Uncertainty Analysis and Application on Smart Homes and Smart Grids: Big Data Approaches

Shi, Heng

Award date:
2018

Awarding institution:
University of Bath

Uncertainty Analysis and Application on Smart Homes and Smart Grids: Big Data Approaches

By

Heng Shi

BEng, sMIEEE

The thesis submitted for the degree of

Doctor of Philosophy

in

The Department of
Electronic and Electrical Engineering
University of Bath

May 2018

Signature:…… Date:……

Contents

Contents ii

Abstract vii

Acknowledgement ix

List of Figures x

List of Tables xiii

List of Abbreviations xiv

Chapter 1 1

1.1 Background and Motivation 2

1.1.1 New Context: Growing Low Carbon Technology 2

1.1.2 New Challenge: Uncertainty in the Power System 3

1.2 Research Contributions of Thesis 5

1.3 Structure of Thesis 7

Chapter 2 9

2.1 Introduction 10

2.2 What is Uncertainty? 10

2.3 Uncertainty Categories and Sources 12

2.4 How to measure uncertainty 13

2.4.1 Moment-based Uncertainty Measures 13

2.4.2 Probability-based Uncertainty Measures 16

2.4.3 Other Uncertainty Measures 16

2.5 Uncertainty Quantification Problems 17

2.6 Reviews on Prior Power System Uncertainty Research 20

2.6.1 Literature of Uncertainty Quantification 20

2.6.2 The literature on Power System Applications considering Uncertainty 21

2.7 Chapter Summary 22

Chapter 3 24

3.1 Introduction 25

3.2 Methodology 26

3.2.1 First Stage: Uncertainty Temporal Components Decomposition, Reconstruction and Quantification 28

3.2.2 Second Stage: Uncertainty Characterization and Visualization 31

3.3 Experiment and Parameters Setting 33

3.4 Demonstration and Results 34

3.4.1 Results of Uncertainty Quantification Stage 35

3.4.2 Visualizations of Uncertainty Characterization Stage 36

3.5 Chapter Summary 42

Chapter 4 43

4.1 Introduction 44

4.2 Introduction on Inverse Uncertainty Quantification 45

4.3 Challenge in Inverse Uncertainty Quantification 45

4.4 Proposed Methodology for Inverse Uncertainty Quantification 47

4.4.1 Probabilistic Modelling with Deep Learning 48

4.4.2 Probabilistic Deep Dropout Generative Nets 52

4.5 Demonstration and Results 60

4.5.1 Experiment Settings 60

4.5.2 Results and Demonstration 62

4.6 Chapter Summary 68

Chapter 5 69

5.1 Introduction 70

5.2 Proposed Methodology 70

5.2.1 Multi-objective Optimization Model for Household EMS Operation 72

5.2.2 Wavelet Auto-decoupling Regularization Term 75

5.2.3 Discussion of Efficacy on WSR regularization term 77

5.3 Experiment Settings 79

5.3.1 Data Description 79

5.3.2 Benchmark Measures 81

5.4 Demonstration and Results 82

5.4.1 Validation of EMS De-coupling Effect on Household Demand Correlations 82

5.4.2 Performance Evaluation Case: 50% Penetration Rate with 100 Households 83

5.4.3 Performance Comparison under Multiple Scenarios 86

5.5 Chapter Summary 88

Chapter 6 90

6.1 Introduction 91

6.2 Point Load Forecasting and Probabilistic Load Forecasting 91

6.3 Deep Learning for Household Load Forecasting-A Novel Pooling Deep RNN 92

6.4 Chapter Summary 117

Chapter 7 118

7.1 Introduction 119

7.2 Literature of prior PLF methods 120

7.3 Modelling Load Uncertainty with Weather, Temporal and Social Features 123

7.4 Producing Probabilistic Forecasts with Inverse UQ Method 126

7.5 Implementing Probabilistic Load Forecasting with Deep Learning Model 129

7.5.1 Data preparation stage: pre-processing and manipulation 130

7.5.2 Model Training Stage 130

7.5.3 Model Testing Stage 131

7.6 Experiment Settings 132

7.6.1 Data Description 132

7.6.2 Parameter Settings 133

7.6.3 Benchmarks 134

7.7 Demonstration 135

7.7.1 Result Visualization and Benchmarking 135

7.7.2 Result Discussion 137

7.8 Chapter Summary 138

Chapter 8 140

8.1 Conclusions 141

8.1.1 Multi-time-scale Forward Uncertainty Quantification for Household Load 142

8.1.2 Inverse Uncertainty Modelling for Household Load based on Deep Learning Techniques 143

8.1.3 Decentralised Energy Management System to Achieve Grid Level goal by Decoupling Uncertainty Correlations 145

8.1.4 Point and Probabilistic Load Forecasting by Modelling Load Uncertainty with Deep Learning 146

8.2 Future Works 147

8.2.1 Extending the Fundamental Research on Uncertainty Quantification 147

8.2.2 Improving the Smart Grid Applications 150

Appendix. A 152

Appendix. B 154

Publications 156

Reference 158

Abstract

Methods for uncertainty quantification (UQ) and mitigation in the electrical power system are very basic, Monte Carlo (MC) method and its meta methods are generally deployed in most applications, due to its simplicity and easy to be generalised. They are adequate for a traditional power system when the load is predictable, and generation is controllable. However, the large penetration of low carbon technologies, such as solar panels, electric vehicles, and energy storage, has necessitated the needs for more comprehensive approaches to uncertainty as these technologies introduce new sources of uncertainties with larger volume and diverse characteristics, understanding source and consequences of uncertainty becomes highly complex issues. Traditional methods assume that for a given system it has a unique uncertainty characteristic, hence deal with the uncertainty of the system as a single component in applications. However, this view is no longer applicable in the new context as it neglects the important underlying information associated with individual uncertainty components. Therefore, this thesis aims at: i) systematically developing UQ methodologies to identify, discriminate, and quantify different uncertainty components (forward UQ), and critically to model and trace the associated sources independently (inverse UQ) to deliver new uncertainty information, such as, how uncertainty components generated from its sources, how uncertainty components correlate with each other and how uncertainty components propagate through system aggregation; ii) applying the new uncertainty information to further improve a range of fundamental power system applications from Load Forecasting (LF) to Energy Management System (EMS).

In the EMS application, the proposed forward UQ methods enable the development of a decentralised system that is able to tap into the new uncertainty information concerning the correlations between load pattern across individual households, the characteristics of uncertainty components and their propagation through aggregation. The decentralised EMS was able to achieve peak and uncertainty reduction by 18% and 45% accordingly at the grid level. In the LF application, this thesis developed inverse UQ through a deep learning model to directly build the connection between uncertainty

components and its corresponding sources. For Load Forecasting on expectation (point LF) and probability (probabilistic LF) and witnessed 20%/12% performance improvement compared to the state-of-the-art, such as Support Vector Regression (SVR), Autoregressive Integrated Moving Average (ARIMA), and Multiple Linear Quantile Regression (MLQR).

Acknowledgement

First of all, I would like to show my respectful appreciation to my supervisor, Prof. Furong Li. She has supported me with constructive suggestions, profound knowledge and helpful encouragement throughout years of my Ph.D. career. Her guidance not only helps me to improve myself in terms of knowledge, techniques, and research, but also provides me great opportunities to sharpen my presentation skills, adapt to team works, carve my intuition to the cutting-edge innovations.

I would like to take this opportunity to express my gratitude to my family and my fiancée Miss Jing Teng, who have supported me and gives me priceless consolation and encouragement when I faced with challenges and frustration.

I am grateful for my colleagues who have collaborated with me in former projects, researches, and articles. Dr. Chen Zhao, Dr. Nathan Smith, Dr. Ran Li, Mr. Minghao Xu, Dr. Zhong Zhang, Miss Qiuyang Ma, Miss Chi Zhang, Mr. Lurui Fang, Mr. Xinhe Yang, Miss Wangwei Kong, Mr. Han Wu, Ms. Lanqing Shan, Mr. Haiwen Qi. I am sincerely grateful for the efforts and times they devoted to our collaboration.

I would like to express my special thanks to Mr Lurui Fang, Dr. Kang Ma, Dr. Chenghong Gu, Mr. Han Wu, Mr. Haiwen Qi, for their assistance and constructive comments on my thesis writing and formatting.

I am also thanks my other colleagues and friends who gave me helps, inspirations, and suggestions during my Ph.D. career. For Dr. Ignacio Hernando Gil, Dr. Chenghong Gu, Dr. Kang Ma, Dr. Zhipeng Zhang, Dr. Lin Zhou, Dr. Fan Yi, Mr. Hantao Wang, Mr. Yuankai Bian, Dr. Shuangyuan Wang, Mr. Xiaohe Yan, Miss Wei Wei, Mrs. Heather Wyman-Pain, Dr. Jie Yan, Mr. Jiangtao Li. I would thank them for sharing their profound knowledge with me in their speciality.

Last but not least, I would like to appreciate for the Graduate Office, Student Services, IT support in our department and university for giving hands when I needed technical instructions and suggestions during my Ph.D. career.

List of Figures

Figure 2-1 Illustration of Skewness 15

Figure 3-1 the flowchart of the DTHUQ approach 26

Figure 3-2 Energy density spectrum of Haar expansion coefficients 30

Figure 3-3 a) residential load profiles, b) aggregated load profiles, c) quantiles of residential load profiles, d) quantiles of aggregated load profiles 34

Figure 3-4 Heatmap of the logarithm of average energy density spectrum across 900 residential households..... 35

Figure 3-5 Heatmap to present logarithm of the sum of squared coefficients in cross-time-scale channels 35

Figure 3-6 Daily uncertainty distribution of different time-scale channels from 14th July. 2009 to 31st Dec. 2010 36

Figure 3-7 Hourly uncertainty distribution of different time-scale channels over a day..... 36

Figure 3-8 Hourly uncertainty distribution of different time-scale channels over a week 37

Figure 3-9 Uncertainty heatmap of multiple time-scale channels: a) Annual distribution in day indexes; b) weekly distribution in hour indexes 38

Figure 3-10 Comparison of uncertainty distributions of two cross channels in different periods: a) annual period, b) daily period, c) weekly period. 39

Figure 3-11 Uncertainty propagation factor in differing aggregation scales 41

Figure 3-12 Normalised uncertainty propagation factor in differing aggregation scales 41

Figure 4-1 Illustration of two types of deep neural networks: a) deterministic model; b) probabilistic model. 50

Figure 4-2 Computational graph of proposed probabilistic deep dropout generative network: a) deterministic model; b) probabilistic model. 54

Figure 4-3 Visualization on Probability Distribution of Uncertainty Component Driven by Time of a Day 63

Figure 4-4 Visualization on Probability Distribution of Uncertainty Component Driven by different Months..... 63

Figure 4-5 Visualization on Probability Distribution of Uncertainty Component Driven by Temperature..... 64

Figure 4-6 Visualization on Probability Distribution of Uncertainty Component Driven by Humidity..... 64

Figure 5-1 Flowchart of Decentralised home EMS Operation Strategy 71

Figure 5-2 Demands and quantiles at aggregated/disaggregated level: a) demand of residential households, b) demand of distribution network 79

Figure 5-3 TOU tariffs employed in differing seasons 81

Figure 5-4 Household demand and battery solution with different EMS strategy 84

Figure 5-5 Quantile plot of distribution network with different EMS strategies: a) without EMS; b) benchmark P1; c) benchmark P2; d) WARP strategy 85

Figure 5-6 EMS performances on a) peak reduction & b) uncertainty reduction across different penetration rate and aggregation scales..... 87

Figure 5-7 Performance of WARP strategy across different weight factors..... 88

Figure 7-1 Demonstration of How Uncertainty Sources Generate Load Uncertainty 124

Figure 7-2 Comparison of Computational graphs of: a) Underlying Uncertainty Model, and b) Deep Learning Model, on Load Uncertainty.....127

Figure 7-3 Flowchart of Proposed Deep Learning Model for Probabilistic Load Forecasting 129

Figure 7-4 The computational graph and unfolded topological graph in two continuous epochs at the training stage.....131

Figure 7-5 The computational graph and unfolded topological graph in two continuous epochs at the testing stage 132

Figure 7-6 Weekly quantile forecasts with vanilla MLQR method.....136

Figure 7-7 Weekly quantile forecasts with proposed PDDGN method 136

List of Tables

Table 3-1 Notations of Methodology 27

Table 3-2 Average Statistics (covariance, variance and PCCs) across 900 Residential Households 40

Table 4-1 Parameters Settings in Training Stage 61

Table 4-2 Parameters Settings in visualization Stage 61

Table 5-1 NOMENCLATURE 72

Table 5-2 BATTERY PARAMETERS 80

Table 5-3 STATISTICS (EXPECTATION, 25% QUANTILE & 75% QUANTILE) OF DEMAND CORRELATIONS IN PCCS ACROSS 900 RESIDENTIAL HOUSEHOLDS 83

Table 5-4 EMS Performance Summary on Scenario with 100 Aggregation Scale and 50% EMS Penetration Rate 86

Table 7-1 Parameters Settings in Training and Testing Stage 134

Table 7-2 Comparison of Average Pinball Loss between Benchmarking method and Proposed Method Across 5000 Customers 137

List of Abbreviations

ANN	Artificial Neural Network
ANOVA	Analysis of variance
BNN	Bayesian Neural Network
CBTs	Customer Behaviour Trials
CER	Commission for Energy Regulation
CLT	Central Limit Theorems
CNN	Convolutional Neural Network
DL	Deep Learning
DRBM	Deep Restricted Boltzmann Machine
DRNN	Deep Recurrent Neural Network
DSA	Deep Sparse Autoencoder
DSM	Demand Side Management
DSR	Demand Side Response
DSRT	Dropout Stochastic Regularization Technique
DTHUQ	Data-driven Temporal-dependency Haar expansions Uncertainty Quantification
DWT	Discrete Wavelet Transform
EDS	Energy Density Spectrum
ELBO	Evidence Lower Bound
EMD	Empirical Mode Decomposition
EMS	Energy Management System
EV	Electric Vehicle

FUQ	Forward Uncertainty Quantification
GAN	Generative Adversarial Net
GLUE	Generalized Likelihood Uncertainty Estimation
gPCs	generalized Polynomial Chaos
GUM	Guide to the Expression of Uncertainty in Measurement
HWT	Haar Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
KL divergence	Kullback–Leibler divergence
LDA	Latent Dirichlet Allocation
LF	Load Forecasting
LSTM	Long Short-term Memory
LV	Low Voltage
MAE	Mean Absolute Error
MC	Monte Carlo
MILP	Mixed-Integer Linear Programming
MLP	Multilayer Perceptron
MLQR	Multiple Linear Quantile Regression
MLR	Multiple Linear Regression
MW	Mega Watt
NRMSE	Normalised Root Mean Squared Error
NUPF	Normalised Uncertainty Propagation Factor
PCC	Pearson Product-moment Correlation Coefficients
PDDGN	Probabilistic Deep Dropout Generative Network
PDDGN	Probabilistic Deep Dropout Generative Nets

PDE	Partial Differential Equations
PDF	Probability Distribution Function
PDRNN	Pooling-based Deep Recurrent Neural Nets
PLF	Probabilistic Load Forecasting
PV	Photovoltaic
QR	Quantile Regression
RMSE	Reduced Mean Squared Error
RMSE	Reduced Mean Squared Error
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SD	Standard Deviation
SG	Smart Grid
SMEs	Small/Medium Enterprises
SOC	State of Charge
SRTs	Stochastic Regularization Techniques
SSR	Stochastic Surface Response
STD	Standard Deviation
STLF	Short-term Load Forecasting
SVR	Support Vector Regression
TOU	Time-of-use
UCM	Uncertainty Components Modelling
UPF	Uncertainty Propagation Factor
UQ	Uncertainty Quantification
VA	Variational Auto-encoder

VI	Variational Inference
Wha	Weiner Haar expansions
WSR	Wavelet Auto-decoupling Regularization
WT	Wavelet Transform

Chapter 1

Introduction

THIS chapter briefly introduces the background, motivations, objectives, challenges and con tributions of this thesis. The outline of this thesis is also provided in this chapter.

1.1 Background and Motivation

1.1.1 New Context: Growing Low Carbon Technology

To move towards a green future, over 100 countries worldwide have set the environmental target to limits the global warming to 2 ℃ at most [1, 2]. Specifically, in the UK, the greenhouse gas emission target is hereafter set to be 80% reduction by 2050 with respect to the standard at 1990 level [1, 2]. As one of the largest proportion of greenhouse gas emission sources, power system, i.e., electricity and heat productions takes around a quarter of the overall emission [3], which is the largest single source of greenhouse gas emission. Therefore, the environmental targets have long become the major motivation for technical innovations in the power sector.

From a technical perspective, traditional fossil energy fuel generates over 60% of the total carbon emission and act as the major environmental threat. Therefore, many low carbon techniques have invented to resolve the adversarial impact of fossil energy. In general, two categories of low carbon techniques are deployed:

- Renewable generations: renewables generations such as wind generation, and photovoltaics are an environmentally friendly substitute for the fossil energy sources. By replacing the traditional fossil energy sources with renewables, the carbon emission will be largely reduced from its origin.

- Flexibility resources: Alongside the renewable generations that attempts to reduce carbon emission by replacing fossil energy, many other innovations are deployed to enhance the efficiency of the power system by providing flexibility. For instance, electric vehicles, battery storages, and so on. These flexible techniques provide the necessary flexibility to the distribution networks. By optimally operates the flexibility, it can create great values to the system, such as deferring the infrastructure investments, reducing network congestions, reducing system demand peak, increasing system utilization, etc. Hence, reduce the required volumes of electricity generation with fossil fuels.

Nowadays, these low carbon techniques have increasingly grown and integrated into the system as the new context of the modern and future power system.

1.1.2 New Challenge: Uncertainty in the Power System

Majority of prior research in the electrical power system, deployed Monte Carlo and it's metamethods to deal with the uncertainty due to its simplicity and easy to be implemented and generalised, when the loads are predictable, and generations are controllable. Efforts in investigating advanced uncertainty quantification and mitigation techniques are not necessarily required in the traditional power system.

However, with the new context of the modern power system, i.e., the growing low carbon technologies. New sources of uncertainties with larger volume and diverse features are introduced to the power system, in particular, to the Low Voltage (LV) networks. However, these uncertainties become one of the major challenges to the operation and planning of LV networks. For system operation, uncertainties have several adversary impacts, such as deteriorating the accuracy of prediction, causing financial risks to operational solutions, breaking the balance between generation and demand, and leading to voltage and frequency instability, etc. For system planning, it may move up network reinforcement, increasing the network loading level, and so on. Developing deep insights towards the natures and properties of uncertainties in the power system can help to improve the existing smart grid applications by: i) adapting existing techniques to the uncertain environments; ii) developing new techniques to reduce uncertainty in the power system. Therefore, understanding uncertainties in the power system is of great importance.

Due to the significance of uncertainty research, extensive efforts are paid in the energy community. In general, prior works generally consist of two types of content:

- **Uncertainty quantification (UQ) in power system**: To support the planning and operation of vast of smart grid applications under uncertain environment, it is of immense importance to understand the uncertainty of interest. This understanding is not only about the numerical quantification of uncertain elements, but also the underlying properties of this uncertainty, such as the component composition, causing sources, and the spatial or temporal characteristic of the uncertainty. In other fields, such as Fluid Dynamics, Chemistry, Applied Mathematics, Complex Environment System, and so on, two general types of problems are defined and investigated for decades, i.e., forward UQ and inverse UQ. While forward UQ

provides tools, such as MC methods, for quantifying the uncertainty of interest numerically, inverse UQ problem models the relationship between the uncertainty and its causing sources. Prior works in the power system mainly focused on forward UQ, and implemented with MC method and its meta approaches. As an example, in prior works [4, 5, 6], the uncertainty distribution of the electric power flow is investigated given assumptions on the uncertainty distribution of electric load of individual households. Some forward UQ methods, such as Surface Response (SR) method are therefore proposed to accelerate the convergence and improve the accuracy of uncertainty quantification. Alongside with forward UQ problem, inverse UQ problem can be regarded as the complementary to the forward UQ that can provide deeper insights into the underlying properties of uncertainty other than the numerical estimation/quantification of uncertainty. These properties include: i) the component composition of the observed uncertainty, ii) and the underlying relationship between observed uncertainty and its causing sources. Tackling the inverse UQ problem will provide natural tools for assessing the impact on uncertainty from its sources, which will ultimately indicate how to mitigate uncertainty to the power system. Although the significance of inverse UQ problem is widely recognised in other fields, it is still untouched in the power system.

- **Improving smart grid applications under uncertainty**: Compared to the limited literature indirectly quantifying uncertainties in the power system, more prior works focused on how to improve the planning and operation of techniques and applications in the smart grids considering the impact from uncertainty. For example, many works [7, 8, 9, 10, 11] investigate how to operate EMS economically and efficiently given uncertainties in electricity demand and distributed generations. Approaches such as stochastic programming, robust optimization, etc, are brought up to achieve the target from a probabilistic perspective. Another example is the probabilistic load forecasting problem, which attempts to generate the probability distribution of future electric load of interest. Some works [12, 13, 14, 15] assumes the strong correlation between load uncertainty and temperature recorded by neighbouring weather stations [16], and propose a scenario-based model that generate a collection of temperature realizations as the input, and predict the possible realizations of future load by input the temperature scenarios into a deterministic forecasting model.

Although the uncertainty research is contributed by worldwide researchers in the energy community, two unsolved limitations and research gaps are still existing:

- **Research gaps in the fundamental science of uncertainty quantification**: As mentioned above, most uncertainty quantification works in the power community are tackling the forward UQ problem. These works provide tools for assessing the numerical quantity of target uncertainty but can rarely provide detail information regarding the components composition of the uncertainty, and how these components are impacted by the causing sources. Whereas, inverse UQ acquires the functionality towards the desired information. However, to the status quo in the power system, none of the existing works has touched inverse UQ problems. Therefore, this thesis firstly attempts to propose practical approach and implementation to tackle the inverse UQ problem in the power system.

