M. He, J. Zhang, and V. Vittal, "Support-Vector-Machine-Enhanced Markov Model for Short-Term Wind Power Forecast," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 791-799, 2015.
- [101] M. J. Sanjari, H. B. Gooi, and N. C. Nair, "Power Generation Forecast of Hybrid PV–Wind System," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 703-712, 2020.
- [102] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind Power Density Forecasting Using Ensemble Predictions and Time Series Models," *IEEE Transactions on Energy Conversion*, vol. 24, no. 3, pp. 775-782, 2009.
- [103] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425-464, 2001.
- [104] M. H. DeGroot and M. J. Schervish, *Probability and Statistics, Fourth Edition*. Pearson Education, 2013.
- [105] S. Benninga and B. Czaczkes, *Financial Modeling, Fourth Edition*. MIT Press, 2014.
- [106] A. Paisios, "Profiling and disaggregation of electricity demands measured in MV distribution networks," PhD thesis, The University of Edinburgh, 2017.
- [107] MetOffice., "MIDAS Open: UK hourly weather observation data, v201908," Centre for Environmental Data Analysis (CEDA), 2019.
- [108] C. A. Hall and W. W. Meyer, "Optimal error bounds for cubic spline interpolation," *Journal of Approximation Theory*, vol. 16, no. 2, pp. 105-122, 1976.
- [109] "Scottish bank holidays," Scottish Government, 2019. [Online]. Available: <https://www.gov.scot/publications/bank-holidays/>.
- [110] J. Zhang, Y. Lai, and J. Lin, "The day-of-the-Week effects of stock markets in different countries," *Finance Research Letters*, vol. 20, pp. 47-62, 2017.
- [111] J. LeSage, "Analyzing the Sun's Path," MATLAB Central File Exchange, 2016. [Online]. Available: <https://uk.mathworks.com/matlabcentral/fileexchange/55510-analyzing-the-sun-s-path>.
- [112] M. Zimecki, "The lunar cycle: Effects on human and animal behavior and physiology," *Postępy Higieny i Medycyny Doświadczalnej (Advances in Hygiene and Experimental Medicine)*, vol. 60, pp. 1-7, 2006.
- [113] M. Mahooti, "Moon Phases," MATLAB Central File Exchange, 2019. [Online]. Available: <https://www.mathworks.com/matlabcentral/fileexchange/55270-moon-phases>.
- [114] D. R. Satterlund, R. C. Chapman, and R. D. Beach, "Modeling the Daily Temperature Cycle," *Northwest Science*, vol. 57, pp. 22-31, 1983.
- [115] S. Makonin, B. Ellert, I. V. Bajić, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Scientific Data*, vol. 3, no. 1, p. 160037, 2016.
- [116] R. Gelaro *et al.*, "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)," *Journal of climate*, vol. Volume 30, no. Iss 13, pp. 5419-5454, 2017.
- [117] G. Papaefthymiou and B. Klockl, "MCMC for Wind Power Simulation," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 234-240, 2008.

- [118] D. Harris and S. Harris, *Digital Design and Computer Architecture, Second Edition*. Morgan Kaufmann Publishers Inc., 2012, p. 712.
- [119] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [120] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980*, 2015.
- [121] M. Shcherbakov, A. Brebels, N. L. Shcherbakova, A. Tyukov, T. A. Janovsky, and V. A. Kamaev, *A survey of forecast error measures*. 2013, pp. 171-176.
- [122] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 35, no. 4, pp. 476-487, 2005.
- [123] D. Opitz and R. Maclin, "Popular ensemble methods: an empirical study," *Journal of Artificial Intelligence Research*, vol. 11, no. 1, pp. 169-198, 1999.
- [124] L. Mason, J. Baxter, P. Bartlett, and M. Frean, "Boosting algorithms as gradient descent," presented at the Proceedings of the 12th International Conference on Neural Information Processing Systems, Denver, CO, 1999.
- [125] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [126] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, no. 1, p. 150007, 2015.