- **Lack of methodology diversity in dealing with uncertainty in applications:** In terms of prior works that aims to enhance the performance of the existing application under uncertainty, the family of Monte Carlo methods is the most commonly deployed methodology. More specifically, in terms of the example of EMS application with stochastic programming or robust optimization, many implementations were to translate the uncertainty in demand and prices into scenarios associated with possibilities. This scenario representation of the uncertainty is essentially equivalent to the Monte Carlo method that samples possible scenarios of uncertainty from the underlying distribution. So as the example of probabilistic load forecasting with temperature scenario methods. Including the examples, existing literature in the power system lack of diversities in methodology application other than the family of Monte Carlo methods. However, limited works deploy alternative methodologies such as spectral approaches that are widely investigated in research communities out of power system.

1.2 Research Contributions of Thesis

Therefore, there are two key contributions of this thesis to the state-of-the-art knowledge: i) this thesis firstly investigates inverse uncertainty quantification problem in the power system, to fill the research gaps in the fundamental science of uncertainty

quantification; ii) this thesis demonstrates how to develop advanced demand-side applications to achieve preferable performances with proposed uncertainty models and approaches under high uncertainty. In details, this thesis presents innovations in both problems and applications.

In terms of contributions to the fundamental science of uncertainty quantification:

- This thesis further extends the forward uncertainty quantification across multi-time-scales. A novel Temporal Dependency Haar expansions Uncertainty Quantification approach is proposed. It integrates wavelet transform decomposition into the Monte Carlo uncertainty quantification scheme. The proposed method is also demonstrated on smart meter data of domestic electricity demand, to characterize the temporal natures of demand uncertainty.

- This thesis for the first time investigates inverse uncertainty quantification problem for the power community, which will provide a natural tool to discriminate diverse uncertainty components and find the relationship between the uncertainty components and its causing sources. In the thesis, a novel Probabilistic Deep Dropout Generative Nets (PDDGN) is proposed based on the Theory of Bayesian Inference and implemented with the promising deep learning techniques. The approach is demonstrated on smart meter data of domestic electricity demand, to formulate the contributions of external sources, such as temperature, humidity, the day of the week, month, etc.

As to contributions in improving existing application in smart homes and smart grids:

- For Energy Management System application, an advanced decentralised home EMS is proposed to achieve both home and grid level goals simultaneously. Compared to prior home EMS, the proposed EMS can explicitly achieve grid-level goals by adding a Wavelet Auto-decoupling Regularization technique. On the other hand, compared to prior EMS at the grid-level, the proposed EMS: i) just needs to solve a simple MILP programming model at home level; ii) does not require investment in ancillary communication devices.

- For Load Forecasting application, two novel deep learning method, are proposed for addressing the point load forecasting and probabilistic load forecasting problems accordingly. For point load forecasting, a novel Pooling Deep Recurrent Neural Nets is proposed. Compared to prior point LF methods, the proposed methods can improve the accuracy by learning the sharing uncertainty between the neighbourhood and tackle the overfitting issue by increasing data diversity with pooling-strategy.

- For probabilistic load forecasting, this thesis deploys the proposed bespoke deep learning method for inverse uncertainty quantification, i.e., probabilistic deep dropout generative network (PDDGN) model. By modelling the contributions of uncertainty sources, this method can accurately estimate the probability distribution of combined uncertainties with given measurements on external uncertainty sources. Both proposed load forecasting approaches are witnessed superior performance for domestic load forecasting compared to prior related works.

1.3 Structure of Thesis

Chapter two introduces the preliminaries of uncertainty research, includes concepts, terminologies, and fundamental knowledge. Prior uncertainty research in the electrical power system is also reviewed and discussed.

Chapter three develops an advanced forward UQ method, to quantify and characterize individual uncertainty components of load uncertainty. The proposed method can disaggregate load uncertainty into components at multiple time scale, and models the correlation and features across different time scales. The result is demonstrated on smart metering data to reveal the temporal features of load uncertainty.

Chapter four for the first time develops a Deep Learning method to achieve inverse UQ for electric load uncertainty. The fundamental idea is to infer the underlying uncertainty model and its parameters with the power of Bayesian Inference theory. The latent model parameters that learned from smart metering data with the promising learning capability of Deep Learning model. The demonstration is implemented on Low

Carbon London smart metering database. The result indicates how uncertainty generated from external uncertainty sources through propagation.

Chapter five improves the Home Energy Management System (EMS) by applying the properties in uncertainty propagation. According to the property that uncertainty will be mitigated through propagation, and the mitigation rate is related to the diversity or more exactly the coupling degree of uncertainty components. An advanced EMS operation strategy is proposed to achieve higher uncertainty mitigation through system demand aggregation. The developed home EMS is reported to achieve peak and uncertainty reduction at grid level with decentralised control.

Chapter six proposes a deep learning technique to improve point load forecasting for domestic demand by directly modelling the temporal connectivity of uncertainty and the sharing uncertainty across neighbourhoods. The testing results prove that the proposed method can achieve 20% accuracy improvement compared to the state-of-the-art.

Chapter seven improves the probabilistic load forecasting with the Deep Learning model developed in Chapter Four. The proposed Deep Learning model can directly model the relationship between uncertainty components and its causing sources. By inputting the observations on these uncertainty sources, such as weather conditions, temporal features, the prospected uncertain load can be predicted with probability distribution associated. By assessed with Pinball loss, the proposed probabilistic load forecasting method can achieve around 12% performance improvement compared to prior state-of-the-art.

Chapter eight summarises the key findings from the research, the major contributions of the work, and discusses the potential research topics in future work.

Chapter 2

Preliminaries and Review on Uncertainty Research

THIS chapter briefly introduces the concepts and terms of uncertainty researches. The prior literature of uncertainty study in electrical power system is also reviewed and discussed.

2.1 Introduction

To clarify the preliminaries for the thesis, this Chapter will formalise the terminologies and introduce common concepts of uncertainty research. The first part will start with the definition of uncertainty, which will get rid of some confusions on the terminologies of uncertainty across different disciplines. For instance, terms such as 'intermittency', 'residual' are often used in power systems with similar meanings to 'uncertainty' but are rarely used by the general community.

In the second part, some essential concepts are introduced afterwards, which give answers to some crucial problems to the understanding of uncertainties, e.g., How to measure and represent uncertain quantity? How to categorize different uncertainty components? What are the sources of uncertainty? And how these uncertainty sources take effect and generate uncertainty?

Next part will formally introduce the problem of uncertainty quantification. In particular, this thesis will show the existing problem regarding uncertainties in the power systems can be summarised into two general types of uncertainty quantification problems, i.e., forward uncertainty quantification (FUQ) and reverse uncertainty quantification (RUQ). As two widely researched problem in the uncertainty community, FUQ and RUQ are built upon Bayesian Probabilistic modelling. Therefore, we would like to use the language of Bayesian Probability to formalise all the discussions on mathematics across the following chapters.

Last part will review and discuss the prior works done in the electrical power system, to recognise the status quo of uncertainty research in the power system. Hence to reveal the limitations of current approaches, and opportunities to improve.

2.2 What is Uncertainty?

'Uncertainty', literally means the lack of certainty, refers to the combined concept with twofold, i.e.: i) the lack of confidence to exactly describe the existing states or future outcomes to the objective due to the limited information; and ii) the nature of randomness that will result in multiple possible states or variabilities in the outcomes. As a mixture of concepts, the uncertainty of a model observation, in practice is the

propagation from two sources corresponds to the definition of uncertainty. On the one hand, the lack of knowledge and information will lead to uncertainties on the hypothesis that supporting the structure of the model. This component of uncertainty is associated with the widely known 'Epistemic Uncertainty', 'Knowledge Uncertainty' or 'Systematic Uncertainty' [17, 18, 19, 20, 21, 22]. On the other hand, with the presence of random natures in almost any physical systems, the output of the system cannot be predicted entirely accurate, even with complete relevant information available. This component of uncertainty acts as variability in observations, which has been described as 'random uncertainty' in many prior studies [17, 18, 23, 24].

Notably, the terminologies 'Uncertainty' is commonly shared by various fields of research, but there exist some similar or alternative terms in specific topics. For instance, 'intermittency', 'residual', 'variability', etc. In some situations, these terms are interchangeable, but there still exists distinctions between these terminologies. Comparing the similarities and differences would help us to understand the concept of 'Uncertainty'.

In terms of 'intermittency', in the energy sector, it is often more emphasizing the temporal discretion in value and continuity in the temporal domain. Whilst in dynamical systems, it refers to the irregular alternation or changes from a phase to another. Unlike 'Uncertainty', the terminology 'intermittency' focuses on the pattern of the signal itself, rather than from the probability or possibility point of view.

Compared to 'intermittency' in specific contexts, the term 'variability' is more similar to 'uncertainty'. Normally, it refers to the random variation in an uncertain quantity. In many prior works [25, 26, 27], authors regards 'variability' as 'aleatory uncertainty', which is a category of 'uncertainty' components that are caused by natural randomness. The categories of uncertainty components will be introduced in latter sections.

The term 'residual', or more generally the term 'error', are also interchangeable with 'uncertainty' in many cases. However, it has more specific meaning that referring to the distance between samples and the true value. Even in some situation the true value cannot be measured directly, and will be alternatively estimated by the arithmetic sample mean. Specifically, in the case of 'uncertainty', it always exists whether the true value (estimates) exists or not.

In general, these alternative terms are widely recognized in many research fields, and discriminated with 'uncertainty' under some specific criteria. In order to formalize the terminologies and discussion in this thesis, this thesis will only use 'uncertainty', rather than the alternatives, as the terminology for quantifying the lack of certainty in both value and confidence. Notably, for some text in the latter chapters, the alternative terms may also be mentioned, but will be discussed with a specific definition. For instance, 'residual' will be introduced in discussing load forecasting problems to refer to the difference between samples and arithmetic sample mean.

To the status quo, there exist two ways of interpretation of probability in the academia, i.e., Frequentist Interpretation of Probability, and Bayesian Interpretation of Probability. While Frequentist Interpretation of Probability defines probability as the limit of the frequency of trials, when the trial times is sufficient. This view connects the probability with experiments, and form the basis of many methods, such as Monte Carlo simulation. Bayesian Interpretation of Probability distinguishes the actual probability with the limit of the frequency of experiments, where the actual probability of interested is named as a posterior distribution on condition of given frequency of experiments, while the frequency of experiments is named as evidence probability that only represents the data samples rather than the actual probability. In this view, the actual probability of interest is considered as a conditional distribution that the sampling from it will form the evidence distribution, i.e., to generate the observations in the trials.

2.3 Uncertainty Categories and Sources

Based on the definition of 'Uncertainty', this part will further discuss the components and categories of uncertainty, to better recognize the meaning of uncertainty. The categories of uncertainty components are firstly reported by the Guide to the expression of uncertainty in measurement in 1995 [17]. In general, it distinguishes uncertainties into systematic (epistemic) and random (aleatory) components [17, 18, 21, 24] according to the uncertainty source types and characteristics. This categorization is commonly recognized by the academia and industries.

As its name, 'epistemic/systematic uncertainty' are caused by systematic effect from the underlying physical system. To trigger the corresponding systematic effect to

generate the epistemic uncertainty, a latent or observable factor often exists. With prior knowledge on the underlying mechanism of this triggering factor and the systematic effect, epistemic uncertainties can be mitigated. For instance, the temperature may result in load uncertainty for households, if the future temperature is available via prediction or simulation, the associated load uncertainty can be predicted more accurately.

'Aleatory uncertainty' refers to the natural variations of the uncertain entity. However, it also indicates none of triggering factor are associated with this uncertainty component. In other words, this part of uncertainty components naturally exists, and irreducible even with the complete systematic knowledge of the underlying system. In most of existing physical systems, the probability of aleatory uncertainties is, or can be approximated with Gaussian Distribution [23, 24].

2.4 How to measure uncertainty

In most scenarios in the engineering, knowing the quantity of uncertainty is essential to get rid of the risk caused by uncertainties. Therefore, it calls for benchmarks and methods to measure uncertainty. Formally, this process is named Uncertainty Quantification (UQ) in the researching fields. Currently, there are many different measures are widely recognized, such as Variance, Quantiles, Intervals, etc. These measures provided tools for quantifying and even characterizing uncertainties in the engineering problems. However, these numerical measures are not sufficient to capture all the features of uncertainty, especially on its structure and distribution. Therefore, it calls for more intuitive and informative methods to represent uncertainties without losing information. In this section, the widely used measures and representation methods for uncertainty are introduced. In details, threefold of measures are presented, i.e., the moment-based measures, probability measures and possibility measures.

2.4.1 Moment-based Uncertainty Measures

The Moment is a specifically defined quantitative measure to describe the shape of a probability distribution with a set of numerical measures, without information loss on the probability distribution. In general, the moments of a probability distribution consist of a set of values with a different order of moments. For instance, the 0^{th} order moment

refers to the total probability of the event for occurring, the first moment then refers to the arithmetic mean of the probability distribution, the second moment is hence represents the variance.

The probability distribution defined in the bounded interval can be uniquely determined once given the collection of all orders of moments. Notably, this proposition will not exist if the probability distribution is unbounded, which is discussed by the Hamburger Moment Problem [28]. In other words, the moment is a tool to capture all the features of a probability distribution with numerical measures, which simplify the difficulty in representing uncertainty. The n^{th} order moment for a continuous probability distribution $p(x)$ is formulated as [29, 30]:

$$\mu^n = E[X^n] = \int_{-\infty}^{\infty} x^n p(x) dx \qquad (2\text{-}1)$$

In details, for a set of sample collections of random quantity, the different order of moments extracts a different feature of the uncertainty, and the commonly deployed moment-based uncertainty measures are the first four order of moments:

Total probability (0^{th} order moment):

The 0^{th} order moment equals to $\int dF(x)$ which refers to the total probability.

Mean (1^{st} order moment):

It is defined as $E[X]$, and known as arthe ithmetic mean or random quantity. It describes the estimate of the true value by data samples.

For second and higher order moments, the central moments are normally investigated rather than the moment with respect to constant zero.

Variance (2^{nd} order moment):

The second order moment is defined as $E[(x-\mu)^2]$, which is also known as Variance of a probability distribution. It describes the degree of variability of the uncertain data around of the true value.

Skewness (3^{rd} order moment):

The third order moment is named as skewness [31, 32, 161], which is formulated as the normalised third order moment as:

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E[(X-\mu)^3]}{(E[(X-\mu)^2])^{3/2}} \quad (2\text{-}2)$$

where it gives the asymmetry of the probability distribution. If the skewness is positive, the probability distribution has a longer right tail and the major proportions of its mass are allocated to the left of its mean, and vice versa. Given in **Figure 2-1** to illustrate the differences between positive and negative skewness.

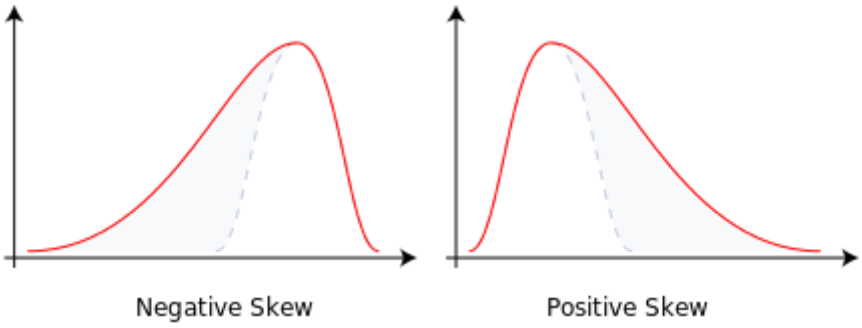


Figure 2-1 Illustration of Skewness [161]

Kurtosis (3^{rd} order moment):

The fourth central moment is named kurtosis [32, 33], which is the measure to present the heaviness/proportions of the distribution tails. In details, a higher kurtosis will refer to lower concentration to the mean of the distribution (heavier tails of the distribution), and oppositely, the light tail distribution will have a small kurtosis measure.

Mathematically, the kurtosis is formulated with $E[(T^2 - aT - 1)^2]$ [33], where $T = (X-\mu)/\sigma$.

Above are the most widely deployed features of moment measure to describe the shape and structure of the probability distribution of uncertainty. Theoretically, all the features of an uncertain quantity can be shaped by the collection of all orders of moments. And for each order of moment, it can be estimated unbiasedly with Monte Carlo methods or its metal methods. However, it is computationally expensive, unrealistic or unnecessary in the practice. Therefore, the first four orders of moments are considered in most cases.

In terms of uncertainty with higher complexity, the first four orders of moment measures are a simplified method but not providing sufficient information. Therefore, other measures are also required.

2.4.2 Probability-based Uncertainty Measures

According to the prior studies, probability distribution function (PDF) is the most common way to describe features of uncertainty, which associates a probability measure to any possible state, values or scenarios of the uncertainty. As a fact, given an uncertain quantity, there will only exist a unique probability distribution that matches the uncertain quantity. In other words, the probability distribution is the way that captures all the details of an uncertainty within its possible range.

Alongside moment-based measures of uncertainty, which is regarded as the numerical abstraction of a probability distribution, more works directly focus on the probability distribution itself. In mathematics, many typical closed form probability distribution functions have been proposed and deeply investigated to understand its probabilistic features and physical models. For instance, Gaussian distribution is introduced by the physical model of the Gaussian process. However, in practice, the closed form function to represents the probability distribution is rarely available in most cases. Therefore, given a set of data collection, we have to estimate the latent distribution type that lies behind the data collection.

Compared to Moment-based measures, the advantage of probability measure is the competence of uncertainty information remains in the measure. While moments at certain order will only hold the relevant information, such as kurtosis only indicate how heavy is the tails of a probability distribution, probability distribution itself contains as much information as possible.

2.4.3 Other Uncertainty Measures

Other widely recognized uncertainty measures include Uncertainty Interval [34, 35], Quantiles [36, 37], Uncertainty Set [38], Fuzzy Set [39, 40], Possibility measures [41, 42], and so on. Compared to a probability distribution that is easy to be generalized to arbitrary problems, these measures are normally ad-hoc measures designed to meet the

industrial needs. For instance, Uncertainty Set is widely used for robust optimization to achieve optimal unit commitment, demand response, etc., under high uncertain and risky environment. The benefits of these measures are witnessed under certain industrial applications, which lies in the simplicity of implementation and computational resources saving. However, the major limitation of these ad-hoc measures is the lack of generality to different scenarios.

2.5 Uncertainty Quantification Problems

Uncertainty Quantification [43, 44, 45, 46] is a research area that aims to formalize the quantitative characterization and mitigation of uncertainty. It has been theoretically researched under computational scenarios with formal mathematical languages, as well as investigated in realistic engineering applications. The importance of UQ problem has been recognized by many fields, such as Fluid Dynamics, Complex Environment System, Applied Mathematics, Chemistry, and so on.

Two general types of UQ problems are defined and investigated for decades, i.e., forward UQ and inverse UQ. While forward UQ provides tools, such as MC methods, for quantifying the uncertainty of interest numerically, inverse UQ problem models the relationship between the uncertainty and its causing sources.

Forward UQ, literally refers to the uncertainty quantification and characterization in the forward uncertainty propagation process. Mathematically, it considers a generalized uncertainty model within multiple input uncertainties, and these uncertainties will propagate and accumulate to the output. Given prior knowledge on the distribution of input uncertainty and the uncertainty model function, forward UQ quantifies and characterize the probability distribution or other measures of the model output.

In the state-of-the-art, three types of methods are deployed:

Systematic Approach: If the uncertainty structure can be featured with a physical model. Or in other words, the distribution types can be modelled theoretically, i.e., Gaussian distribution, Weibull distribution, etc. It can be directly described with the parameters and hyperparameters of this type of uncertainty distribution. For instance, the estimate of mean μ and variance σ^2 once given normal distribution as its type.

Sampling-based methods: For an arbitrary probability distribution, the model output uncertainty can be quantified with Monte Carlo method and its meta sampling-based methods. Even if the probability distribution of interest cannot be described with closed-form functions or the system is complete a black box, MC and other sampling-based methods can be easy generalised to these scenarios since it only requires the model can generate output realizations once feed-in realizations of inputs. By sampling under the known distribution of input uncertainty, these realizations of model inputs are easy to be obtained. Despite the simplicity and easy to be generalized, MC methods often suffer from efficiency and accuracy, with a comparable volume of realizations sampled from underlying distribution, MC methods converges much slower than another family of approaches, such as polynomial chaos expansion.

Spectral Approach: If the probability distribution types of the uncertainty are unknown, one alternative is to approach the probability distribution with a set of simpler distribution functions. These approaches basically estimate the complex form of target probability distribution with the combination of a set of functional, which holds elegant mathematic attributes to simplify the quantification costs. For example, generalized Polynomial Chaos (gPCs), Weiner Haar expansions (WHa), etc. Given sufficient volume of data samples, the latent probability distribution can be then reconstructed.

Inverse UQ is also known as Inverse Problem, is regarded as a much more challenging problem compared to forwarding uncertainty propagation. Since the underlying function of the uncertainty model is entirely unknown for the inverse problem. The model structure, distribution type, and model parameters are all needed to be estimated. To the status quo, the majority of research out of power system is to solve forward uncertainty propagation, very limited works have investigated the inverse problem. While in the electrical power system, none of the works has touched inverse problem. Even in Applied Mathematics, the inverse problem still suffers from some issues that remain unsolved.

Currently, two families of approaches are proposed and deployed for inverse UQ problem in academia:

Statistical Inference: Statistical Inference, or more specifically refers to Frequentist Inference, is a method to infer the proposition from sampled data. The framework is

built upon several well-established methodologies, such as statistical hypothesis testing, confidence, and probability intervals, and so on. It can be used to infer a proposition with the most possibility.

Bayesian Inference: Other than Statistical Inference, the majority research focus Bayesian Inference, which aims to infer the underlying uncertainty model by regarding the underlying probability distribution of model parameters as a posterior distribution on condition of model input and output variables. By invoking Bayes' theorems, the formulation of the aimed posterior distribution will become an optimization problem of minimization on Evidence Lower Bound (ELBO) [46, 47, 48].

Compared to Frequentist Inference that infers the probability as the limit of its relevant frequency with a vast number of trials, or data samples, Bayesian Inference considers probability as a reasonable expectation in the form of posterior probability distributions, and can be tested or estimated indirectly by assigning prior distributions to the hypotheses that connects the prior probability distribution with the posterior distribution of interest. This attribute enables Bayesian Inference to infer the underlying uncertainty model from input and output observations without prior knowledge to its inherent mechanism and model functions.