- [127] T. Kato and M. Uemura, "Period Analysis using the Least Absolute Shrinkage and Selection Operator (Lasso)," *Publications of the Astronomical Society of Japan*, vol. 64, no. 6, 2012.
- [128] A. Incremona and G. D. Nicolao, "Spectral Characterization of the Multi-Seasonal Component of the Italian Electric Load: A LASSO-FFT Approach," *IEEE Control Systems Letters*, vol. 4, no. 1, pp. 187-192, 2020.
- [129] G. J. A. M. M. Goertzel, "An Algorithm for the Evaluation of Finite Trigonometric Series," *American Mathematical Monthly*, vol. 65, p. 34, 1958.
- [130] D. Wu, X. Wang, J. Su, B. Tang, and S. Wu, "A Labeling Method for Financial Time Series Prediction Based on Trends," *Entropy*, vol. 22, no. 10, p. 1162, 2020.
- [131] S. Mahfoud and G. Mani, "Financial forecasting using genetic algorithms " *Applied Artificial Intelligence*, vol. 10, no. 6, pp. 543-566, 1996.
- [132] J. Hauke and T. Kossowski, *Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data*. 2011, pp. 87-93.
- [133] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 4580-4584.
- [134] D. K. Duvenaud, "Automatic model construction with Gaussian processes," PhD thesis, University of Cambridge, 2014.
- [135] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: no regret and experimental design," presented at the Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 2010.
- [136] N. Batra *et al.*, "NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring," in *5th International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.

- [137] S. J. Watson, B. J. Xiang, W. Yang, P. J. Tavner, and C. J. Crabtree, "Condition Monitoring of the Power Output of Wind Turbine Generators Using Wavelets," *IEEE Transactions on Energy Conversion*, vol. 25, no. 3, pp. 715-721, 2010.
- [138] Z. Zhang, Y. Sun, D. W. Gao, J. Lin, and L. Cheng, "A Versatile Probability Distribution Model for Wind Power Forecast Errors and Its Application in Economic Dispatch," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3114-3125, 2013.
- [139] M. Ghofrani, A. Arabali, M. Etezadi-Amoli, and M. S. Fadali, "Energy Storage Application for Performance Enhancement of Wind Integration," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4803-4811, 2013.
- [140] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. Prem Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 452-460, 2014.
- [141] C. Carrillo, M. A. F. Obando, J. Cidrás, and E. Díaz-Dorado, "Review of power curve modelling for wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 21, pp. 572-581, 2013/05/01/ 2013.
- [142] B. P. Hayes, I. Ilie, A. Porpodas, S. Z. Djokic, and G. Chicco, "Equivalent power curve model of a wind farm based on field measurement data," in *2011 IEEE Trondheim PowerTech*, Trondheim, Norway, 2011, pp. 1-7.
- [143] L. Ye, Y. Zhao, C. Zeng, and C. Zhang, "Short-term wind power prediction based on spatial model," *Renewable Energy*, vol. 101, pp. 1067-1074, 2017.
- [144] M. Xu, P. Pinson, Z. Lu, Y. Qiao, and Y. Min, "Adaptive robust polynomial regression for power curve modeling with application to wind power forecasting," *Wind Energy*, vol. 19, no. 12, pp. 2321-2336, 2016.
- [145] S. Gill, B. Stephen, and S. Galloway, "Wind Turbine Condition Assessment Through Power Curve Copula Modeling," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 94-101, 2012.
- [146] A. Kusiak, H. Zheng, and Z. Song, "Models for monitoring wind farm power," *Renewable Energy*, vol. 34, no. 3, pp. 583-590, 2009.
- [147] A. Kusiak and A. Verma, "Monitoring Wind Farms With Performance Curves," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 192-199, 2013.
- [148] S. Sarkar and V. Ajjarapu, "MW Resource Assessment Model for a Hybrid Energy Conversion System With Wind and Solar Resources," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 4, pp. 383-391, 2011.
- [149] D. Dhungana and R. Karki, "Data Constrained Adequacy Assessment for Wind Resource Planning," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 219-227, 2015.
- [150] S. Wang, X. Zhang, L. Ge, and L. Wu, "2-D Wind Speed Statistical Model for Reliability Assessment of Microgrid," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1159-1169, 2016.
- [151] E. Sainz, A. Llombart, and J. J. Guerrero, "Robust filtering for the characterization of wind turbines: Improving its operation and maintenance," *Energy Conversion and Management*, vol. 50, no. 9, pp. 2136-2147, 2009.
- [152] S. Wang, Y. Huang, L. Li, and C. Liu, "Wind turbines abnormality detection through analysis of wind farm power curves," *Measurement*, vol. 93, pp. 178-188, 2016.

- [153] M. Javadi, A. M. Malyscheff, D. Wu, C. Kang, and J. N. Jiang, "An algorithm for practical power curve estimation of wind turbines," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 1, pp. 93-102, 2018.
- [154] M. Schlechtingen and I. Ferreira Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1849-1875, 2011.
- [155] B. Stephen, S. J. Galloway, D. McMillan, D. C. Hill, and D. G. Infield, "A Copula Model of Wind Turbine Performance," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 965-966, 2011.
- [156] X. Shen, X. Fu, and C. Zhou, "A Combined Algorithm for Cleaning Abnormal Data of Wind Turbine Power Curve Based on Change Point Grouping Algorithm and Quartile Algorithm," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 46-54, 2019.
- [157] Y. Zhao, Y. Liu, and R. Wang, "Fuzzy scalar quantisation based on hidden Markov model and application in fault diagnosis of wind turbine," *The Journal of Engineering*, vol. 2017, no. 14, pp. 2685-2689, 2017.
- [158] J. Park, J. Lee, K. Oh, and J. Lee, "Development of a Novel Power Curve Monitoring Method for Wind Turbines and Its Field Tests," *IEEE Transactions on Energy Conversion*, vol. 29, no. 1, pp. 119-128, 2014.
- [159] N. Yampikulsakul, E. Byon, S. Huang, S. Sheng, and M. You, "Condition Monitoring of Wind Power System With Nonparametric Regression Analysis," *IEEE Transactions on Energy Conversion*, vol. 29, no. 2, pp. 288-299, 2014.
- [160] X. Xu, Y. Lei, and Z. Li, "An Incorrect Data Detection Method for Big Data Cleaning of Machinery Condition Monitoring," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2326-2336, 2020.
- [161] P. Cambron, R. Lepvrier, C. Masson, A. Tahan, and F. Pelletier, "Power curve monitoring using weighted moving average control charts," *Renewable Energy*, vol. 94, pp. 126-135, 2016.
- [162] A. Kusiak, H. Zheng, and Z. Song, "On-line monitoring of power curves," *Renewable Energy*, vol. 34, pp. 1487-1493, 2009.
- [163] 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.
- [164] E. Mangalova and E. Agafonov, "Wind power forecasting using the k-nearest neighbors algorithm," *International Journal of Forecasting*, vol. 30, no. 2, pp. 402-406, 2014.
- [165] C. Lowery and M. O. Malley, "Impact of Wind Forecast Error Statistics Upon Unit Commitment," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 760-768, 2012.
- [166] Y. Liu, Q. Liu, S. Han, J. Yan, and L. Li, "The detection method of wind turbine operation outliers based on multi-dimensional cluster," in *8th Renewable Power Generation Conference (RPG 2019)*, Shanghai, China, 2019, pp. 1-7.
- [167] L. Zheng, W. Hu, and Y. Min, "Raw Wind Data Preprocessing: A Data-Mining Approach," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 11-19, 2015.

- [168] Y. Wang, D. G. Infield, B. Stephen, and S. J. Galloway, "Copula based model for wind turbine power curve outlier rejection," *Wind Energy*, Article vol. 17, no. 11, pp. 1677-1688, 2014.
- [169] Y. Hu, Y. Qiao, J. Liu, and H. Zhu, "Adaptive Confidence Boundary Modeling of Wind Turbine Power Curve Using SCADA Data and Its Application," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1330-1341, 2019.