The major unsolved difficulties of inverse UQ problem lies in twofold:

Curse of Dimensionality: The existing of frameworks proposed to solve inverse UQ problem will need to infer the model structure, model function types and related parameters, which is an extremely complex problem with large amount variables to be calculated. Therefore, with the increase of problem dimensionality, the computational costs will grow dramatically. Therefore, most research discusses simplified scenarios and models for inverse UQ.

The concern of Identifiability: Although the underlying uncertainty model can be inferred by existing approaches of inverse UQ, the solution may not be unitary. Instead, a range of solutions could lead to the same observation result. In this scenario, the methodology or data will cause an unprecedented error in inferring the underlying uncertainty model.

2.6 Reviews on Prior Power System Uncertainty Research

2.6.1 Literature of Uncertainty Quantification

In terms of the fundamental research of uncertainty quantification methods, the mainstream in power system is the family of MC methods. Therefore, this review will categorize existing researches into MC methods and other methods.

For the family of MC methods, many meta methodologies of MC method have been produced and investigated in power system. In work [49], the author deploys the vanilla MC method to emulate the probability in the energy trading market. Work [50] assesses the impacts of wind generation in the distribution network, by invoking MC sampling from wind historical data. Work [51] employs a meta MC method named Latin Hypercube sampling for reliability evaluation in the transmission system.

The rest of UQ methods emerged in power system are mainly analytical methods established on certain mathematical theorems. In work [52, 53] investigate the uncertainty of a simple linear model, and the model output uncertainty is obviously the convolution of the input uncertainties. Work [54] further implement this method in the framework of Discrete Fourier Transform (DFT). However, this model is heavily constrained by its strong linear assumption and expensive computational cost, hence, this method is seldom applied in latter research. Work [55, 56, 57, 58] further simplifies the uncertainty quantification of the linear model, with cumulant method to replace the convolution operation.

In latter research, to generalise the UQ problem into non-linear models, other analytical methods are explored. The major efforts are made on methods based on expansion approximation. In details, works [56, 59] approximate uncertainty distribution with Edgeworth expansion [60, 61], and applied to the quantification of probabilistic power flow. In works [62, 63], Taylor Series are deployed to compress and simplify the numerical quantification of uncertainty by translating the quantification problem into the solution of polynomial functions. Works [68, 69, 70] deploys point estimation method by estimating the first order and second order moment.

In conclusion, the existing works for the fundamental methodology of uncertainty quantification can be classified into twofold: numerical methods, and analytical methods. For numerical methods, the existing research lies in MC method and its meta methods. For analytical methods, the major idea based on the expansion and approximation of the underlying probability with simple, closed-form probability functions, to further simplify and accelerate the quantification. Notably, these methods are forward UQ methods, inverse UQ methods are not developed and investigated in any literature in the power system.

2.6.2 The literature on Power System Applications considering Uncertainty

As discussed in the research motivation, the necessity of considering uncertainty in the power system applications are widely recognized by researchers. Therefore, more and more power system applications have developed new implementations considering the impact of uncertainty. Since the uncertain elements in the new context of the power system are mainly introduced by low carbon technologies integrated to distribution networks. The major efforts lie in the applications in distribution networks. In details, these applications include:

Robust Optimization (RO): As renewable generations and flexible load emerges in the distribution networks, many operational applications are heavily affected by uncertainty, such as Energy Management System, Demand Response, unit commitment and Active Network Management. To respect the uncertain scenario, update the previous version of implementation on these applications are of great necessity. Robust Optimization (RO) is a widely recognized alternative to obtain an optimal solution that is robust and stable enough to immune against risks from uncertainty. Well-developed implementations of RO include Minimax Regret Optimization [67, 68, 70], Scenario-based RO [69, 70], Uncertainty Set based RO [71, 72, 73, 74], and so on. This categorization lies in how the RO method expresses and deals with uncertainty.

Stochastic Programming (SP): This concept and technique is a direct interpretation of maximizing the expectation, which has the same range of applications. Compared to deterministic programming used in traditional power system applications, SP simply

replaces the deterministic objective with the expectation, and attempts to find the solution that can achieve maximum/minimum objective expectation.

Probabilistic Load Forecasting: Due to the integration of uncertain elements to the households, the inherent uncertainty lies in electric load has been unprecedented significant compared to the past. Therefore, the prediction of load profiles considering uncertainty has become a heatedly discussed research topics and are formally named as Probabilistic Load Forecasting. According to the needs of industry, the forms of forecasts can be intervals, quantiles, number of variances, uncertainty scenarios, or even the entire distribution.

Probabilistic Power Flow: Similar to probabilistic load forecasting, the quantification of power flow is also deteriorated by the increasing uncertainty in the distribution system. Therefore, to characterize the uncertainty associated with power flow is crucial to the system operation. Most of the existing works are developed with MC methods, to sampling the uncertainty in each network nodes, and simulate the realization of power flows by MC framework. Recently, novel UQ method, such as Stochastic Surface Response, Polynomial Chaos method [75, 76], has been proposed and considered to accelerate the convergence rate and relieve the computational burden of traditional MC methods. The fundamental idea is to approximate the uncertainty of power flow by a simple and closed form of expansions.

Alongside with the above applications and techniques that encountered uncertainty, there still exist vast of applications needed to be transformed into probabilistic version, and immense techniques can be developed to improve these applications. However, in the power system, the uncertainty research is driven by the industrial needs and the urgency to address risks from uncertainty. Therefore, the current status quo of uncertainty research in the electrical power system mainly lies in the applications in the distribution system.

2.7 Chapter Summary

The chapter introduces the preliminaries of uncertainty in terms of concepts and terminologies. And the relevant research in the electrical power system is also reviewed.

The concept of ‘uncertainty’ is firstly introduced with the interpretations from Frequentist and Bayesian point of view. In the context of the electrical power system, some confusions in terminologies are discussed to clarify the differences between ‘uncertainty’ and its similar terminologies used in power system, such as ‘residual’, ‘intermittency’. Afterwards, the measures of uncertainty are briefly introduced, included moment-based measures, probability measures, and other ad-hoc measures, such as Intervals, Quantiles, Fuzzy Set, Uncertainty Set, and so on.

The Concept of Uncertainty Quantification is firstly formalised in this chapter. According to the research in other fields, forward UQ problem and inverse UQ problems are defined accordingly. The research status quo, the main streams of methodologies, and the existing challenges are all discussed help readers to understand the motivations and ideas discussed in latter chapters about forwarding UQ and inverse UQ. Related research in the electrical power system is discussed to show the status quo of uncertainty study, to present evidence of support on the innovativeness of the works done in this thesis.

Chapter 3

Forward Uncertainty Quantification on Domestic Load Uncertainty

THIS chapter proposes a novel forward uncertainty quantification method on decomposed load components, to capture useful information of uncertainty components across multiple time scales. This method is demonstrated on the dataset of domestic energy demand.

3.1 Introduction

In previous chapters, we discussed the necessity of investigating advanced uncertainty quantification approaches, introduced concepts/terminologies of uncertainty research, and reviewed the related works in the electrical power system. In the traditional power system where the load is predictable, and generation is controllable, uncertainties in most of the real-world problems can be quantified with intuitive and simple forward uncertainty quantification methods. Therefore, the majority of prior research deploys MC method and its metamethods, and limited works use analytical methods for uncertainty quantification, to achieve higher convergence rate and accuracy.

However, above forward UQ methods are built upon an idea that regarding uncertainty as a unitary entity, which will limit the useful information we can obtain from forward uncertainty quantification. In this chapter, we will develop a novel method for forward UQ that regards load uncertainty as comprising uncertainty components across different time scales, and quantify/characterize the uncertainty at each time scales. Alongside with the numerical quantification results that also can be obtained from traditional forward UQ methods, the expected outcomes of proposed method will also include the temporal dependency and property comparisons between uncertainty components at different time scales from half-hourly, hourly, two-hourly, to daily, two-daily, and four-daily.

The chapter is organized as follows: Section 3.2 introduces the proposed methodology for multi-time-scale forward UQ, Section 3.3 discusses the details of experiment setting, and introduces the smart metering dataset that records household load data from Irish residents, Section 3.4 demonstrates the results on domestic load uncertainty, and discusses the findings on temporal dependencies and features of uncertainty in half-hourly, hourly, two-hourly, daily, two-daily and four-daily time scales. According to the comparison of uncertainty components properties. An extended test is conducted in the process of load aggregation from domestic level to the system, which indicates uncertainty components at different time scales will be mitigated in distinguished ratio.

The content of this chapter is cited from a published article [77] of the thesis author in IEEE Transactions of Smart Grid, titled as ‘Data-driven Uncertainty Quantification and Characterization for Household Energy Demand Across Multiple Time-scales’.

3.2 Methodology

The fundamental idea of the proposed forward UQ method is to integrate the spectral decomposition technique into the framework of uncertainty quantification. In this part, a novel data-driven UQ method based on Haar expansions are proposed. The flowchart that indicates the procedure of this method is presented in **Figure 3-1**:

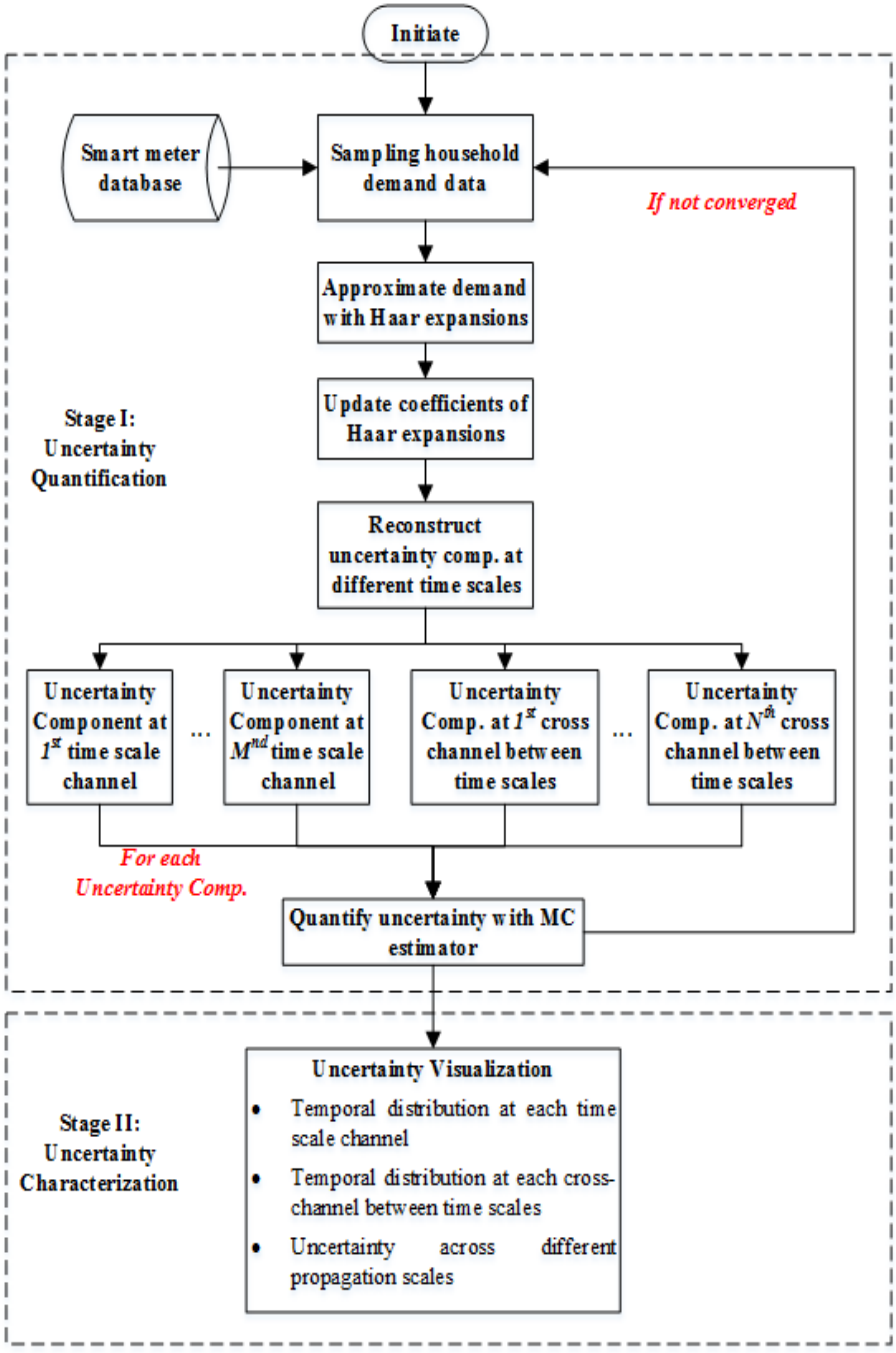


Figure 3-1 the flowchart of the DTHUQ approach

In general, the proposed method is a two-stage UQ approach: i) Firstly, a Haar expansion will be invoked to approximate the load uncertainty along the temporal domain. This process will decompose the original uncertain load into multiple components at different time scales. Afterwards, the decomposed uncertainty load approximation will be reconstructed into uncertainty components that represent the corresponding time scale channel. Quantification with Monte-Carlo UQ framework will be performed on these reconstructed uncertainty components. ii) In the second stage, with the quantification results from the first stage, detail characterization and visualization will be performed on each uncertainty components to extract useful uncertainty information and findings. In details, three types of visualizations are presented: i) uncertainty temporal distribution of each time scale; ii) uncertainty temporal distribution of each cross-channels between two-time scales; iii) and how uncertainty propagates under different aggregation scales from households to the system.

Table 3-1 Notations of Methodology

Notations	***Description***
$\mathbb{R}$	Set of real numbers
$\mathbb{Z}$	Set of integers
$L^2(\mathbb{R}^2)$	Hilbert space, square integrable functions with respect to L^2-norm
ξ	Uncertain input of uncertainty model
Y	Output of uncertainty model
$\tilde{Y}$	Actual uncertain output of uncertainty model
$\tilde{Y}'$	Estimated uncertain output of uncertainty model
τ	Discretized time variable
T	Interval or set of discretized time variable τ
Θ_T	Set of collections of sampled households
θ	Index of a certain realization in Θ_T
$c_k^{(d)}$	Coefficient of Haar expansions at decompose level d and shifting index k
$\phi_k^{(d)}$	Basis of extended Haar mother wavelet
$\Psi_k^{(d)}$	Basis of standard Haar mother wavelet
d	Decompose level of Haar expansions
k	Shifting index of Haar mother wavelet
E_θ	Expectation of given random uncertainty over the marginal integration over θ
μ_i	Expectation of random variable ξ_i

σ_i	Standard deviation of random variable ξ_i
v_i	Variance of random variable ξ_i
U	Reconstructed uncertainty components
N	Sample size
ρ	Pearson product-moment correlation coefficients (PCCs)

Notations used in this chapters are demonstrated in Table 3-1. Details of the two stages are discussed in following parts.

3.2.1 First Stage: Uncertainty Temporal Components Decomposition, Reconstruction and Quantification

Considering an uncertainty model that represents the uncertain domestic electric load with formulation $Y = f(\xi)$, in the practise of powthe er system, smart meters are the only devices that can enable the visibility of domestic electric load. Therefore, the observations of uncertain load Y is the only data available. The underlying physical system featured with synthesized function f and most of uncertainty sources ξ are unknown.

In this section, we attempt to quantify and characterize the temporal features of load uncertainty. Therefore, the time indexes needed to be considered. Denoting the time index as $\tau \in T$, where the time index could be continuous or discretized. However, the demonstrated data in this section is collected every half hour with 48 indexes per day. For mathematic convenience, we only consider the time index as a discretized index, and T is a set. For example, $\tau \in T = [0,24) \cap \mathbb{Z}$ refers to hourly indexes in a day.

Alongside with the time indexes defined for smart metering data, another crucial index used in this chapter is the index for realization collections, which is determined and corresponded by time index. Given a time index τ and its set T, the realization collection index is denoted as $\theta \in \Theta_T$. It refers to all instances that are indexed by T. For instance, if the time index are hours of a day, then the collection index are different customers in different dates. If the time index ranges from the starting hour of the first date to the last hour in the last date without daily, weekly or monthly periodicity, the collection index should be set as customer index. In other words, by iterating both time index and collection index, all data records should be dealt for exactly once.

Decomposition: According to the Haar expansion introduced in Appendix A. The uncertain demand can be expanded by Haar expansion with decompose level D:

$$Y(\tau;\theta) = Y_0(\theta) + \sum_{d=0}^{D}\sum_{k=0}^{2^d-1} c_k^{(d)}(\theta)\phi_k^{(d)}(\tau) \quad (3\text{-}1)$$

where coefficients can be obtained by:

$$c_k^{(d)}(\theta) = \frac{1}{card(\Theta_T)}\sum_{\tau\in T} Y(\tau;\theta)\phi_k^{(d)}(\tau) \quad (3\text{-}2)$$

In formulation 3-2, $\phi_k^{(d)}$ refers to the scaled Haar wavelet from standard Haar mother wavelet $\Psi_k^{(d)}$. The original unit interval [0,1] will be expanded into a closed interval that corresponds to time index set T. Further extends to two-dimensional version, the load uncertainty should be denoted as $Y(\tau_1,\tau_2)$, the corresponding expansions can be also rewritten into:

$$Y(\tau_1,\tau_2) = Y_0(\tau_1,\tau_2) + \sum_{d_2=0}^{D_1}\sum_{d_2=0}^{D_2}\sum_{k_1=0}^{2^{d_1}-1}\sum_{k_2=0}^{2^{d_2}-1} c_{k_1,k_2}^{(d_1,d_2)}\psi_{k_1,k_2}^{(d_1,d_2)}(\tau_1,\tau_2) \quad (3\text{-}3)$$

with coefficients:

$$c_{k_1,k_2}^{(d_1,d_2)} = \sum_{\tau_1,\tau_2\in T_1,T_2} Y(\tau_1,\tau_2)\psi_{k_1,k_2}^{(d_1,d_2)}(\tau_1,\tau_2) \quad (3\text{-}4)$$

The topology of decomposed uncertainty components at different time scales is illustrated in **Figure 3-2**, in the form of a matrix measured as energy density. As shown in the coefficients matrix, coefficients of any time scales or any cross channels between two-time scales can be easily indexed. Considering two independent time index τ_1,τ_2, uncertainty at daily (τ_2) time scale, can be reconstructed by coefficient sets $\{c^{(k,3)}\}_{k\in[0,3]\cap\mathbb{Z}}$.

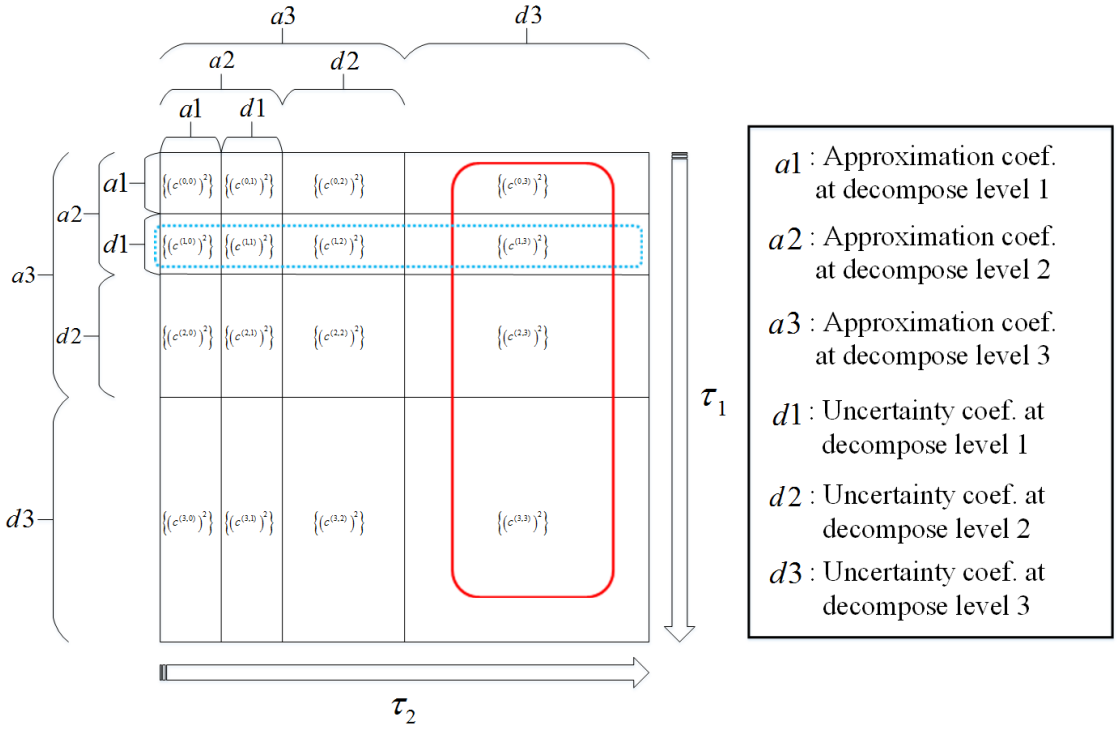


Figure 3-2 Energy density spectrum of Haar expansion coefficients

Reconstruction: Afterwards, we will deploy these coefficients matrices for uncertainty components reconstruction. Given timescale level D_0, and reconstructed uncertainty component will be:

$$Y(\theta,\tau;D_0)=\sum_{k=0}^{2^{D_0}-1} c_k^{(D_0)}(\theta)\phi_k^{(D_0)}(\tau) \qquad (3\text{-}5)$$

Extending this formulation to cross channels between any two-time scales τ_1,τ_2. Given time-scale level D_1, D_2. The reconstruction of uncertainty components can be obtained as:

$$Y(\theta,\tau;D_1,D_2)=\sum_{k_1=0}^{2^{D_1}-1}\sum_{k_2=0}^{2^{D_2}-1} c_{k_1,k_2}^{(D_1,D_2)}\psi_{k_1,k_2}^{(D_1,D_2)}(\tau_1,\tau_2) \qquad (3\text{-}6)$$

Quantification: After reconstruction of uncertainty components, the quantification of uncertainty will be performed with Monte-Carlo method. By collecting all realizations of load uncertainty indexed by time index τ, the uncertainty in standard deviation

(kWh) at time scale D_0 can be formulated as:

$$U(\tau; D_0) = E_\theta\left[Y(\theta, \tau; D_0) - E_\theta\left(Y(\theta, \tau; D_0)\right)\right]^2 = E_\theta[Y(\theta, \tau; D_0)^2] - (E_\theta[Y(\theta, \tau; D_0)]^2) \quad (3\text{-}7)$$

For cross-channel uncertainty components $Y(\theta, \tau; D_1, D_2)$, the standard deviation can be quantified as:

$$U(\tau; D_1, D_2) = E_\theta\left[Y(\theta, \tau; D_1, D_2) - E_\theta\left(Y(\theta, \tau; D_1, D_2)\right)\right]^2 = E_\theta[Y(\theta, \boldsymbol{\tau}; D_1, D_2)^2] - (E_\theta[Y(\theta, \boldsymbol{\tau}; D_1, D_2)]^2) \quad (3\text{-}8)$$

3.2.2 Second Stage: Uncertainty Characterization and Visualization

In this stage, three aspects of content will be presented, i.e.: i) uncertainty temporal distribution of each time scale; ii) uncertainty temporal distribution of each cross-channels between two-time scales; iii) and how uncertainty propagates under different aggregation scales from households to the system.