- [170] M. Mohammadpourfard, A. Sami, and Y. Weng, "Identification of False Data Injection Attacks With Considering the Impact of Wind Generation and Topology Reconfigurations," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1349-1364, 2018.
- [171] M. Yesilbudak, "Partitional clustering-based outlier detection for power curve optimization of wind turbines," in *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, Birmingham, UK, 2016, pp. 1080-1084.
- [172] Y. Zhou *et al.*, "A semi-supervised anomaly detection method for wind farm power data preprocessing," in *2017 IEEE Power & Energy Society General Meeting*, Chicago, IL, United States, 2017, pp. 1-5.
- [173] Z. Sun and H. Sun, "Stacked Denoising Autoencoder With Density-Grid Based Clustering Method for Detecting Outlier of Wind Turbine Components," *IEEE Access*, vol. 7, pp. 13078-13091, 2019.
- [174] X. Ye, Z. Lu, Y. Qiao, Y. Min, and M. O. Malley, "Identification and Correction of Outliers in Wind Farm Time Series Power Data," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4197-4205, 2016.
- [175] H. Long, L. Sang, Z. Wu, and W. Gu, "Image-Based Abnormal Data Detection and Cleaning Algorithm via Wind Power Curve," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 938-946, 2020.
- [176] Y.-H. Wan, E. Ela, and K. Orwig, "Development of an equivalent wind plant power curve," in *Windpower 2010 Conference & Exhibition*, Dallas, Texas, United States, 2010.
- [177] L. T. Paiva, R. C. Veiga, and J. M. L. M. Palma, "Determining wind turbine power curves based on operating conditions," *Wind Energy*, vol. 17, no. 10, pp. 1563-1575, 2014.
- [178] H. Chen, F. Li, and Y. Wang, "Wind power forecasting based on outlier smooth transition autoregressive GARCH model," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 3, pp. 532-539, 2018.
- [179] M. Morshedizadeh, M. Kordestani, R. Cariveau, D. S. Ting, and M. Saif, "Power production prediction of wind turbines using a fusion of MLP and ANFIS networks," *IET Renewable Power Generation*, vol. 12, no. 9, pp. 1025-1033, 2018.
- [180] W. K. Härdle and L. Simar, "Principal Components Analysis," in *Applied Multivariate Statistical Analysis* Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 319-358.
- [181] A. Albers, "Turbulence and Shear Normalisation of Wind Turbine Power Curve," in *European Wind Energy Conference and Exhibition (EWEC)*, Warsaw, Poland, 2010.
- [182] A. Albers, T. Jakobi, R. Rohden, and J. Stoltenjohannes, "Influence of meteorological variables on measured wind turbine power curves," in

European Wind Energy Conference and Exhibition 2007 (EWEC), Milan, Italy, 2007.

- [183] G. King, J. Honaker, A. Joseph, and K. Scheve, "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *American Political Science Review*, vol. 95, no. 1, 2001.
- [184] J. B. Gayo, "Reliability focused research on optimizing Wind Energy systems design, operation and maintenance: Tools, proof of concepts, guidelines & methodologies for a new generation (Reliawind)," Gamesa Innovation and Technology, Madrid, Spain, 2011.
- [185] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier Detection for Temporal Data: A Survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250-2267, 2014.
- [186] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, p. Article 15, 2009.
- [187] J. W. Tukey, *Exploratory data analysis*. Boston, MA, United States: Reading, Mass. : Addison-Wesley Pub. Co., 1977.
- [188] 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.
- [189] Y. Zhou, A. T. K. Wan, S. Xie, and X. Wang, "Wavelet analysis of change-points in a non-parametric regression with heteroscedastic variance," *Journal of Econometrics*, vol. 159, no. 1, pp. 183-201, 2010.
- [190] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, USA, 1996.
- [191] J. Gao, W. Hu, W. Li, Z. Zhang, and O. Wu, "Local Outlier Detection Based on Kernel Regression," in *2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey, 2010, pp. 585-588.
- [192] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," presented at the Proceedings of the 2000 ACM SIGMOD international conference on Management of data, Renton, WA, USA, 19–21 October, 2000.