3.2.2.1 Temporal Distribution of Each Time Scale and Cross Channels between Two Time Scales

To understand how uncertainty varies along the time domain, the temporal distribution of each uncertainty components is presented first in the curve graphs and heat maps. For an uncertainty component $U(\tau; D_0)$ at D_0 time scale and indexed with time index τ. By changing the time index into two ancillary time indexes, the curve graph of uncertainty distribution along with index τ can be translated into heatmap indexed by $\tau_1 = g_1(\tau)$ and $\tau_2 = g_2(\tau)$. The uncertainty distribution in the heatmap will be reformulated into $U(\tau_1, \tau_2; D_0) = U(g_1(\tau), g_2(\tau); D_0)$.

For cross-channel between any two-time scales, the similar steps can be done by considering the time scale indexes are D_1, D_2, instead of D_0. The curve graph to visualize these components are: $U(\tau; D_1, D_2)$.

3.2.2.2 Uncertainty Propagation Analysis

Considering a scenario with N random variables $\{\xi_1, \xi_2, \dots, \xi_N\}$ as uncertain household loads of N households. The expectation μ_i, standard deviation σ_i and variance v_i of any ξ_i can be formulated statistically:

$$v_i = E[(\xi_i - \mu_i)^2], \quad i \in [1, N] \cap \mathbb{Z} \tag{3-9}$$

$$\sigma_i = \sqrt{v_i}, \quad i \in [1, N] \cap \mathbb{Z} \tag{3-10}$$

Similarly, for any two household loads ξ_i and ξ_j, the covariance between them can be quantified as:

$$v_{ij} = E\big[(\xi_i - \mu_i)(\xi_j - \mu_j)\big] = \rho_{ij}\sigma_i\sigma_j, \quad i, j \in [1, N] \cap \mathbb{Z} \tag{3-11}$$

where ρ_{ij} is the Pearson product-moment correlation coefficients (PCCs) [78]. To the system level, these household load can propagated through system aggregation: $\xi_p = \sum_{i=1}^{N} \xi_i$, since the household numbers in a substation are normally large enough, this aggregated load can be approximated by Normal distributions based on Central Limit Theorem [79]:

$$\xi_p = \sum_{i=1}^{N} \xi_i = N\left(\sum_{i=1}^{N} \mu_i\,, \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{ij}\sigma_i\sigma_j}\right) \tag{3-12}$$

Two measures are proposed in this thesis to feature the uncertainty propagation process. The first one is named as Uncertainty Propagation Factor (UPF), which refers to the ratio of propagated uncertainty with respect to the sum of uncertainty components of each household, in form of standard deviation, which is calculated by:

$$UPF = \frac{\sigma_p}{\sum_{i=1}^{N} \sigma_i} \tag{3-13}$$

Notably, this measure is complementary to the ratio that uncertainty mitigated through aggregation process. In details, UPF equals to 0.1, means 90% of uncertainty mitigated through system aggregation.

Another measure is Normalised Uncertainty Propagation Factor (NUPF), with the formulation as:

$$NUPF = \sqrt{N} \cdot \frac{\sigma_p}{\sum_{i=1}^{N} \sigma_i} \quad (3\text{-}14)$$

This measure represents independency degree between households' load. In particular, NUPF benchmark is equivalent to constant 1 in the scenario of N independent, identically distributed random variables (i.i.d. R.V.), since:

$${\sigma_p}^2 = \sum_{i=1}^{N} {\sigma_i}^2 = N^2 \sigma_i \quad (3\text{-}15)$$

3.3 Experiment and Parameters Setting

The proposed method is implemented in a publicly available smart metering database, which records half-hourly readings from 900 Irish domestic households. This database is authored by Commission for Energy Regulation. The commission period of these smart meters is from 14th July. 2009 to 31st Dec. 2010 [37]. Notably, the involved households are not incentivised by any technical or commercial incentives, hence maintains the nature of the household load. **Figure 3-3** shows the load profiles at household and aggregated level. The corresponding quantiles plots are presented in subplots c) and d).

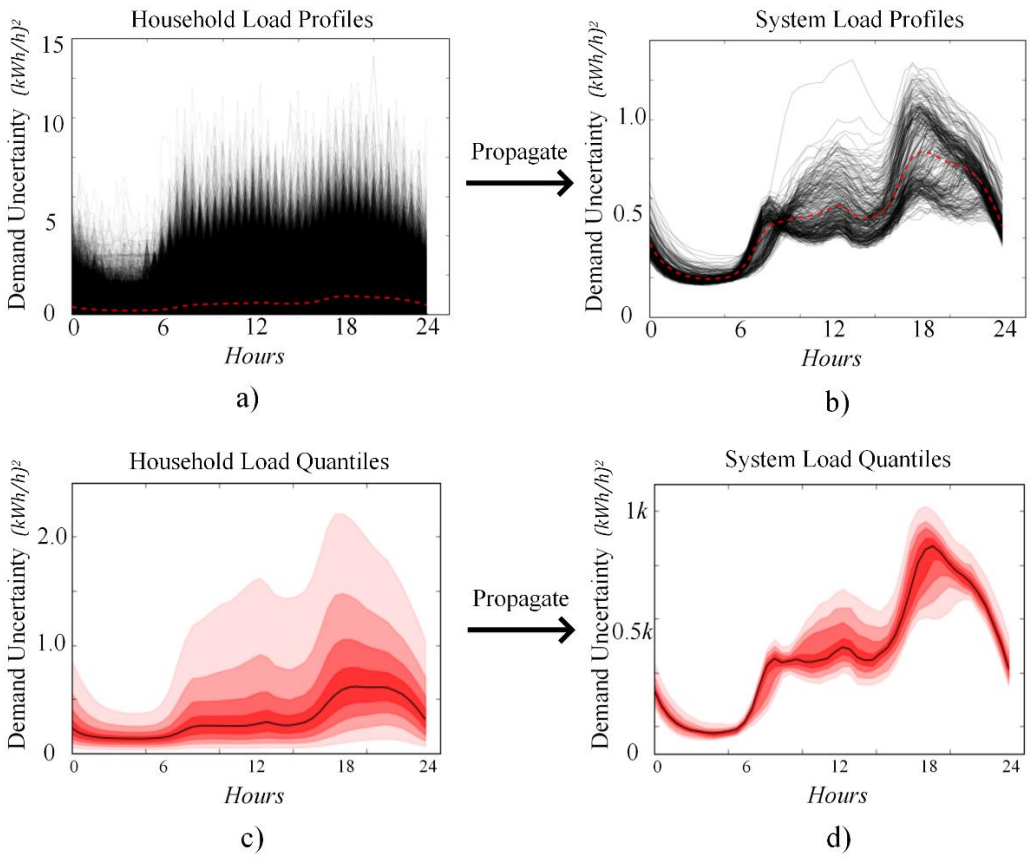


Figure 3-3 a) residential load profiles, b) aggregated load profiles, c) quantiles of residential load profiles, d) quantiles of aggregated load profiles

3.4 Demonstration and Results

In this section, the uncertainty quantification result for each component will be firstly presented as energy density spectrum (EDS) matrices. The visualization result provided in uncertainty characterization stage will be followed.

3.4.1 Results of Uncertainty Quantification Stage

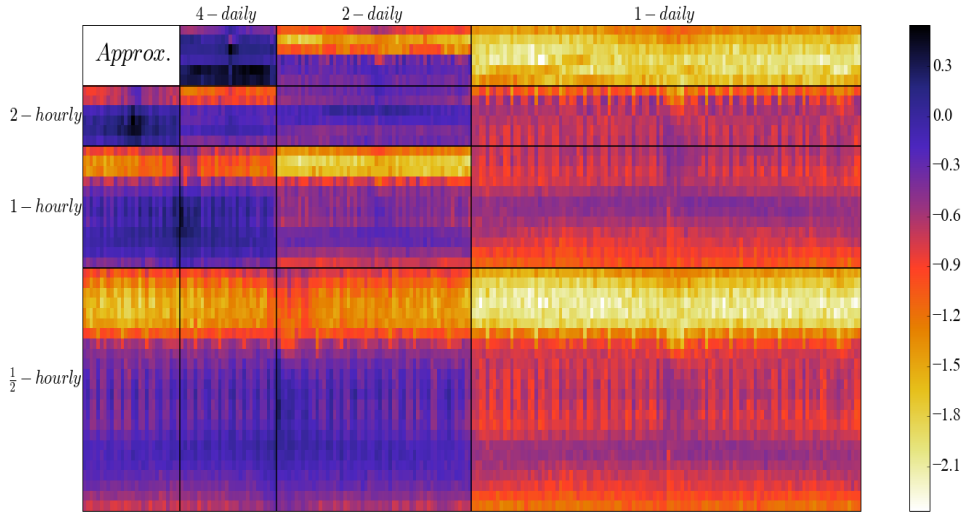


Figure 3-4 Heatmap of the logarithm of average energy density spectrum across 900 residential households

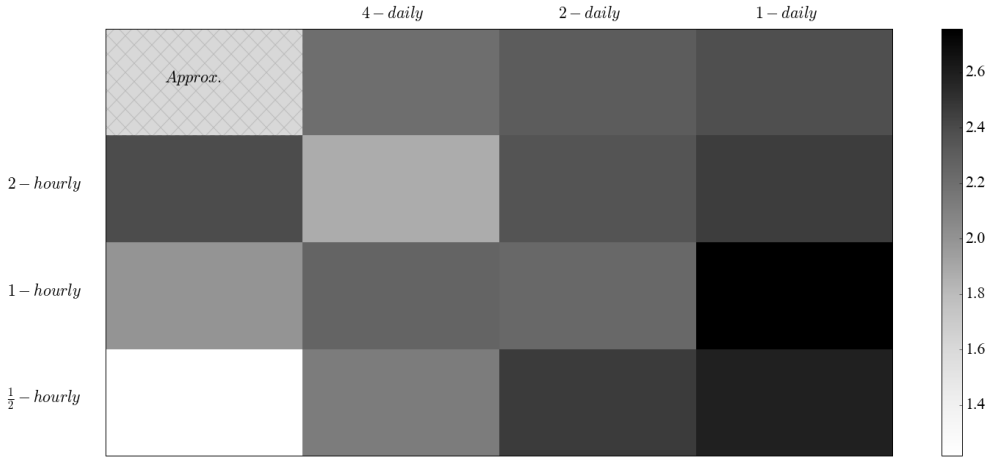


Figure 3-5 Heatmap to present logarithm of the sum of squared coefficients in cross-time-scale channels

As shown in Heatmaps in **Figure 3-4**, the coefficients matrix of uncertainty components at each time scales are presented. These coefficients are quantified with the logarithm of mean energy density spectrum (EDS). Rows of index represent different resolution in hourly scales, i.e., half-hourly, hourly, and two-hourly time scales. Similarly, Columns of the matrices means daily, two-daily and four-daily time scales. Notably, the top left block noted with 'Approx.' is the approximation of Haar expansion after three levels of decomposition, which holds the lower frequency components of the time series data. Summing up the mean-squared coefficients in each selected time scale,

uncertainty in variance can be quantified accordingly, and this result is presented in **Figure 3-5 3-5**. Since this cross-channels between time scales can indicate the correlation degree, or in other words, the dependency. Lighter blocks refer to the stronger connection between the involved two-time scales [80].

3.4.2 Visualizations of Uncertainty Characterization Stage

As mentioned in the methodology section, this part will look into: i) uncertainty temporal distribution of each time scale and uncertainty temporal distribution of each cross-channels between any two-time scales; ii) and how uncertainty propagates under different aggregation scales from households to the system.

3.4.2.1 Uncertainty Temporal Distributions

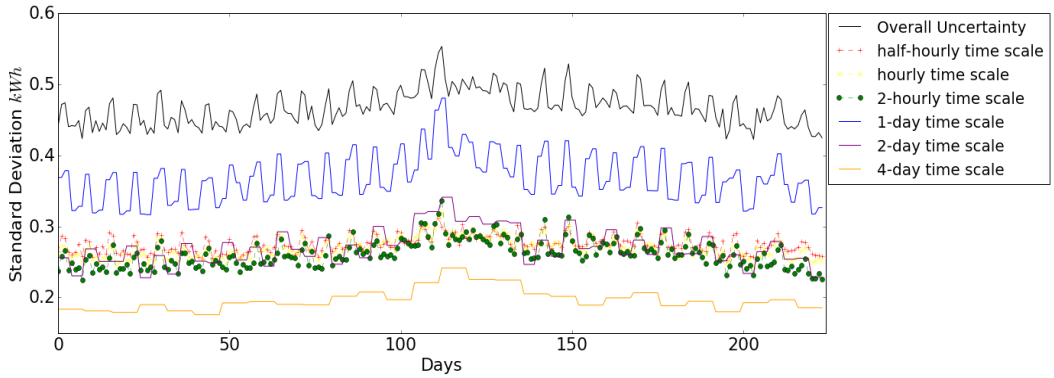


Figure 3-6 Daily uncertainty distribution of different time-scale channels from 14th July. 2009 to 31st Dec. 2010

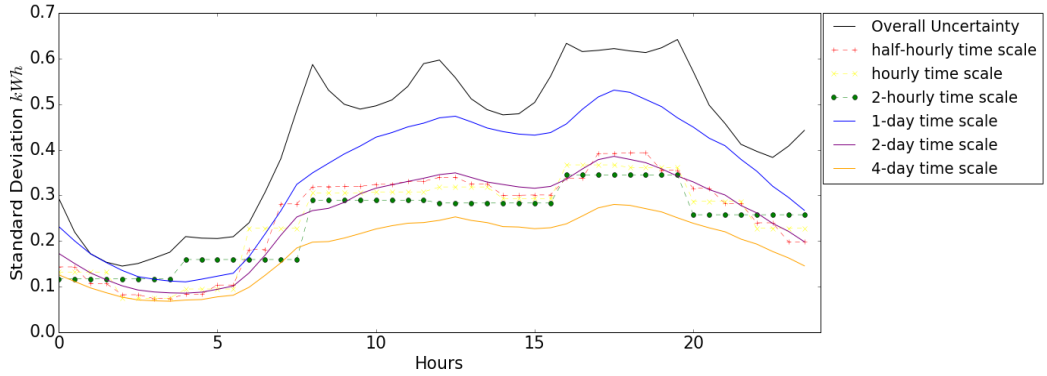


Figure 3-7 Hourly uncertainty distribution of different time-scale channels over a day

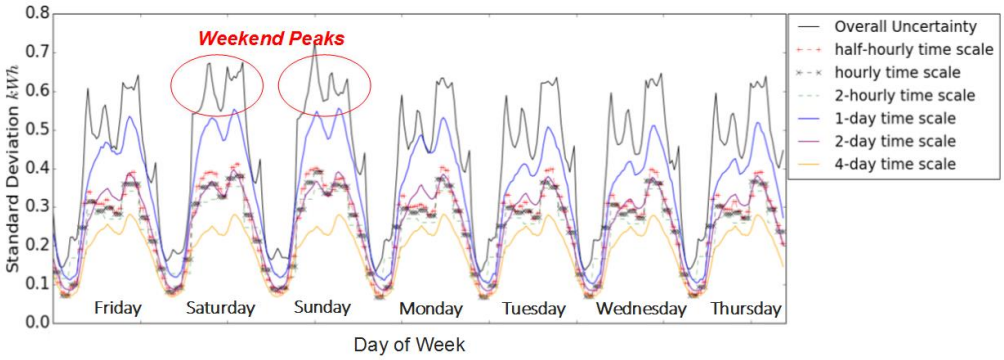


Figure 3-8 Hourly uncertainty distribution of different time-scale channels over a week

Figure 3-6 presents the temporal uncertainty distribution along with 224 days. Two findings are found from the results: i) daily uncertainty component has the largest proportion; ii) half-hourly, hourly, two-hourly and two-daily time scale have similar uncertainty quantity.

Figure 3-7 shows the temporal uncertainty distribution along with daily periodicity. Two findings are reported: i) half-hourly, hourly and two-hourly time scales have similar uncertainty quantity, while daily time scales have larger uncertainty than two-daily and four-daily; ii) a common pattern is witnessed across all time scales, with noon and evening peaks of uncertainty.

Figure 3-8 presents the weekly uncertainty temporal distribution from Friday to Thursday. As annotated in red circles, it indicates load peaks on weekends are slightly higher than peaks in the weekdays. And Mondays and Fridays have its unique pattern compared to the rest three weekdays.

Figure 3-9 presents two heatmaps on two-dimensional uncertainty temporal distribution. **Figure 3-9** a) shows the uncertainty temporal distribution in days across 32 weeks in columns and 7 days of the week in rows.

b) shows the uncertainty temporal distribution in hours of a week periodicity, with 7 days of the week in rows and 48 time-points (half-hourly sampled) of the day in columns. The results indicate: i) demand uncertainty has two peaks, a higher one in the evening between 16:00 p.m. - 19:30 p.m., and the daytime peak is between 9:00 a.m. - 14:00 p.m.; ii) uncertainty peak emerges on any of Wednesdays, Thursdays and Winter season.

Figure 3-10 presents the uncertainty temporal distribution at cross channels between two different time scales. This result is in consistent with **Figure 3-5** that stronger dependencies between two-time scales will lead to less uncertainty in the corresponding cross-channels.

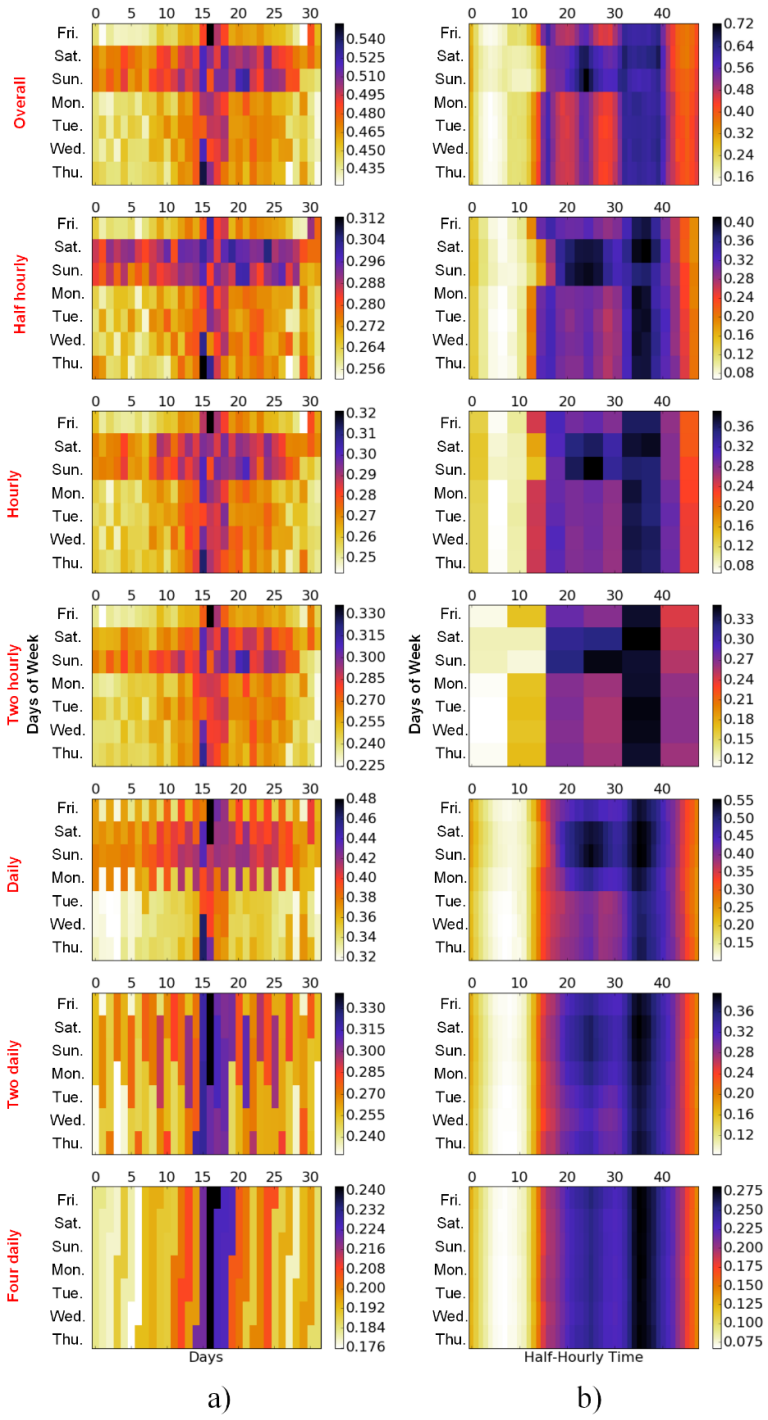
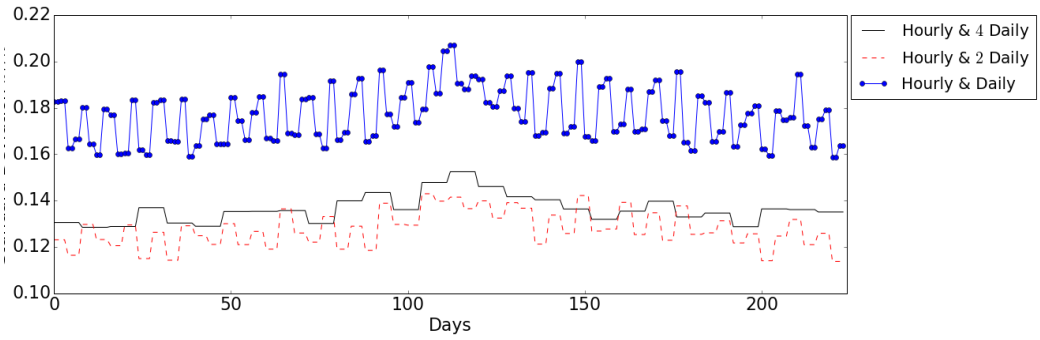
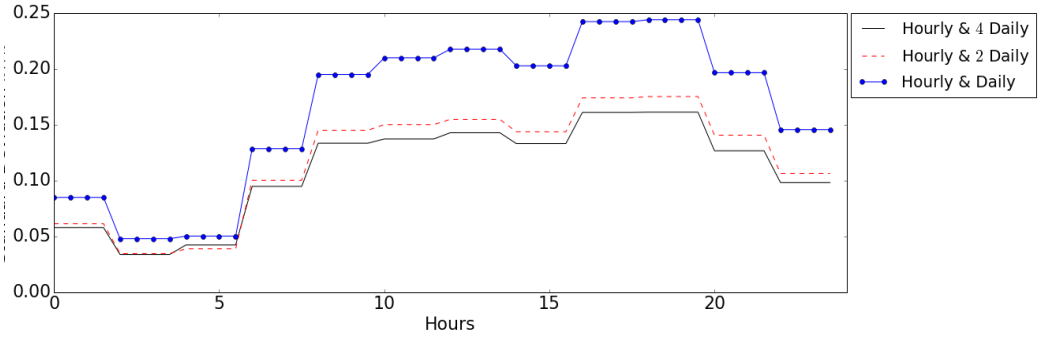


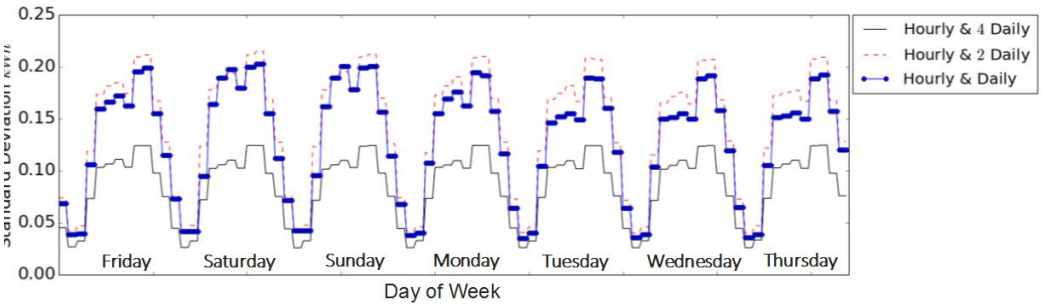
Figure 3-9 Uncertainty heatmap of multiple time-scale channels: a) Annual distribution in day indexes; b) weekly distribution in hour indexes



a)



b)



c)

Figure 3-10 Comparison of uncertainty distributions of two cross channels in different periods: a) annual period, b) daily period, c) weekly period.

3.4.2.2 Uncertainty Propagation Analysis

By simulating the system aggregation process, comparisons can be done on domestic loads and system aggregated load, to see how much ratio of uncertainty is mitigated through the propagation. As discussed in the former section of methodology, the ratio of mitigation through propagation is expected to be correlated to the average pair-wise

PCCs across all domestic households. Therefore, the average statistics of variances, covariances, and PCCs are summarized in **Table 3-2**

Table 3-2 Average Statistics (covariance, variance and PCCs) across 900 Residential Households

Time scales	*AVERAGE PAIR-WISE COVARIANCE*	*Variance*	*Average Pair-wise PCCs*
Overall Uncertainty	0.0082	0.2178	0.0394
Half-hourly time scale	0.0006	0.0760	0.0085
Hourly time scale	0.0025	0.0732	0.0356
Two-hourly time scale	0.0051	0.0685	0.0774
Daily time scale	0.0012	0.1368	0.0081
Two-daily time scale	0.0013	0.0743	0.0159
Four-daily time scale	0.0007	0.0392	0.0171

Figure 3-11 shows the Uncertainty Propagation Factor (UPF) measured at different aggregation scales with simulations, the aggregation scales vary in the range between 10 households and 900 households. The result in **Figure 3-11** and Table I indicates that: i) uncertainties in disaggregated level can mitigate through system aggregation, the mitigating rate will approach an upper limit when the aggregation scale is large enough. In the given example, aggregation with 200 households is the threshold to the mitigation limit, with 9o% of the uncertainty mitigation effect (0.1 in UPF); ii) uncertainty components at different time scales have different potential UPF (infimum of UPF). In particular, Half-hourly and daily time scales have largest UPF limits at around 90%; iii) the potential UPF limit is positively related to the average PCCs. The possible range of the limit of UPF measure is between 0 to 1; iv) two extreme scenarios are: if UPF=0, household loads are independent, identically distributed random variables; if UPF =1, all the household loads are clones to one load profile prototype.

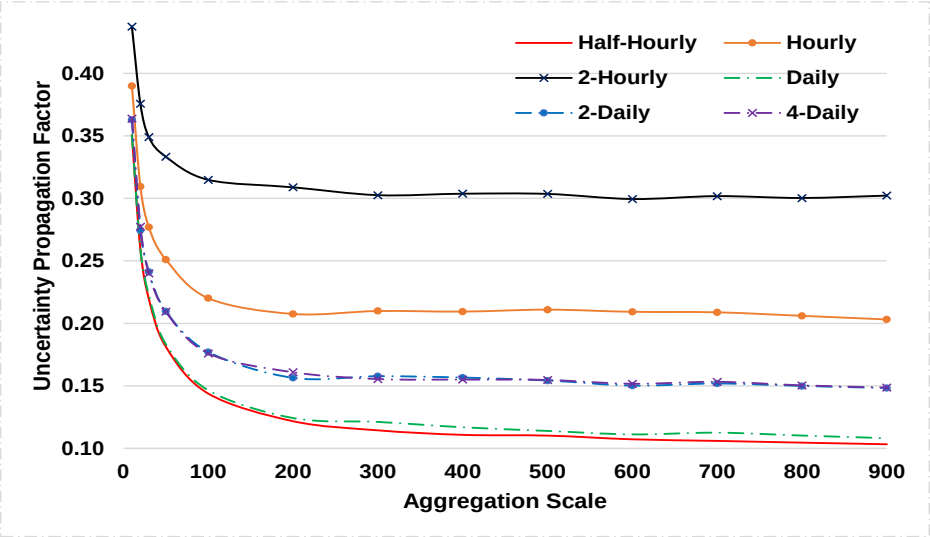


Figure 3-11 Uncertainty propagation factor in differing aggregation scales

Figure 3-12 shows the NUPF measures corresponds to the aggregation scales investigated in **Figure 3-11**. The NUPF measures are witnessed strong linear relationship with the aggregation scale. According to formulation (10) and (11), the approaching line of NUPF is determined by the average PCCs across all households. Notably: i) for the scenario when household loads are independent, identically distributed random variables, NUPF will be transformed into a horizontal line $l_{SUMR} = 1$; ii) if all demands are clones to a single prototype, NUPF is proportional to $\sqrt{N}$.

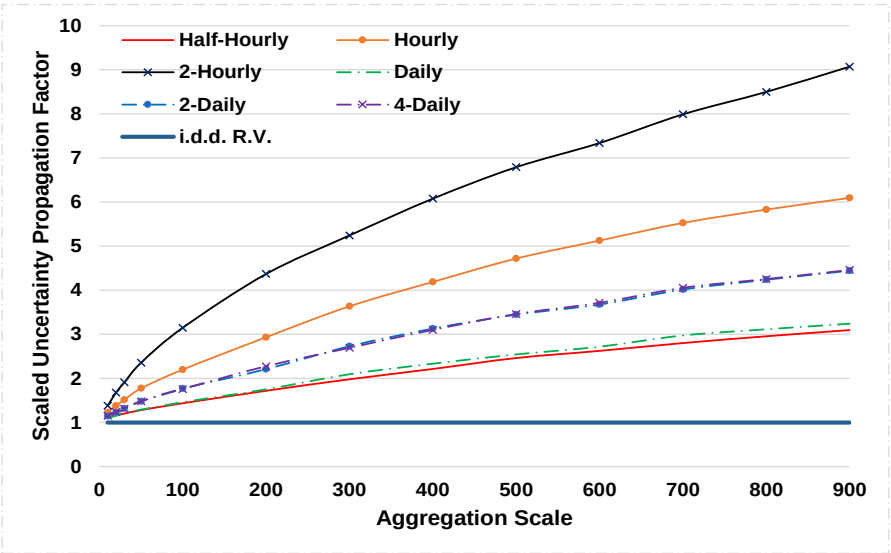


Figure 3-12 Normalised uncertainty propagation factor in differing aggregation scales

3.5 Chapter Summary

This chapter proposes a novel analytical tool for performing forward UQ on individual uncertainty components at multiple time scales. The results shown in this chapter is demonstrated particularly on a smart metering dataset of Irish residential households collected nationwide [37]. According to the results, temporal characteristics of load uncertainties can be summarized as: i) demand uncertainties is roughly proportional to the average demand consumption; ii) demand uncertainty has two peaks, a higher one in the evening between 16:00 p.m. - 19:30 p.m., and the daytime peak is between 9:00 a.m. - 14:00 p.m.; iii) uncertainty peak emerges on any of Wednesdays, Thursdays and Winter season; iv) stronger dependencies (inter-day uncertainties) between two time scales will lead to less uncertainty in the corresponding cross-channels. The results in uncertainty propagation analysis indicate how uncertainty mitigated through system aggregation. In details: i) the rate of uncertainty mitigated through aggregation has an upper limit, and this limit will be approached with sufficient aggregation scale; ii) the potential UPF limit is positively related to the average PCCs. Specifically, half-hourly and daily time scales have the most UPF limits at around 90%.

This finding in the uncertainty propagation process provides useful information in the latter chapter V, to develop advanced home EMS and let uncertainty components mitigated automatically to achieve peak and uncertainty reduction at grid level.

Chapter 4

Inverse Uncertainty Quantification for Modelling Domestic Load Uncertainty

THIS proposed a deep learning model for achieving inverse uncertainty quantification, demonstration is given on domestic load uncertainty, to modelling the underlying uncertainty components of load uncertainty, and track its causing sources.

4.1 Introduction

As mentioned in former chapters, inverse UQ is still an untouched area in the electrical power system, due to the lack of motivation in traditional power system and the difficulty to implement in a real-world problem. As discussed in Chapter 4, the uncertainty of electricity demand is a combination of various components caused by different external sources. This complex mixture largely sets obstacles for understanding demand uncertainty easily and intuitively.

In Chapter 4, this thesis focuses on one alternative for exploring quantification and characterization on components of uncertainty. In details, it disaggregates demand along with temporal domain, with decomposition techniques. Therefore, each component of uncertainty represents a single timescale or a cross-channel between two individual time scales.

In this chapter, we attempt to decompose uncertainty according to a different angle, i.e., its uncertainty sources, and models the relationship between uncertainty components and corresponding uncertainty sources. Inverse UQ provides natural tools for understanding the relationship between uncertainty and its causing sources, hence, this chapter will develop and implement a novel method for forward UQ, and demonstrated in domestic load uncertainty.

The later parts of this Chapters are followed by: Section 4.2 introduces the concept of inverse UQ. Section 4.3 presents the major challenges that lie in the current literature of inverse UQ in other fields. Section 4.4 introduces the proposed method of inverse UQ implemented with Deep Learning model, named as Probabilistic Deep Dropout Generative Nets (PDDGN). The experiments and results are demonstrated in Section 4.5. Chapter Summary follows in Section 4.6.

The content of this chapter is cited from the author's manuscript to be submitted to IEEE Transactions on Smart Grid, named as 'Understand Load Uncertainty with Deep Learning-A Probabilistic Deep Dropout Generative Nets'.

4.2 Introduction on Inverse Uncertainty Quantification

Uncertainty Component is a widely discussed concept in uncertainty measurement, which is firstly proposed in the Guide to the Expression of Uncertainty in Measurement (GUM) [11] in 1993. It regards uncertainty as a combination of multiple components, and these components can be classified into two categories according to its causing sources and natures, i.e., knowledge uncertainty or random uncertainty.

Knowledge Uncertainty is also called as systematic uncertainty and epistemic uncertainty, which means knowledge in Greek [18, 19, 20, 21, 22, 23]. The sources of this type of uncertainties normally come from systematic effect from certain causing factors, and reducible with deep understandings of the mechanism and knowledge behind the systematic effect. For instance, the impact of temperature on electricity demand is a kind of knowledge uncertainty, where in the winter, lower temperatures normally indicate higher electricity consumption for heat. Random Uncertainty has another name as aleatoric uncertainty, which means dice player in Latin [18, 23]. This type of uncertainties is caused by random factors or pure randomness, and irreducible due to its random nature. For instance, the measurement error of electricity demand caused by smart meters devices is a typical random uncertainty. Although we know the sources and causes of measurement error, it is still inevitable.

Therefore, uncertainty can be regarded as the combination of various components contributed from different uncertainty sources. Some components are caused by systematic effect and can be classified into knowledge uncertainties, whilst the other components are random uncertainties. To this end, an essential problem is how to model these uncertainty components, and further discriminate contributions from different uncertainty sources, i.e., Uncertainty Components Modelling (UCM).

4.3 Challenge in Inverse Uncertainty Quantification

Out of energy community, UCM is also an essential problem ranging from social sciences, economics, chemistry to complex environmental systems. To the existing literature, prior methods can be summarized into threefold:

Statistical Approaches: some works deploy statistical approaches to quantifies the proportions of uncertainty components numerically. For instance, Zhang, Jing, et al [81] employs Random Forest and feature selection to evaluate the impact different factors to the stock price variability; AD Atkinson, et al, [82] attempts to model the system model biases with ANOVA methods. Despite advantages of these statistical methods, such as simplicity and generality, however, it remains two limitations: i) outputs of the prior statistical method are a numerical number to indicate the number of uncertainty components rather than intuitive probability distributions; ii) and it also cannot evaluate the components caused by unobservable/unknown factors.

Analytical Approaches: analytical approaches are also widely deployed to model uncertainty components. Work [83] adopted Latent Dirichlet Allocation (LDA) to transform the complex textual data into latent representation and allocate the aggregated variability into various causing factors. K. Beven et al [84], employs the generalized likelihood uncertainty estimation (GLUE) methods to decompose the uncertainties into components in the form of probability distribution. Compared to statistical approaches, it can obtain intuitive probability distribution as output. However, the major limitation of analytical methods is the trade-off between generality and modelling capability affected by model selection [85]. More specifically, simple models can be easily generalized to different data but have inadequate modelling capability, while complex models can perform well at certain datasets but lack of generality. Due to this limitation, most of the existing methods are ad-hoc methods specifically designed for certain datasets.

Given the physics that demand uncertainty is caused by a variety of external factors as uncertainty sources, ranging from uncertain customer activities, measurement errors of monitoring devices, interventions of techniques, and so on. Some of these uncertainty sources are observable, while others are unobservable/unmeasurable. For instance, the temperature can be observed and collected as a companion with the demand consumption readings at the time, whilst measurement error cannot be collected and predicted. Therefore, the major challenges for the UCM of electricity demand are mainly threefold:

- the high complexity and non-linearity between demand uncertainty and its sources;

- the mutual dependency between different uncertainty sources;
- and the lack of data on unobservable/unmeasurable uncertainty sources.

These challenges limit the practicality of existing analytical methods due to the lack of generality. In addition, existing statistical approaches are not of best choice since they cannot give intuitive results of interests.

4.4 Proposed Methodology for Inverse Uncertainty Quantification

To achieve uncertainty components modelling on electricity demand, it calls for an advanced tool that has both high generality and modelling capability. Therefore, this thesis for the first time to explore the possibility of deploying the state-of-the-art deep learning techniques to achieve uncertainty components modelling. There are two rationales for using deep learning techniques.

On the one hand, deep learning can achieve superior learning capabilities for modelling highly non-linear and complex functions. Given sufficient scales of network structure and data volumes, the deep neural network is capable to approximate any complex non-linear relationships between inputs and outputs [86, 87, 88]. Hence, this thesis makes use of the learning capability of deep learning to model and learn demand uncertainty directly regarding its highly complex and non-linear relationships between the demand uncertainty and its sources.

On the other hand, deep learning neural networks do not map the inputs directly to outputs. In fact, it firstly maps inputs to the features space, and hereafter maps to the outputs. With this feature, not only the observable sources can be modelled as the inputs of deep neural networks, but also the unknown/unobservable sources can be modelled and estimated as the latent uncertainties over the feature space.

In this part, the author will present the proposed PDDGN model in three steps: i) briefly introduces the concepts of probabilistic modelling and its application in Deep Learning community; ii) introduce the architecture of proposed PDDGN model; iii) and proves the efficacy of proposed PDDGN model theoretically.

4.4.1 Probabilistic Modelling with Deep Learning

Deep neural networks literally refer to the multi-layer neural network models connecting multiple processing layers sequentially, to learn complex and abstract features from given datasets. Traditionally, classical deep neural networks are a deterministic model rather than probabilistic models, since it is normally designed to predict the best estimation of uncertain outputs, which generates a deterministic number/vector as the output. However, this thesis seeks for a probabilistic model of deep neural networks, in other words, the proposed deep neural networks should be able to learn and generate probability distribution. This section discusses the probabilistic deep learning model and its theoretical basis: Bayesian Modelling and Variational Inference (VI). Bayesian Modelling is highly-rated since it can cope with most complex problems in the real world due to its model capacity. However, the implementation of Bayesian Modelling heavily suffers from its expensive computational costs caused by extensive sigma functions and cross-corpora calculations. In addition, they perform poorly with very small data sets. Therefore, in this section, Deep Learning models are employed as the model basis for Bayesian Modelling. Dropout Regularization units are specifically applied to introduce random features to the enable the deterministic deep learning model for probabilistic Bayesian Modelling. The efficacy and effectiveness are discussed and demonstrated in following sections.

4.4.1.1 Probabilistic Modelling based on Bayesian Modelling

Classical deep learning models map model inputs to the corresponding outputs in the form of deterministic numerical vectors. Given model input variables and output variables as variables x and y, classical deep neural networks can be simply regarded as a synthesized function $f(*)$. Specifically, this function f is the combination of matrices multiplication and activation functions of each processing layer. The model parameters are denoted as $\theta = \{W, b, \dots\}$, which consists of weight matrices, bias matrices and other parameters of certain network architectures. Therefore, the model outputs can be obtained given input variable x on condition of parameter vectors θ: $y = f(x; \theta)$, where both x, y are deterministic variables, which indicates classical deep learning models are deterministic models.

In terms of the probabilistic model, both model inputs and outputs are further considered as random variables. To express notations in consistency, a tilde accent is added to notations to replace the corresponding deterministic variables with random variables, i.e., $\tilde{y}$ refers to the uncertainty of model output. In this case, the deterministic model is predicting an estimate on the expectation of probability distribution of uncertain output $\tilde{y}$, i.e.:

$$\mathbb{E}_\theta[\tilde{y}] = \mathbb{E}_\theta[f(x;\theta)] = \mathbb{E}_\theta[p(\tilde{y}|x,\theta)] \quad (4\text{-}1)$$

Assuming the probabilistic model of the corresponding deep neural network is f', the probabilistic outputs of the deep neural network can be then formulated as:

$$\tilde{y} = f'(x;\tilde{\theta}) \sim p(\tilde{y}|x;\tilde{\theta}) \quad (4\text{-}2)$$

where the conditional distribution $p(\tilde{y}|x;\tilde{\theta})$ is the probabilistic model of the deep neural network. The comparison between the classical deterministic model and probabilistic model of deep neural networks are shown in **Figure 4-1** below:

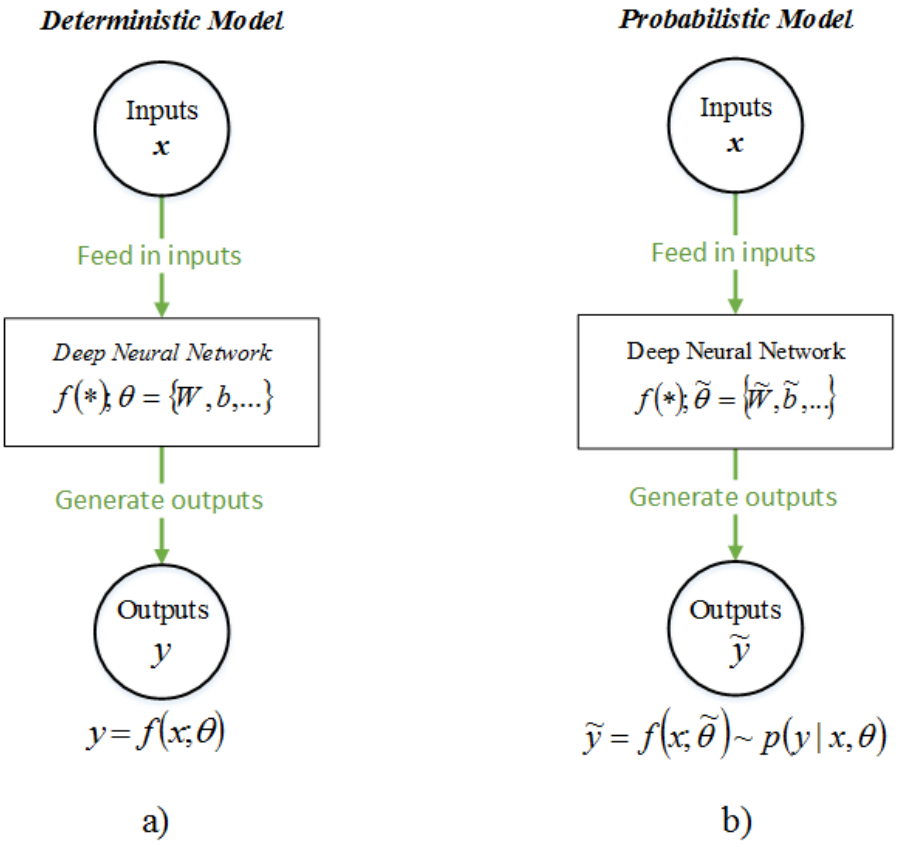


Figure 4-1 Illustration of two types of deep neural networks: a) deterministic model; b) probabilistic model.