- [193] Ruiyu Jia and Y. Li, "K-means algorithm of clustering number and centers self-determination," *Computer Engineering and Applications*, vol. 54, no. 7, pp. 152 - 158, 2018.
- [194] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough Deep Neural Architecture for Short-Term Wind Speed Forecasting," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2770-2779, 2017.
- [195] J. Lu and Q. Zhu, "An Effective Algorithm Based on Density Clustering Framework," *IEEE Access*, vol. 5, pp. 4991-5000, 2017.
- [196] B. Wu and B. M. Wilamowski, "A Fast Density and Grid Based Clustering Method for Data With Arbitrary Shapes and Noise," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1620-1628, 2017.
- [197] I. H. Witten, E. Frank, and M. A. Hall, "Chapter 3 - Output: Knowledge Representation," in *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*, I. H. Witten, E. Frank, and M. A. Hall, Eds. Boston, United States: Morgan Kaufmann, 2011, pp. 61-83.

- [198] A. Mitra, *Fundamentals of Quality Control and Improvement, Third Edition*. New Jersey, United States: John Wiley & Sons, Inc, 2008.
- [199] A. Tewari, M. J. Giering, and A. Raghunathan, "Parametric Characterization of Multimodal Distributions with Non-gaussian Modes," in *2011 IEEE 11th International Conference on Data Mining Workshops*, Vancouver, BC, Canada, 2011, pp. 286-292: IEEE Computer Society 1730 Massachusetts Ave., NW Washington, DC, USA.
- [200] R. G. Brown and R. F. Meyer, "The Fundamental Theorem of Exponential Smoothing," *Operations Research*, vol. 9, no. 5, pp. 673-685, 1961.
- [201] C. A. Markowski and E. P. Markowski, "Conditions for the Effectiveness of a Preliminary Test of Variance," *The American Statistician*, vol. 44, no. 4, pp. 322-326, 1990.
- [202] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191-203, 1984.
- [203] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764-766, 2013/07/01/ 2013.
- [204] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image Analysis Using Mathematical Morphology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 532-550, 1987.
- [205] H. Ming-Kuei, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179-187, 1962.
- [206] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129-136, 1982.
- [207] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *Neurocomputing*, vol. 48, no. 1, pp. 85-105, Oct, 2002.
- [208] H. A. David, "Early Sample Measures of Variability," *Statistical Science*, vol. 13, no. 4, pp. 368-377, 1998.
- [209] Q. Lin and J. Wang, "Vertically Correlated Echelon Model for the Interpolation of Missing Wind Speed Data," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 804-812, 2014.
- [210] Z. Yang, Y. Liu, and C. Li, "Interpolation of missing wind data based on ANFIS," *Renewable Energy*, vol. 36, no. 3, pp. 993-998, 2011/03/01/ 2011.
- [211] Y. Zhang, S. Kim, and G. B. Giannakis, "Short-term wind power forecasting using nonnegative sparse coding," in *2015 49th Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, Maryland, United States, 2015, pp. 1-5.
- [212] B. Efron, "Missing Data, Imputation, and the Bootstrap," *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 463-475, 1994.
- [213] H. Takagi and I. Hayashi, "NN-driven fuzzy reasoning," *International Journal of Approximate Reasoning*, vol. 5, no. 3, pp. 191-212, 1991/05/01/ 1991.
- [214] J. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665-685, 1993.

- [215] C. Sherlock, P. Fearnhead, and G. O. Roberts, "The Random Walk Metropolis: Linking Theory and Practice Through a Case Study," (in en), *Statist. Sci.*, vol. 25, no. 2, pp. 172-190, 2010.
- [216] C. G. Anderson, "North Harris Community Wind Farm: Wind Measurements at Moana Site," North Harris Trading Company Ltd, Perth, United Kingdom, 4, May 2007.
- [217] O. Bozorg-Haddad, M. Solgi, and H. A. Loáiciga, "Genetic Algorithm," in *Meta - Heuristic and Evolutionary Algorithms for Engineering Optimization* Hoboken, New Jersey, United States: John Wiley & Sons, Inc., 2017, pp. 53-67.