Given historical datasets X, Y as training data, we can obtain model parameters that are most likely to generate label of dataset Y from probabilistic deep neural network given inputs X. Mathematically, we look for the posterior distribution of model parameter $p(\tilde{\theta}|X,Y)$ on given dataset X, Y. By invoking Bayes' theorem, the posterior distribution of interest can be formulated by:

$$\tilde{y} = f'(x;\tilde{\theta}) \sim p(\tilde{y}|x;\tilde{\theta}) \quad (4\text{-}3)$$

$$p(\tilde{\theta}|X,Y) = \frac{p(Y|X;\tilde{\theta})p(\tilde{\theta})}{p(Y|X)} \quad (4\text{-}4)$$

With appropriate assumption on the prior distribution of parameters $p(\tilde{\theta})$, the posterior distribution of model parameters can be more easily estimated with integrals over training datasets. This process is widely known as Bayesian Modelling [89, 90, 91].

For the testing realizations of inputs x^*, the prediction on the targeted distribution of y^* can be accordingly inferred as:

$$p(y^*|x^*;X,Y) = \int p(y^*|x^*;\tilde{\theta})p(\tilde{\theta}|X,Y)d\tilde{\theta} \quad (4\text{-}5)$$

Notably, $p(y^*|x^*;\tilde{\theta})$ is the output distribution of the probabilistic model of the deep neural network, which can be simply sampled and simulated by feed in realizations of x^* to the well-tuned deep neural network. Therefore, the targeted posterior distribution $p(y^*|x^*;X,Y)$ can be easily implemented by sampling approaches such as Gibbs Sampling [92].

4.4.1.1 Variational Inference to Train Probabilistic Models

As discussed in the previous section, to invoke feed-forward prediction of the probabilistic model, the key component is to obtain the posterior distribution $p(\tilde{\theta}|X,Y)$ with given data samples. Normally, this distribution cannot be modelled directly by deep networks due to the fixed structure of deep networks. To approximate the targeted distribution with the probability model of deep network $q(\tilde{\theta})$, variational inference is introduced in the training process to minimize the difference between $p(\tilde{\theta}|X,Y)$ and $q(\tilde{\theta})$.

A measure that indicates the similarity between the targeting distribution $p(\tilde{\theta}|X,Y)$ and the distribution of the deep neural network $q(\tilde{\theta})$ can be evaluated by the KL divergence:

$$KL(q(\tilde{\theta})||p(\tilde{\theta}|X,Y)) = \int q(\tilde{\theta})\frac{q(\tilde{\theta})}{p(\tilde{\theta}|X,Y)}d\tilde{\theta} \quad (4\text{-}6)$$

The KL divergence approaches its minimum when the distribution of deep learning model $q(\tilde{\theta})$ is close to the targeting posterior distribution $p(\tilde{\theta}|X,Y)$, which can be denoted as $q^*(\tilde{\theta})$. Therefore, we can replace the posterior distribution with neural network model in equation (4-13) to (4-19):

$$p(y^*|x^*, X, Y) = \int p(y^*|x^*, \tilde{\theta}) q^*(\tilde{\theta})\, d\tilde{\theta} \quad (4\text{-}7)$$

This process means the probability model of output y^* with given inputs x^* can be quantified by sampling through the deep learning model with variational parameter $\tilde{\theta}$ that follows the distribution $q^*(\tilde{\theta})$.

Notably, the KL divergence is intractable in many cases due to the posterior distribution in the integral form. Hence, minimizing the KL divergence is replaced by an equivalent formulation, i.e., evidence lower bound (ELBO) [93].

$$\mathcal{L}(\tilde{\theta}) := \int q(\tilde{\theta}) log\, p(Y|X, \tilde{\theta}) d\tilde{\theta} - KL(q(\tilde{\theta})||p(\tilde{\theta})) \quad (4\text{-}8)$$

Through optimization that maximizes the ELBO term, the KL divergence between $q^*(\tilde{\theta})$ and $p(\tilde{\theta})$ can be accordingly minimized. Hence, the feed-forward of proposed model can be further simplified into compact form as:

$$p(y^*|x^*, X, Y) = \int p(y^*|x^*, \tilde{\theta}) q^*(\tilde{\theta})\, d\tilde{\theta} \approx q^*(y^*|x^*, \tilde{\theta}) \quad (4\text{-}9)$$

4.4.2 Probabilistic Deep Dropout Generative Nets

According to the probabilistic modelling discussed in the last section, deep generative networks can: i) represent uncertainties with proper settings on network architecture and the parameter; ii) learn uncertainties from historical datasets with VI; iii) and predict uncertainties by invoking inference in the feed-forward network.

In a deep learning community, there exists many generative architectures, ranging from Variational Auto-encoder (VA), Generative Boltzmann Machine (GBM), Generative Adversarial Nets (GAN), and so on. Among these architectures, the most simple and intuitive approach to introduce probabilistic natures into specific models is to replace the deterministic model parameters, such as matrices weights, biases, i.e., $\{\widetilde{W}, \tilde{b}, \dots\}$, with probabilistic parameters. Typically, put a standard Gaussian distribution $\mathcal{N}(0,1)$ on those matrices of parameters. For any single epochs at the training and testing stage, make sampling on these random parameters to form a deterministic network. The mentioned models that adopts this approach is categorized in the family of Bayesian

Neural Network (BNN), which has been proposed for decades. However, two fatal shortages of this approach are the excessive computational costs that: i) paid for performing variational inference [47, 48, 94, 95]; and ii) required to in order to capture the detail features of the posterior distribution of outputs by sampling.

This thesis attempt to model uncertainty components of electricity demand under realistic challenging circumstance: i) lack of systematic knowledge, and ii) lack of data on many unknown/latent uncertainty sources. BNN is hence impractical for our needs. Therefore, this for the first time to explore alternative feasible deep learning tool to implement UCM under realistic challenges.

The solution of this thesis is a well-developed and well-investigated technique in a deep learning community, i.e., dropout Stochastic Regularization Techniques (DSRTs). Traditionally, dropout is an ad-hoc technique specially designed for regularizing ill-posed solutions due to the over-fitting problem [87, 96, 97]. In this thesis, this technique is newly deployed to introduce randomness in the feature space and alternatively implement UCM of electricity demand with deep neural networks. This section will further review the practicality of DSRT for the purpose of probabilistic model and VI [88, 91, 93].

In this section, the proposed probabilistic deep dropout neural network (PDDGN) is presented for uncertainty components modelling of electricity demand. The proposed method is discussed with three parts of contents: i) network architecture of proposed deep learning model based on multilayer perceptron (MLP) and dropout stochastic regularization techniques (DSRTs); ii) learn uncertainty components during training process; iii) extract uncertainty components with model feed-forward sampling.

4.4.2.1 Network Architecture of proposed PDDGN model

The fundamental idea of proposed PDDGN model is to integrate classical multilayer deep neural network with dropout SRT units. The computational graph is illustrated in **Figure 4-2**:

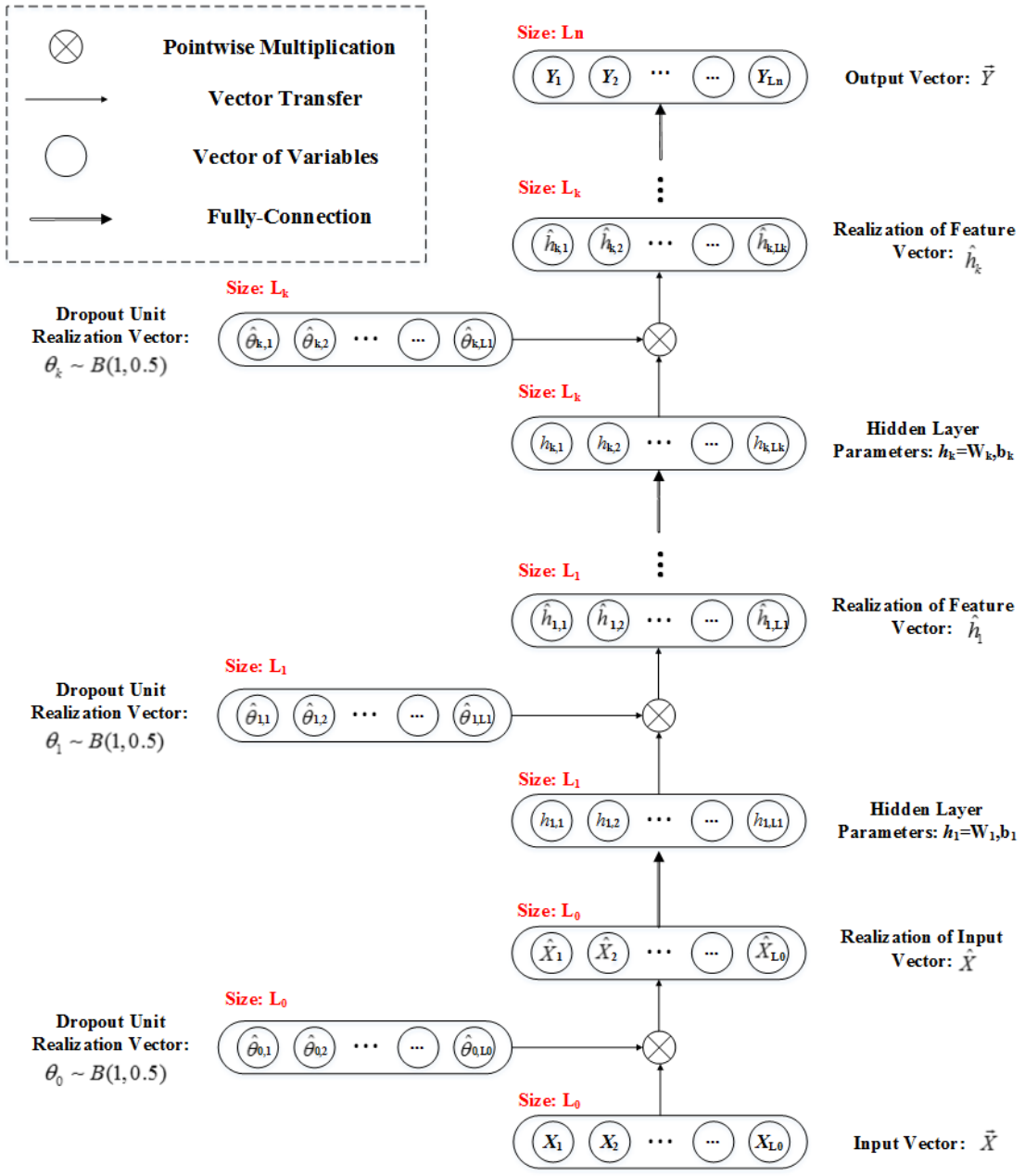


Figure 4-2 Computational graph of proposed probabilistic deep dropout generative network: a) deterministic model; b) probabilistic model.

As shown in the computational graph, the basic network architecture is a multilayer perceptron (MLP) [87, 98] classifier removing the Softmax [87, 99] unit before output Y, to achieve regression purpose to the electricity demand readings. The original equations that describes this basic network architecture with N layers can be formulated as:

$$h_1 = \Phi(b_1 + W_1 \cdot x) \qquad (4\text{-}10)$$

$$h_k = \Phi(b_k + W_k \cdot h_{k-1}) \qquad for\ l = 2,3,\dots,N \tag{4-11}$$

$$y = \Phi(b_{N+1} + W_{N+1} \cdot h_N) \tag{4-12}$$

$$L = L2\_loss(y, y_{target}) \tag{4-13}$$

where $\Phi(*)$ refers to the activation function, which is normally deploying Sigmoid, ReLu, Tanh functions.

In the proposed model, we added dropout SRTs unit between layers, to introduce probabilistic natures to the model. In general, dropout units are well developed and tested in an extensive literature. It is designed and proved to effectively alleviate over-fitting issue in regression problems regularizing ill-posed solutions [96]. However, it was firstly researched by Y. Gal [88, 91, 93] to introduce uncertainty into the deterministic neural network models. In recent Gal's works [91], the attributes and natures of model uncertainty brought by dropout SRTs units have been comprehensively examined. The efficacy and validity DSRTs are also proved in the work. Compared to traditional BNNs, deep learning models with dropout SRTs units have three advantages: i) DSRTs is a decent technique that can well integrate with mini-batch training method to guarantee training efficiency; ii) the probabilistic nature of DSRTs are represented by a random quantity rather than a probability distribution, which can largely reduce the data volume required to converge to the solution; iii) DSRTs is simple and practical to be implemented with existing deep learning platforms, such as Tensorflow [100] in this thesis.

With DSRT integrated, the equations of proposed model can be re-written as:

$$\hat{x} = \hat{\omega} \cdot x \tag{4-14}$$

$$h_1 = \phi(b_1 + W_1 \cdot \hat{x}) \tag{4-15}$$

$$\hat{h}_1 = \hat{\omega}^{(1)} \cdot h_1 \tag{4-16}$$

$$h_k = \sigma(b_k + W_k \cdot \hat{h}_{k-1}) \qquad for\ l = 2,3,\dots,N \tag{4-17}$$

$$\hat{h}_k = \hat{\omega}^{(k)} \cdot h_k \tag{4-18}$$

$$y = \phi(b_{N+1} + W_{N+1} \cdot \hat{h}_N) \quad (4\text{-}19)$$

$$L = L_2\_loss(y, y_{target}) \quad (4\text{-}20)$$

Notably, the cap notation added to variables indicate the vector of realizations of this random variable masked by dropout units. For instance, $\hat{x}$ refers to a realization of x by considering x as a random quantity. Specifically, binary vector $\hat{\omega}$ refers to the realization of dropout units, which is obtained given dropout rate p and vector length l. By sampling a random germ r that follows the standard uniform distribution $U(0,1)$, each binary element $\hat{\omega}^{(k)}$ of k^{th} level depth can be accordingly determined as:

$$\hat{\omega} = \{\omega_1, \omega_2, \dots, \omega_l\} \quad (4\text{-}21)$$

$$\omega_j = \begin{cases} 1 & r > p, r \sim U(0,1), 0 \le p < 1 \\ 0 & r \le p, r \sim U(0,1), 0 \le p < 1 \end{cases} \quad (4\text{-}22)$$

In other words, by introducing DSRT units to the network, each neuron will averagely have $1 - p$ probability to make impact the output at each epoch. Therefore, with sufficient data volume and training epochs, the probabilistic nature of proposed model can be alternatively statistically simulated by realizations of random quantities generated by dropout units.

4.4.2.2 Learning Uncertainty Components with PDDGN model

This section discusses the mini-batch training process [87] of the proposed PDDGN model and proves the equivalence to the VI. Notably, Y Gal [91, 93] has already proved that dropout SRT can be used as the MC estimator for performing VI in the framework of Deep Learning, this provides the theoretical basis for the proposed PDDGN model in this thesis. This subpart will follow the proof presented in [91], to generalise to the application of uncertainty components modelling (inverse UQ).

With respect to the realizations of model output: $\hat{y}$, it can be determined by sampling realizations of dropout units $\hat{\omega}^{(k)}$ from the prior distribution $p(\omega)$.

$$\begin{aligned} \hat{y} = \phi(b_{N+1} + W_{N+1} \cdot \hat{h}_N) = \phi(b_{N+1} + \widehat{W}_{N+1} \cdot h_N = f(x; \widehat{W}_1, \widehat{W}_2, \dots, \widehat{W}_{N+1}, b) \\ = f(x; \hat{\omega}^{(1)}, \dots, \hat{\omega}^{(k)}, \theta) \end{aligned} \quad (4\text{-}23)$$

Where the model parameter θ refers to the weight and bias matrices $\theta = \{W_1, \dots, W_{N+1}, b\}$. In other words, the realization of model outputs is a function of input x, on condition to model parameters θ and dropout SRT units Boolean realization vectors $\widehat{\omega}^{(1)}, \dots, \widehat{\omega}^{(N+1)}$. The model function $f(*)$ describe the feed-forward network of proposed PDDGN model. However, the optimization will minimize the L_2 loss, which can be rewritten as log-likelihood [91]:

$$\begin{aligned} L\left[f\left(x; \widehat{W}_1, \widehat{W}_2, \dots, \widehat{W}_{N+1}, b\right)\right] &= \frac{1}{2}\left\|y - f\left(x; \widehat{W}_1, \widehat{W}_2, \dots, \widehat{W}_{N+1}, b\right)\right\|^2 \\ &= -\sigma^2 logp\left(y|f\left(x; \widehat{W}_1, \widehat{W}_2, \dots, \widehat{W}_{N+1}, b\right)\right) + const \\ &= -\sigma^2 logp\left(y|f\left(x; \widehat{\omega}^{(1)}, \dots, \widehat{\omega}^{(k)}, \theta\right)\right) + const \end{aligned} \quad (4\text{-}24)$$

Notably, the prior distribution of residual between model output and target are gaussian distribution follows the unbiased expectation and standard deviation σ.

$$p\left(y|f\left(x; \widehat{W}_1, \widehat{W}_2, \dots, \widehat{W}_{N+1}, b\right)\right) = \mathcal{N}\left(y; f\left(x; \widehat{\omega}^{(1)}, \dots, \widehat{\omega}^{(k)}, \theta\right), \sigma^2\right) \quad (4\text{-}25)$$

With mini-batch training process, assuming the sample size for each batch is N while the size for the whole dataset S is M. For each epoch in the mini-batch Monte Carlo optimization, a realization of mini-batch MC estimator with respect to model parameters θ is formulated as [91]:

$$\hat{\mathcal{L}}_{mini-batch}(\theta) = -\frac{N}{M}\sum_{i \in S} logp\left(y|f(x; \widehat{\omega}, \theta)\right) \quad (4\text{-}26)$$

Comparably, the ELBO realization of VI processes with a size N batch can be denoted as:

$$\hat{\mathcal{L}}_{VI}(\omega, \theta) = -\frac{N}{M}\sum_{i \in S}\int q_\theta(\omega) logp\left(y|f(x; \widehat{\omega}, \theta)\right) d\omega + KL(q_\theta(\omega)||p(\omega)) \quad (4\text{-}27)$$

Since $\omega \sim q_\theta(\omega)$ in the model, the ELBO of VI can be reformulated as:

$$\hat{\mathcal{L}}_{VI}(\omega,\theta) = -\frac{N}{M}\sum_{i\in S} logp\big(y|f(x;\hat{\omega},\theta)\big) + KL(q_\theta(\omega)||p(\omega)) \quad (4\text{-}28)$$

In each parameter updates:

$$\widehat{\Delta\theta}_{mini-batch} = \frac{\partial \mathcal{L}_{mini-batch}(\theta)}{\partial\theta} = -\frac{N}{M}\sum_{i\in S}\frac{\partial}{\partial\theta} logp\big(y|f(x;\hat{\omega},\theta)\big) \quad (4\text{-}29)$$

Compared to parameter update in VI:

$$\widehat{\Delta\theta}_{VI} = \frac{\partial \mathcal{L}_{VI}(\theta)}{\partial\theta} = -\frac{N}{M}\sum_{i\in S}\frac{\partial}{\partial\theta} logp\big(y|f(x;\hat{\omega},\theta)\big) + \frac{\partial}{\partial\theta}KL(q_\theta(\omega)||p(\omega)) \quad (4\text{-}30)$$

The condition that the efficacy of dropout SRT for modelling the interested posterior distribution $p(\theta|X,Y)$ holds when $\frac{\partial}{\partial\theta}KL(q_\theta(\omega)||p(\omega)) = 0$, which indicates the KL divergence between the prior distribution and model distribution on random variables of the model is constant. A widely used assumption for this condition is known as KL condition [91, 93] proven by Y Gal, which constraints the expectation of $\mathbb{E}_\omega[q_\theta(\omega)]$ is a Gaussian distribution. Since the dropout SRT units are Boolean random variables that follow Bernoulli distribution that can well approximate Gaussian distribution. Therefore, the equation $\widehat{\Delta\theta}_{mini-batch} = \widehat{\Delta\theta}_{VI}$ holds for dropout SRT.

4.4.2.3 Model Uncertainty Components with PDDGN model

This section discusses: i) how to obtain probability distributions of uncertainty components through network feed-forward sampling, ii) proves the feed-forward sampling can correctly estimate the probability distribution of uncertainty components of interest.

Considering a new input x^* and parameter realization $\hat{\theta}$, the corresponding output realization is $\hat{y} = f(x^*,\hat{\theta})$. The method to obtain uncertainty distribution from the model is to perform feed-forward sampling to the model. Given sampling size S, the collection set of output realizations can be denoted as:

$$\{\hat{y}_1,\hat{y}_2,\dots,\hat{y}_S\} = \{f\big(x^*,\hat{\theta}_1\big), f\big(x^*,\hat{\theta}_2\big),\dots,f(x^*,\hat{\theta}_S)\} \quad (4\text{-}31)$$

This thesis is interested in the posterior distribution $p(y^*|x^*;X,Y)$, which can also be regarded as the likelihood distribution. In particular, it indicates the model are most likely to generate datasets X,Y. Given sufficient sampling size, we need to prove the targeted posterior distribution is equivalent to the distribution estimated by feed-forward sampling. In other words, distribution given by feed-forward sampling can match the first two moment of $p(y^*|x^*;X,Y)$.

According to the outcomes given by variational inference at training stage, the variational distribution $q^*(\theta)$ that implied by the model is approaching the posterior distribution of parameters $p(\theta|X,Y)$. Therefore, we can manipulate the interested posterior distribution $p(y^*|x^*;X,Y)$ with formulation (4-31).

$$p(y^*|x^*;X,Y) = \int p(y^*|x^*;\theta)p(\theta|X,Y)d\theta \approx \int p(y^*|x^*;\theta)q^*(\theta)d\theta \qquad (4\text{-}32)$$

Therefore, two Lemmas are given henceforth, the proofs to these Lemmas are given in Appendix B:

Lemma I: Given that $f(x^*;\theta)\sim p(y^*|x^*;\theta)$, the first moment $\mathbb{E}_{p(y^*|x^*)}[y^*]$ can be estimated with unbiased estan imator $\frac{1}{S}\sum_{s=1}^{S} f(x^*;\hat{\theta}_s)$.

Lemma II: Given that $f(x^*;\theta)\sim p(y^*|x^*;\theta)$ and $p(y^*|x^*;\theta) = \mathcal{N}(y^*: \mathbb{E}_\theta[f(x^*;\theta)], \sigma I)$, the second moment $\mathbb{E}_{p(y^*|x^*)}[(y^*)^T y^*]$ can be estimated with an unbiased estimator $\frac{1}{S}\sum_{s=an\ 1}^{S}\left[f\left(x^*;\hat{\theta}_s\right)\right]^T f\left(x^*;\hat{\theta}_s\right) + \sigma^2 I$.

The proofs of the above two lemmas are given in the appendix B. Above derivation proves that the interested posterior distribution of output uncertainty $p(y^*|x^*;X,Y)$ can be estimated by making sampling to the model output, given new input x^* and stochastic realization of model parameters θ.

In addition, to model one of the uncertainty components caused by a given input factor as uncertainty source, e.g., $x_k^* \subset x^*$, it is alternatively to formulate the posterior distribution given the interested input factor $p(y^*|x_k^*;X,Y)$. The sampling inputs will be x_k^* as deterministic part of input and the realization of the rest part of the input vector $x^* - x_k^*$ sampled from given dataset X. The rationale is sampling from dataset can

maintains the prior distribution $p(x^* - x_k^*)$ to achieve unbiased estimate to the uncertainty components.

4.5 Demonstration and Results

4.5.1 Experiment Settings

In this thesis, the proposed PDDGN model for uncertainty component learning is demonstrated on realistic smart metering data from Low Carbon London (LCL) project [101]. In details, this dataset recorded over 5000 local domestic households in London, UK. The time duration covers from November 2011 to February 2014. Notably, in order to naturally characterise the uncertainty of domestic loads, impacts from external interventions are excluded in the data selection stage. In details, customers considered in the demonstration are purely unrestricted domestic households (defined by ELEXON [114]) that without any Time-of-use (TOU) tariffs as incentives and DSR as interventions. However, this thesis investigates the natural demand readings from unrestricted domestic households since the impact of price signals and DSR interventions are not focused in this thesis.

In addition to demand readings, several factors are considered as the uncertainty sources to be investigated in this thesis, include temperature, wind speed, humidity, rainfall, days of week, months, holidays and hours of the day. Besides low carbon London dataset for electricity demand, a weather dataset from third-party weather station is deployed, i.e., Hampstead weather station, London, UK [102]. The holiday information related to UK bank holidays are referenced on the website of UK government.

In the training process, mini-batch Monte Carlo optimization strategy is deployed as the training framework. In each epoch, the optimization is done by Adaptive Moment Estimation algorithm (AdamOptimizer) [103], which is easy to tune and was witnessed with superior performance under higher noise in practice. The sampling is performed across all considered customers together, the detailed parameters and techniques used in training process is summarised in following **Table 4-1**. Notably, these model hyperparameters are found through trial and error method under a set of possible parameters.

Table 4-1 Parameters Settings in Training Stage

Battery parameters	***Unit***
Net Layer Number	4 (Layers)
Epochs Number	50000
Early Stopping Threshold	1.0E-4
Batch Size	672
Input Dimension	48
Output Dimension	1
Hidden Size Each Layer	{128,128,128,64}
Dropout Rate	0.83
Optimization Algorithm	AdamOptimizer (Adaptive moment estimation)
Learning Rate	1.0E-3
Adam Hyper-Parameter Beta1	0.8
Adam Hyper-Parameter Beta2	0.7

In the uncertainty visualization stage after training process done, the model is used in a feed-forward mode to generate realizations of model output given inputs sampled from the original dataset through sampling techniques to maintain the natural prior distribution of model inputs. In details, Gibbs Sampling is deployed to perform efficient sampling across high-dimensional input space (48 dimensions). The related parameters are shown in **Table 4-2**.

Table 4-2 Parameters Settings in visualization Stage

Battery parameters	***Unit***
Sampling methods	Gibbs Sampling
Sampling Size of Realizations	30000*14*48
Number of Bins in Contour Plot (Demand Dimension)	30

In order to visualize the result of uncertainty modelling intuitively, Contour plots are presented in the demonstration of the most factors. For instance, to show the impact of temperature on electricity demand, two-dimensional contour plot and surface plot are presented with the temperature at the x axis and electricity demand at the y axis.

Accordingly, the z value is the probability distribution normalized over y axis. In other words, for any specific temperature, the integral of this probability is accumulating to one across all temperatures. The rationale for this normalization is given by formulation (4-31) that the uncertainty component caused by temperature is the marginal integral of the conditional probability distribution over all the dimensions except for temperature factor.

4.5.2 Results and Demonstration

In order to demonstrate the efficacy and validity of proposed PDDGN deep learning model in modelling uncertainty components, this section demonstrates the result implemented on Low Carbon London household electricity consumption dataset. The result consists of two parts of content: i) uncertainty component modelling results that visualize contributions from each external factor to the demand uncertainty, in this part, the uncertainty impact caused by two temporal factors (hours of the day, months) and two weather factors (temperature, humidity) are visualized by Contour plots; ii) presents the performance evaluation of proposed method by evaluating the reconstruction accuracy in the procedure of uncertainty propagation. The reconstruction accuracy is assessed on two measure, Pinball Loss function [104, 105, 106]. The pinball loss L can be formulated as:

$$L(y,o) = \begin{cases} \tau(y-o) & if\ y > o \\ (1-\tau)(o-y) & if\ o > y \end{cases} \quad (4\text{-}33)$$

4.5.2.1 To visualize Uncertainty Epistemic Components

By modelling the relationship between uncertainty and observable uncertainty components as the model inputs, this part will demonstrate how temperature, humidity, time of day and months make contributions to the load uncertainty. The modelling results are illustrated in Contour plots from Figure 4-3 to Figure 4-6.

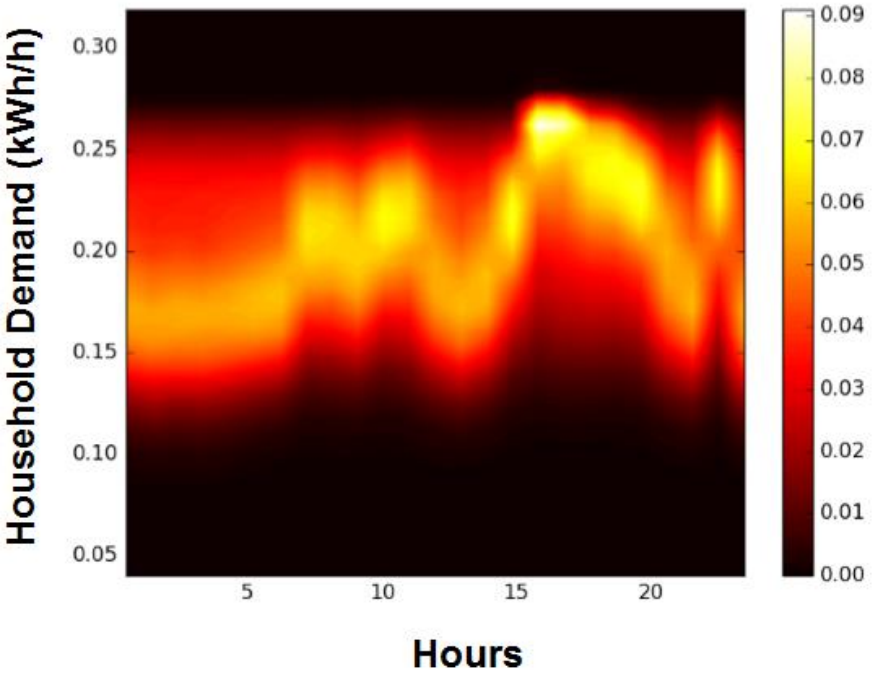


Figure 4-3 Visualization on Probability Distribution of Uncertainty Component Driven by Time of a Day

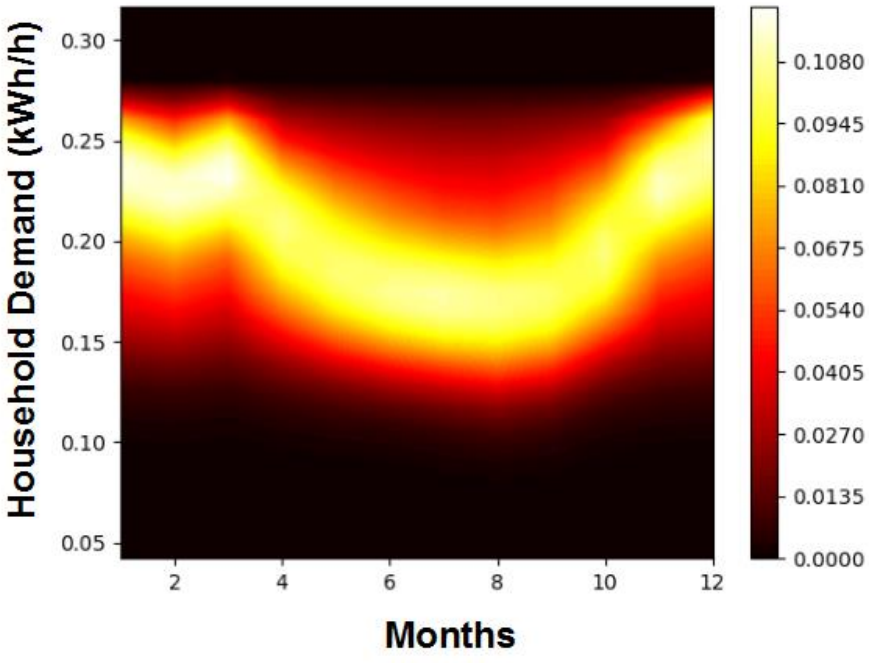


Figure 4-4 Visualization on Probability Distribution of Uncertainty Component Driven by different Months

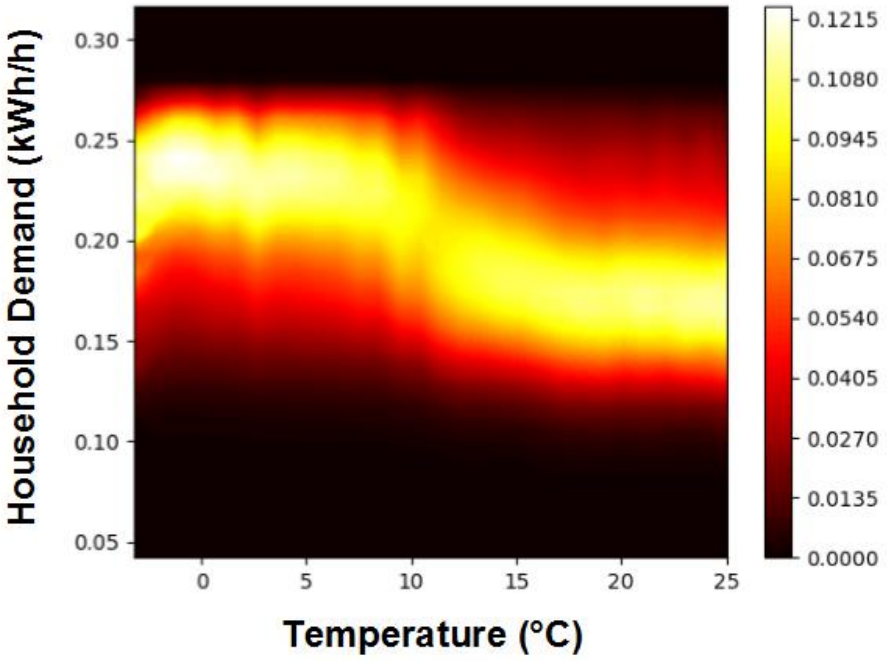


Figure 4-5 Visualization on Probability Distribution of Uncertainty Component Driven by Temperature

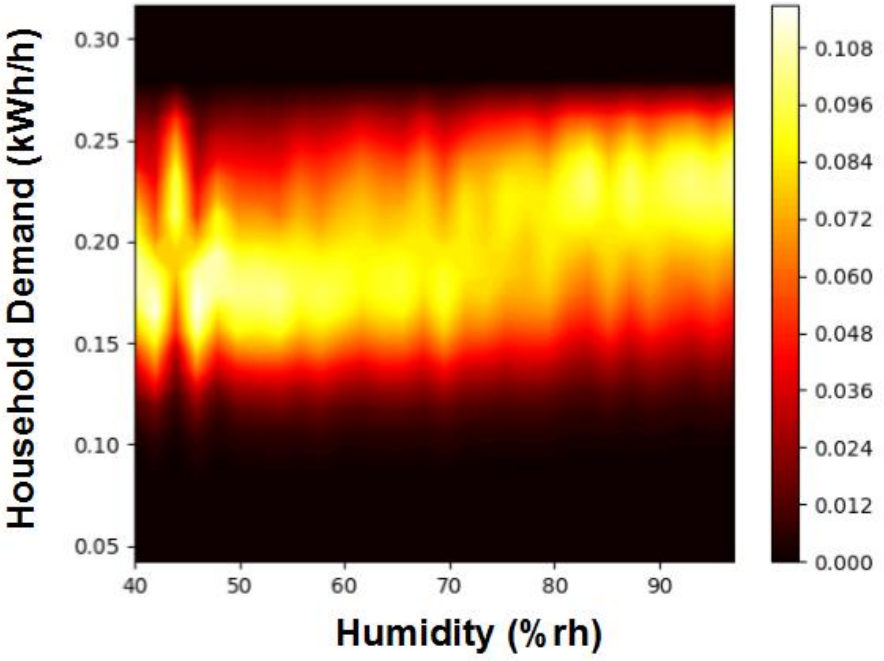


Figure 4-6 Visualization on Probability Distribution of Uncertainty Component Driven by Humidity

In **Figure 4-3**, the visualization result of probability change along with the time of the day generally follows the pattern of typical load profiles in the UK nationwide. It has two peaks per day, one is a lower peak during the noontime, when some residents start cooking and washing during the time. In the evening, a higher peak emerges after working hours, which is associated with the timing when residents come back home from work. Regarding the uncertainty, it has longer upper tails in the demand valley, and longer lower tails are associated with demand peak.

In **Figure 4-4**, the seasonal patterns of load uncertainty generally follow the pattern of seasonal load changes, where peaks are witnessed in Winter seasons, and valleys in Summer. Notably, a reverse change emerges between February and March, when load uncertainty has an unprecedented increase from February to March. This phenomenon is expected to be correlated to customer behaviours: in February, the temperatures are much lower than March, therefore, the usage of electric heaters in domestics is stable and certain, whilst in March, some customers started to shorten the usage of electric heaters.

Figure 4-5 & **Figure 4-6** demonstrates the impacts of temperature and humidity. Comparably, a higher temperature will result in lower load uncertainty, and higher humidity will result in higher load uncertainty. Associated with the needs from customers, lower temperature tends to result in various usage of electric heaters, whilst higher temperature at most result in the usage increase of electric fans (the load is more stable and continuous) [107, 108]. In terms of humidity, rainy days will increase the usage of air-dryer, washing machine and lights, hence increase the load uncertainty. An interesting phenomenon can be found from the results that both for temperature and humidity, a threshold (highlighted in black line) can be found that segments high uncertainty load and low uncertainty load. This threshold in temperature and humidity can be also the threshold that triggers the customer behaviour changes. Interestingly, a significant spike and intermittency are witnessed between 40%rh to 50%rh of humidity in **Figure 4-6**. This phenomenon may reflect the specific result in the UK. In details, lower humidity corresponds to sunny days, whilst most days in UK are rainy at a year basis. This causes insufficient sampling size at low humidity.

4.5.2.1 To Evaluate Uncertainty Components Propagation

In order to validate the proposed uncertainty model, an evaluation is performed by reconstructing the uncertainty from the four uncertainty components associated with the time of day, months, temperature, and humidity. The reconstruction result of uncertainty propagation is compared with Multiple Linear Quantile Regression and measured with Pinball Loss.

The reconstruction performance is presented in Table 4-2 and compared with MLQR in quantile loss. As shown in the results, using two features to modelling the uncertainty, the reconstruction accuracy of proposed method will be improved by 7.2% to MLQR. When using more features, the proposed method will become more accurate than MLQR. With all the four features used in the demonstration, the accuracy of uncertainty reconstruction will increase 10.8% by using proposed deep learning model. This result indicates the proposed method can better model the uncertainty components by observing the uncertainty sources. While the linear assumption of MLQR deteriorates when considering more uncertainty sources, the proposed method can more efficiently model and simulate the mutual influence and non-linearity between multiple uncertainty sources.

Notably, Results presented in Table 4-2 are not all the features considered in this experiment. Overall, 8 external uncertainty sources in total are considered and tested, i.e., Time of day, days of week, Months, Temperature, Humidity, Rainfall, Windspeed, Holidays. However, constrained by the available space for demonstration experiment results corresponding to arbitrary source composition, Table 4-2 mainly presents the composition of sources that are most impactable to the load uncertainty. In other words, Temperature, Humidity, Time of day, Months are the 4 most significant external uncertainty sources of load uncertainty. Beyond the presented results in Table 4-2, the performances are consistently improved by adding more sources as features.

TABLE 4-2 Average Pinball Loss of Uncertainty Reconstruction with Different Input Features: i) MLQR method; ii) proposed PDDGN method

Features used in the reconstruction		*MLQR method*	*Proposed PDDGN method*	*Performance Improvement (%)*
2 Features	***Temperature + Time of day***	36.218	33.491	**7.185**
	Temperature + Humidity	92.291	85.929	
	Temperature + Months	90.886	84.578	
	Humidity + Time of day	37.335	34.665	
	Humidity + Months	95.284	88.982	
	Time of day + Months	34.217	31.487	
3 Features	***Humidity + Time of day + Months***	29.684	26.809	**8.981**
	Temperature + Time of day + Months	28.852	26.294	
	Temperature + Humidity + Months	79.128	72.234	
	Temperature + Humidity + Time of day	34.248	31.282	
4 Features	***Temperature + Humidity + Time of day + Months***	26.724	23.828	**10.837**

4.6 Chapter Summary

This chapter develops an advanced deep learning model to achieve purely data-driven inverse uncertainty quantification. The proposed model is demonstrated on smart metering data of 5000 households in London, UK, in order to model the uncertainty components lies in the load uncertainty and identify its correlated causing sources.

The result indicates a strong correlation between load uncertainty and external factors, such as temperature, humidity, time of day, and months. In general, temporal features make a great impact on load uncertainty, while the temperature is more significant than humidity.

Through performance comparison in uncertainty reconstruction test, the proposed deep learning model can better model the non-linearity and mutual influence between uncertainty sources. The proposed method outperforms linear method MLQR, in different input features dimensionalities. Notably, the performance improvement from a linear model to the proposed deep learning model keep increased with more input features. This phenomenon indicates the proposed method is effective to model the non-linearity and mutual influence.

Chapter 5

Uncertainty Model Applied in Energy Management System

THIS chapter demonstrates an application of previous UQ methods and findings for improving home Energy Management System.

5.1 Introduction

In previous chapter, this thesis developed advanced methodologies for both forward and inverse uncertainty quantification. In addition, many new learnings and information are discovered in the demonstration, on the temporal features, components composition, uncertainty sources. These new learnings will reveal two essential problems: how load uncertainty generated from its uncertainty sources, i.e., weather conditions, temporal features and social information, and how load uncertainty at domestic level propagates to the grids through system aggregation.

This chapter will introduce an application of obtained new learnings of load uncertainty, which aims at improving Energy Management System. Compared to previous works in EMS that attempt to find optimal operation solution under uncertainty, with techniques such as Stochastic Programming, Robust Optimization, etc., the developed EMS explores from a different angle, to achieve grid level goals from decentralised home EMS based on the properties in uncertainty propagation procedure. Traditionally, to achieve system level goal, the centralised control strategy for EMS are required to coordinates the resources of households. However, centralized approaches require additional investments in Information & Communication Techniques (ICTs) and pose a huge computational burden to the central system. This section attempts to achieve centralised performance with the decentralised system without excessive computational burdens and coordination communications.

The rest content of this chapter is organized as follows: Section 5.2 proposes the strategy to improve EMS with learnings in uncertainty propagation. Section 5.3 discusses the experiment and parameter settings, the results and demonstration are presented in section 5.4.

The content of this chapter is cited from the author's manuscript to be submitted to IEEE Transactions on Smart Grid, titled as 'Decentralised Home Energy Management System to Reduce System Peak and Uncertainty by Wavelet Auto-decoupling Optimization'.

5.2 Proposed Methodology

The Introduction of preliminary technique 'Wavelet Transforms' for this chapter is introduced in Appendix A. In this thesis, the household EMS system is enabled by a battery storage that can be charged and discharged at any time of a day within its feasible ranges of State of Charge (SoC). For mathematical convenience, the demand, renewable generation profiles, charging/discharging efficiency and Time-of-use (TOU) price are regarded as deterministic. The EMS is operated in a day-ahead scheme where the battery operations (charging and discharging) are determined by proposed WARP optimization. It is notable that the predicted customer demand is assumed to be accurate for simulation convenience. The EMS performance is thus evaluated simply based on the historical demand of smart metering data.