- [218] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *arXiv:1708.02002 [cs.CV]*, pp. 2999-3007, 2017.
- [219] I. Pobocikova and Z. J. A. m. s. Sedlib ková, "Comparison of four methods for estimating the Weibull distribution parameters," *Applied mathematical sciences*, vol. 8, pp. 4137-4149, 2014.
- [220] P. Bhattacharya and R. Bhattacharjee, "A Study on Weibull Distribution for Estimating the Parameters," *Wind Engineering*, vol. 33, no. 5, pp. 469-476, 2009.
- [221] P. Ramírez and J. A. Carta, "Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study," *Energy Conversion and Management*, vol. 46, no. 15, pp. 2419-2438, 2005.
- [222] F. N. Nwobi and C. A. Ugomma, "A Comparison of Methods for the Estimation of Weibull Distribution Parameters," *Metodološki zvezki*, vol. 11, no. 1, pp. 65-78, 2014.
- [223] G. W. Cran, "Moment estimators for the 3-parameter Weibull distribution," *IEEE Transactions on Reliability*, vol. 37, no. 4, pp. 360-363, 1988.
- [224] J. Heckenbergerova, P. Musilek, and P. Krömer, "Optimization of Wind Direction Distribution Parameters Using Particle Swarm Optimization," in *Afro-European Conference for Industrial Advancement*, Cham, 2015, pp. 15-26.
- [225] A. Carpinone, M. Giorgio, R. Langella, and A. Testa, "Markov chain modeling for very-short-term wind power forecasting," *Electric Power Systems Research*, vol. 122, pp. 152-158, May 2015.
- [226] Y. Wang, H. Wang, D. Srinivasan, and Q. Hu, "Robust functional regression for wind speed forecasting based on Sparse Bayesian learning," *Renewable Energy*, vol. 132, pp. 43-60, 2019.
- [227] M. Cui, V. Krishnan, B. Hodge, and J. Zhang, "A Copula-Based Conditional Probabilistic Forecast Model for Wind Power Ramps," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3870-3882, 2019.
- [228] R. Blonbou, "Very short-term wind power forecasting with neural networks and adaptive Bayesian learning," *Renewable Energy*, vol. 36, no. 3, pp. 1118-1124, 2011.
- [229] K. B. Sahay and P. Shukla, "One Hour-Ahead Wind Speed Forecasting of Sotavento Galicia SA Wind Farm by Using ANN," in *2018 International Electrical Engineering Congress (iEECON)*, 2018, pp. 1-4.
- [230] R. Mbuva, M. Jonsson, N. Ehn, and P. Herman, "Bayesian neural networks for one-hour ahead wind power forecasting," in *2017 IEEE 6th International*

- Conference on Renewable Energy Research and Applications (ICRERA)*, 2017, pp. 591-596.
- [231] H. Wen, J. Gu, J. Ma, and Z. Jin, "Probabilistic Wind Power Forecasting via Bayesian Deep Learning Based Prediction Intervals," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, 2019, vol. 1, pp. 1091-1096.
 - [232] Y. Liu *et al.*, "Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model," *Applied Energy*, vol. 260, p. 114259, 2020.
 - [233] R. J. Bessa, "On the quality of the Gaussian copula for multi-temporal decision-making problems," in *2016 Power Systems Computation Conference (PSCC)*, 2016, pp. 1-7.
 - [234] M. Zou *et al.*, 18 Jan 2022, "Bayesian CNN-BiLSTM and Vine-GMCM Based Probabilistic Forecasting of Hour-Ahead Wind Farm Power Outputs (Input dataset)", IEEE DataPort, <https://dx.doi.org/10.21227/y2rj-xb73>.
 - [235] Z. Sun and M. Zhao, "Short-Term Wind Power Forecasting Based on VMD Decomposition, ConvLSTM Networks and Error Analysis," *IEEE Access*, vol. 8, pp. 134422-134434, 2020.
 - [236] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight Uncertainty in Neural Network," presented at the 32nd International Conference on Machine Learning (ICML), 2015.
 - [237] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," presented at the The 33rd International Conference on Machine Learning (ICML), 2016.
 - [238] Y. Wen, P. Vicol, J. Ba, D. Tran, and R. B. J. A. Grosse, "Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches," presented at the Sixth International Conference on Learning Representations (ICLR), 2018.
 - [239] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79-86, 1951.
 - [240] G. Malsiner-Walli and H. Wagner, "Comparing Spike and Slab Priors for Bayesian Variable Selection," *Austrian Journal of Statistics*, vol. 40, no. 4, pp. 241-264, 2016.
 - [241] T. M. Letcher, *Wind Energy Engineering: A Handbook for Onshore and Offshore Wind Turbines*. Elsevier Science, 2017.
 - [242] T. M. Beasley, S. Erickson, and D. B. Allison, "Rank-based inverse normal transformations are increasingly used, but are they merited?," *Behavior genetics*, vol. 39, no. 5, pp. 580-595, June, 2009.
 - [243] N. I. Fisher, *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press, 1993.
 - [244] J. Ma, M. Yang, and Y. Lin, "Ultra-Short-Term Probabilistic Wind Turbine Power Forecast Based on Empirical Dynamic Modeling," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 906-915, 2020.
 - [245] J. Wolff, L. Nance, L. Bernardet, and B. Brown, "WRF Reference Configurations," presented at the 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction, Omaha, US, 2009.
 - [246] V. Sundarapandian, *Probability, Statistics and Queuing Theory*. New Delhi, India: PHI Learning, 2009.

- [247] S. Tewari, C. J. Geyer, and N. Mohan, "A Statistical Model for Wind Power Forecast Error and its Application to the Estimation of Penalties in Liberalized Markets," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2031-2039, 2011.
- [248] M. Santhosh, C. Venkaiah, and D. M. Vinod Kumar, "Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review," vol. 2, no. 6, p. e12178, 2020.
- [249] U. Focken, M. Lange, K. Mönnich, H.-P. Waldl, H. G. Beyer, and A. Luig, "Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 3, pp. 231-246, 2002.
- [250] L. Landberg, "Short-term prediction of the power production from wind farms," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 80, no. 1, pp. 207-220, 1999.
- [251] S. Scher and J. Molinder, "Machine Learning-Based Prediction of Icing-Related Wind Power Production Loss," *IEEE Access*, vol. 7, pp. 129421-129429, 2019.
- [252] Y. Zhang, C. Qin, A. K. Srivastava, C. Jin, and R. K. Sharma, "Data-Driven Day-Ahead PV Estimation Using Autoencoder-LSTM and Persistence Model," *IEEE Transactions on Industry Applications*, vol. 56, no. 6, pp. 7185-7192, 2020.
- [253] C. A. Hans and E. Klages, "Very Short Term Time-Series Forecasting of Solar Irradiance Without Exogenous Inputs," *arXiv:1810.07066*, 2018.
- [254] "Global Energy Forecasting Competition 2014 Probabilistic Wind Power Forecasting," CrowdANALYTIX, Available: <https://www.crowdanalytix.com/contests/global-energy-forecasting-competition-2014-probabilistic-wind-power-forecasting>, Accessed on: 22 July, 2021.
- [255] Y. Zhang and S. You, "Chapter 5 - Use probabilistic forecasting to model uncertainties in electricity markets—A wind power example," in *Mathematical Modelling of Contemporary Electricity Markets*, A. Dagoumas, Ed.: Academic Press, 2021, pp. 75-94.
- [256] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1074-1080, 2016.
- [257] R. Juban, H. Ohlsson, M. Maasoumy, L. Poirier, and J. Z. Kolter, "A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1094-1102, 2016.
- [258] M. Landry, T. P. Erlinger, D. Patschke, and C. Varrichio, "Probabilistic gradient boosting machines for GEFCom2014 wind forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1061-1066, 2016.
- [259] K. Hatalis, A. Lamadrid, K. Scheinberg, and S. Kishore, "Smooth Pinball Neural Network for Probabilistic Forecasting of Wind Power," *arXiv:1710.01720*, 2017.