As shown in the flowchart, the proposed EMS attempts to achieve global performance without coordination signals from households and aggregators. In detail, it achieves the goal by integrating proposed wavelet auto-decoupling regularization term into a two-objective optimization model: i) the primary objective is to minimize household energy cost by shift energy from high price to low price period; ii) the

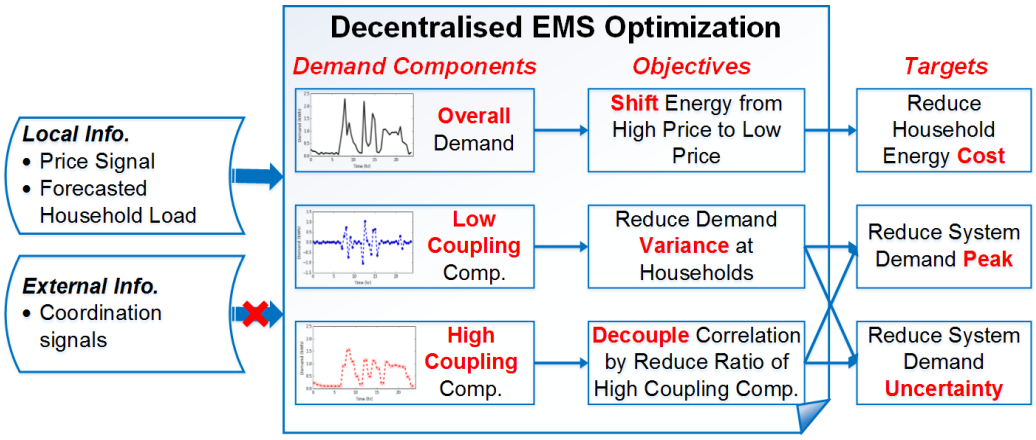


Figure 5-1 Flowchart of Decentralised home EMS Operation Strategy

secondary objective is to reduce peak and uncertainty at household level by reducing demand variances at households; iii) and the wavelet auto-decoupling term is integrated to decoupling correlations of household groups in the same service area by reducing the proportion of high coupling demand components in each household, in order to reduce system level peak and uncertainty through propagation from households to the system.

5.2.1 Multi-objective Optimization Model for Household EMS Operation

The programming model considers optimizing the battery operations of a single household within the optimization horizon. The optimization horizon is set to be a day duration that is evenly divided into M time intervals. The length of each time interval is represented by T. In this thesis, $T = 0.5(hr)$ and $M = 48$. In the programming model, time intervals are denoted by discretized time index $t \in [1,48] \cap \mathbb{Z}$.

Table 5-1 NOMENCLATURE

Notations		*Description*
Constant parameters	T	Time intervals for each operation, in hr
	M	A number of time intervals in the programming horizon.
	$\mathbb{Z}^+$	Set of natural numbers
	P_h	Thermal limit of the household circuit, as expressed as maximum household power, in kWh/h.
	P_{dch}	Maximum battery discharging rate, in kWh/h.
	P_{ch}	Maximum battery charging rate, in kWh/h.
	η_{ch}	The battery charging efficiency factor
	η_{dch}	Battery discharging efficiency factor
	E_{ini}	The initial energy of battery storage, in kWh.
	E_{max}	Maximum energy limit of battery storage, in kWh.
	E_{min}	Minimum energy limit of battery storage, in kWh.
	C_t	Time-of-use electricity tariff at time t, in $£/hr$, $\forall t$.
Variables	t	Time index
	d_t	Household load at time t, in kWh/h, $\forall t$.
	$\boldsymbol{p_{ch,t}}$	Charging power at time t, in kWh/h, $\forall t$.
	$p_{dch,t}$	Discharging power at time t, in kWh/h, $\forall t$.
	$b_{ch,t}$	Charging state boolean variable at time t, $\forall t$.
	$b_{dch,t}$	Discharging state boolean variable at time t, $\forall t$.
	e_t	Battery energy at time t, in kWh, $\forall t$.
	x_t	Net demand at time t, in kWh/h, $\forall t$.

		$x_t = d_t + p_{ch,t} - p_{dch,t}$
	$\bar{x}$	Average net demand across the programming horizon, $\bar{x} = \frac{1}{M}\sum_{t=1}^{M} x_t$
	λ	Weight factor of demand variance
	γ	Weight factor of the regularization term
Vectors	$\boldsymbol{d}$	Vector of household load, $\boldsymbol{d} = (d_1, d_2, \dots, d_M)^T$
	$\boldsymbol{p}_{ch}$	Vector of battery charging power, $\boldsymbol{p}_{ch} = (p_{ch,1}, p_{ch,2}, \dots, p_{ch,M})^T$
	$\boldsymbol{p}_{dch}$	Vector of battery discharging power, $\boldsymbol{p}_{dch} = (p_{dch,1}, p_{dch,2}, \dots, p_{dch,M})^T$
	$\boldsymbol{e}$	Vector of battery energy, $\boldsymbol{e} = (e_1, e_2, \dots, e_M)^T$
	$\boldsymbol{C}$	Vector of daily time-of-use tariff,$\boldsymbol{C} = (C_1, C_2, \dots, C_M)^T$
	$\boldsymbol{x}$	Vector of net demand $\boldsymbol{x} = (x_1, x_2, \dots, x_M)^T$
	$\boldsymbol{u}$	Unit column vector $\boldsymbol{u} = (1,1, \dots,1)^T \in \mathbb{R}^M$
	$\boldsymbol{Q}$	Projection matrix to wavelet basis, $\boldsymbol{Q} = \boldsymbol{\Psi}_k^{(j)} \boldsymbol{\Psi}_k^{(j)^T}$

5.2.1.1 Objective I: to minimize household energy bills

The first objective for the EMS optimization is to minimize the electric energy bills for the household customer under given TOU tariffs. Assuming the predicted daily demand is $\{d_t\}_{t\in[1,48]\cap\mathbb{Z}}$ and daily TOU tariffs is $\{C_t\}_{t\in[1,48]\cap\mathbb{Z}}$.Therefore, this objective can be formulated as the sum of the energy cost of each time interval within the whole optimization horizons:

$$(P1) \quad Objective\ I \coloneqq min \ \ T\sum_{t=1}^{M} C_t x_t \qquad (5\text{-}1)$$

Subjected to:

$$0 \leq p_{ch,t} \leq b_{ch,t} P_{ch}, \forall t \in [1, M] \cap \mathbb{Z} \quad (5\text{-}2)$$

$$0 \leq p_{dch,t} \leq b_{dch,t} P_{dch}, \forall t \in [1, M] \cap \mathbb{Z} \quad (5\text{-}3)$$

$$b_{ch,t} + b_{dch,t} \leq 1, \forall t \in [1, M] \cap \mathbb{Z} \quad (5\text{-}4)$$

$$b_{ch,t}, b_{dch,t} \in \{0,1\} \quad (5\text{-}5)$$

$$E_{min} \leq e_t \leq E_{max}, \forall t \in [1, M] \cap \mathbb{Z} \quad (5\text{-}6)$$

$$e_0 = e_M = E_{ini} \quad (5\text{-}7)$$

$$e_t = e_{t-1} + T \cdot (p_{ch,t} \eta_{ch} - \frac{p_{dch,t}}{\eta_{dch}}), \forall t \in [1, M] \cap \mathbb{Z} \quad (5\text{-}8)$$

$$0 \leq x_t \leq P_h, \forall t \in [1, M] \cap \mathbb{Z} \quad (5\text{-}9)$$

where the net demand of the household is defined as $x_t = d_t + p_{ch,t} - p_{dch,t}$. d_t refers to the inherent household electric demand. $p_{ch,t}$ and $p_{dch,t}$ refer to control variables of battery charging and discharging power, which alter the demand to reduce the household demand peak and uncertainty whilst reduce energy bills.

5.2.1.2 Objective II: to smooth household demand

The second objective term attempts to smooth the household demand, to reduce demand peak and uncertainty at the household level. This goal is achieved by minimizing the variance:

$$Objective\ II \coloneqq min \sum_{t=1}^{M} (x_t - \bar{x})^2 \quad (5\text{-}10)$$

Where $\bar{x}$ refers to the arithmetic mean of the household net demand (inherent demand and battery operations).

5.2.1.3 Programming model with two objectives

Considering thermal, power and capacity constraints with respect to battery storages and household circuits, the single objective programming model ($P1$) can be formulated as:

$$(P2) \quad min \ \ T\sum_{t=1}^{M} C_t\, x_t + \lambda \sum_{t=1}^{M} (x_t - \bar{x})^2 \qquad (5\text{-}11)$$

Subjected to:

(5-2)—(5-9)

The parameter λ is the weight factor [109, 110], of the second objective term compared to first objective term. This model is a quadratic convex programming problem which can be solved efficiently, especially in decentralised problem scale at the household level. The control variables $p_{ch,t}$ and $p_{dch,t}$ will indicate how to operate the battery storage within the optimization horizon, hence alter the demand to reduce the household demand peak and uncertainty whilst reduce energy bills.

5.2.2 Wavelet Auto-decoupling Regularization Term

The fundamental idea of wavelet auto-decoupling regularization (WSR) term is to alter the demand shape to de-couple the correlations of the population of household demands. This part formulates the WSR regularizer and discusses the efficacy. Considering three level of decomposition on the household demand, we can expand the demand series with Haar expansions shown in equation (5-12):

$$x_t = \sum_{k} a_k^{(j_0)}\, \phi_{k,t}^{(j_0)} + \sum_{j=1}^{j_0} \sum_{k} d_k^{(j)} \Psi_{k,t}^{(j)} \qquad (5\text{-}12)$$

It is notable that to simplify the formulation, time index t are moved to subscripts. Since the first term of Haar expansions is the approximation part that approximates the demand at decomposition level j_0. The second terms are detail components in higher decomposition levels $j > j_0$. Prior evidences [77] have indicated the higher level of decomposition, the lower diversity lies in the decomposed components. To reduce the energy of low resolution channel components, a L_2 norm regularization term is proposed and formulated for decomposition levels lower than $j_0 - 1$ as:

$$regularization\ term = \sum_{j=1}^{j_0-1} \sum_{k} \left\| d_k^{(j)} \Psi_{k,t}^{(j)} \right\|_2 \quad (5\text{-}13)$$

According to the absolutely scalable property of norm [111, 112] and normality of Haar wavelets that $\left\| \Psi_k^{(j)} \right\| = 1$, the regularization term(5-14) can be reformulated as:

$$regularization\ term = \sum_{j=1}^{j_0-1} \sum_{k} d_k^{(j)^2} = \sum_{j=1}^{j_0-1} \frac{1}{2^j} \left[\sum_{k} \left(\sum_{t=1}^{M} x_t \Psi_{k,t}^{(j)} \right)^2 \right] \quad (5\text{-}14)$$

Therefore, by adding the regularization term, the multi-objective programming model is:

$$(P3) \quad min \ \ T \sum_{t=1}^{M} C_t\, x_t + \lambda \sum_{t=1}^{M} (x_t - \bar{x})^2 + \gamma \sum_{j=1}^{j_0-1} \frac{1}{2^j} \left[\sum_{k} \left(\sum_{t=1}^{M} x_t \Psi_{k,t}^{(j)} \right)^2 \right] \quad (5\text{-}15)$$

Subjected to:

(5-2)—(5-9)

As shown in the P2 programming model, the constraints are linear while the objective is quadratic. In matrices representation, the P2 model can be rewritten into a more compact element-wise counterpart. Denoting the demand with EMS as $\boldsymbol{x} = (x_1, x_2, \dots, x_M)^T$, $\boldsymbol{u} = (1,1,\dots,1)^T \in \mathbb{R}^M$, and …

The element-wise counterpart is:

$$(P2) \quad min \ \ T\boldsymbol{c}^T\boldsymbol{x} + \lambda(\boldsymbol{x} - \bar{x}\boldsymbol{u})^T(\boldsymbol{x} - \bar{x}\boldsymbol{u}) + \gamma \sum_{j=1}^{j_0-1} \frac{[\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}]}{2^j} \quad (5\text{-}16)$$

Subjected to:

$$\boldsymbol{Q} = \boldsymbol{\Psi}_k^{(j)} \boldsymbol{\Psi}_k^{(j)^T} \quad (5\text{-}17)$$

$$\mathbf{0} \leq \boldsymbol{p}_{ch} \leq P_{ch}\boldsymbol{u} \quad (5\text{-}18)$$

$$\mathbf{0} \leq \boldsymbol{p}_{dch} \leq P_{dch}\boldsymbol{u} \quad (5\text{-}19)$$

$$E_{min}\boldsymbol{u} \leq \boldsymbol{e} \leq E_{min}\boldsymbol{u} \quad (5\text{-}20)$$

$$e_0 = e_M = E_{ini} \quad (5\text{-}21)$$

$$e_t = e_{t-1} + T \cdot (p_{ch,t}\eta_{ch} - \frac{p_{dch,t}}{\eta_{dch}}), \forall t \in [1, M] \cap \mathbb{Z} \quad (5\text{-}22)$$

$$\mathbf{0} \leq \boldsymbol{x} \leq P_h\boldsymbol{u} \quad (5\text{-}23)$$

5.2.3 Discussion of Efficacy on WSR regularization term

This part discusses the efficacy of proposed wavelet auto-decoupling regularization term, which consists of two part of the discussion: i) WSR term can increase the proportion of high decomposition components; ii) higher proportion of high decomposition components will result in less correlated household demands, hence WSR term can de-couple correlations between household demands.

The proposed WSR term attempts to minimize the L_2 norm of low and medium decomposition level of uncertain demand, which is:

$$\sum_{j=1}^{j_0-1} \sum_{k} \left\| d_k^{(j)} \Psi_{k,t}^{(j)} \right\|_2 = \sum_{j=1}^{j_0-1} \sum_{k} {d_k^{(j)}}^2 \quad (5\text{-}24)$$

Since the integral of wavelets across time horizon is constantly equal to zero, the L_2 norm of a certain level j is equivalent to the variance of the decomposed component at level j. According to Central Limit Theorem [79], we can formulates the variance of sum of random variables in Lemma I:

Lemma I: Considering N random variables $\{\xi_i\}_{i=1,2,\dots,N}$, the sum of the random variable ξ_p can be approximated by Normal the distribution that:

$$\xi_p = \sum_{i=1}^{N} \xi_i = normal\left(\sum_{i=1}^{N} \mu_i, \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_{ij}\sigma_i\sigma_j}\right) \quad (5\text{-}25)$$

The variance of the random variable sum is:

$$Var(\xi_p) = \sum_{i=1}^{N} Var(\xi_i) + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \rho_{ij}\sigma_i\sigma_j \quad (5\text{-}26)$$

where ρ_{ij} is the pair-wise PCC between variable i and j. Since wavelet components at different decomposition level have orthogonal property, that variance of demand is the sum of variances of decomposed demand components at different decompose level. Since the WSR term is designed to reduce the variance at low and medium decomposition level whilst does not directly affect high-level components. The proposed WSR thus can increase the proportion of high-level decomposed components.

In terms of electric demand, components at higher decomposition level tend to be less correlated and vice versa. The physics behind this property is:

i) There exist many highly correlated demand patterns caused by external factors, such as weather conditions, job-related activities, social events, energy prices and technologies. However, these factors usually impact the electric demand at daily basis. For instance, job-related activities tend to have the daily periodical pattern and impact electric demand on daily basis.

ii) With a high level of decomposed demand components, causes such as randomness, measurement error, etc. tend to be less correlated.

Therefore, high decomposition level of components of household demand is less correlated and vice versa. The proposed WSR term thus can de-couple the correlations between household demands.

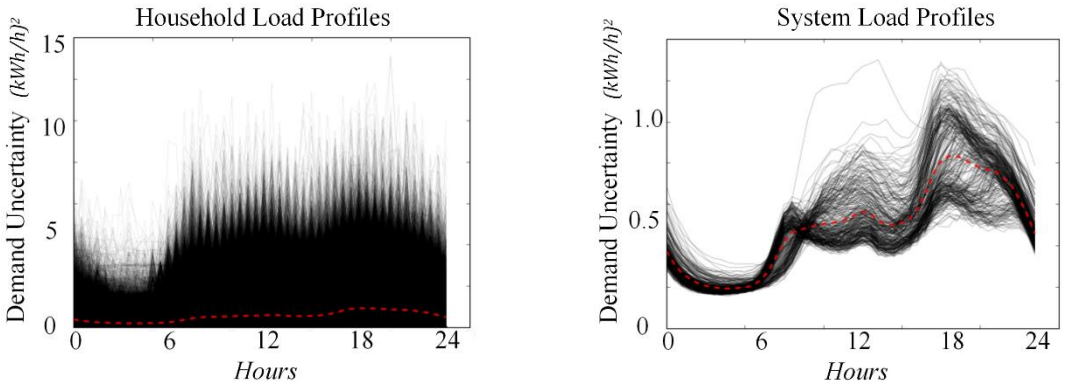
In order to validate the proposed WARP strategy, two benchmark programming strategies are encountered in this thesis, i.e., classical EMS programming model that solely considers energy bill reduction as the problem $(P1)$, and a multi-objective

programming model that minimizes energy bill and demand variance as the problem ($P2$).

5.3 Experiment Settings

5.3.1 Data Description

In this thesis, the proposed WARP strategy is implemented on realistic smart metering data recorded from over 5000 Irish domestic households and small/medium enterprises (SMEs). The database is published by the Smart Metering Electricity Customer Behaviour Trials (CBTs) initiated by Commission for Energy Regulation (CER). The time duration of this database dates back to 1st July 2009 – 31st December 2010. Specifically, 929 1-E-E consumers have considered in this thesis, who represents residential (1) customers with the controlled stimulus (E) and controlled tariff (E). The rationale for employs these group of customers is that they are billed on flat rate tariff without any DSR interventions exist, which can naturally reflect the actual household demand without biases. The data of both aggregated and disaggregated demand are presented in **Figure 5-2**:



a) b)

Figure 5-2 Demands and quantiles at aggregated/disaggregated level: a) demand of residential households, b) demand of distribution network

In order to investigate the performance of proposed EMS strategy comprehensively, different aggregation scales of customer group are encountered in the demonstration, ranging from 50 to 900. In addition, different levels of penetrations are also considered, i.e. low, medium and high penetration rate (30%, 50% and 80%).

In the household with EMS employed, an in-home battery is installed to enable the EMS system. We first assume all the EMS system employs same type of battery storage with 5 (kWh) capacity and 1.5 (kW) maximum charging/discharging rate. The initial SOCs of the battery storage is generated by a normal distribution $\mathbb{N}(0.5, 0.1^2)$. For simplicity, the charging and discharging efficiency are all set to be 0.93. The maximum and minimum limit of battery SOC is set at 90% and 10% [113]. According to the periodical similarity of different dates, the final SOC level ($T = 48$) is equal to the corresponding initial SOC as well. The parameters setting details are presented in **Table 5-2** below:

Table 5-2 BATTERY PARAMETERS

Battery parameters	***Unit***
Capacity	5 kWh
Charging Limit	1.5 kW
Discharging Limit	1.5 kW
Max SOC	0.9
Min SOC	0.1
Initial SOC	$N(0.5, 0.1^2)$
Charging Efficiency	0.93
Household Circuit Limit	5 kW
Discharging Efficiency	0.93

In the demonstration, a set of typical UK TOU tariff is considered, which consists of several price levels at the different time of the day. This TOU tariff is invented by ELEXON in the 1990s which consists of multiple tariffs corresponding to different seasons and types of customers (Economic 7 and Non-economic 7). The Economic 7 customers employ electric heater that is working hard at night, their tariffs are hence largely different to the normal customers. This thesis employs a TOU tariff to incentivize the EMS operations. The TOU tariff was proposed in work [114] which were derived from the wholesale price of UK. A set of different tariffs for different seasons is adopted as the economic signals to alter electricity consumption behaviour. The four seasonal tariffs are shown in **Figure 5-3**.

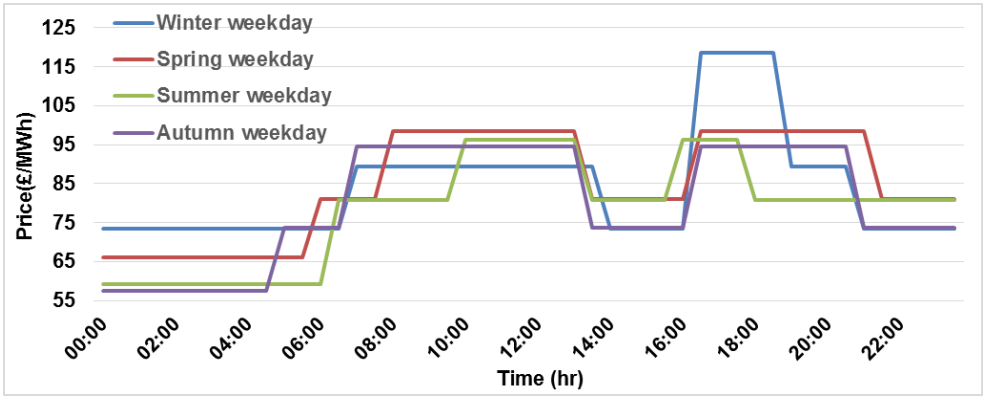


Figure 5-3 TOU tariffs employed in differing seasons

5.3.2 Benchmark Measures

In this thesis, the performance of EMS system for the purpose of a network supporting consists of three parts: i) average energy bill savings across individual households; ii) Annually and daily network peak reduction; and iii) system demand uncertainty in standard deviation (SD). To demonstrate the value of system peak reduction, network investment deferral is also employed in the evaluation. Both the saving of energy bill and network investment deferral is evaluated in UK pounds and percentages of bill reduction.

In terms of the network peak reduction, both annually and average daily peak are considered and examined in the demonstration. Daily network peak reduction can evaluate the capability of reducing demand peak by shifting demand from system peak period to the rest of periods. However, in the system infrastructure planning in UK industry, the investment is determined by the maximum demand of the whole year [115], which is normally on a particular day in the winter. Therefore, the reduction of annual system peak is also considered and translated into network investment deferral.

For a period of L days and N residential households, the propagated system demand can be denoted as:

$$D_{l,t} = \sum_{n=1}^{N} d_{l,t,n} \ , \forall l \in [1, L] \cap \mathbb{Z}, \forall n \in [1, N] \cap \mathbb{Z}, \qquad \forall t \in [1,48] \cap \mathbb{Z} \tag{5-27}$$

where sub-notion t represents the time of the period, n denotes the customer index, and l refers to the date index. The typical load profile (TLPs) [116] of propagated network demand is hence calculated as the arithmetic mean across days, i.e., $E[D_{l,t}] = \frac{\sum_{l=1}^{L} D_{l,t}}{L}, t \in [1,48] \cap \mathbb{Z}$. The uncertainty of network demand can be represented as the standard deviation (σ) with respect to the arithmetic mean of network demand.

$$\sigma(D) = E[(D - E[D])^2] = \sqrt{\frac{\sum_{t=1}^{T}\sum_{l=1}^{L}(D_{l,t} - E[D_{l,t}])^2}{LT}} \quad (5\text{-}28)$$

5.4 Demonstration and Results

In order to validate the performance of proposed WARP strategy, we compare the proposed strategy with the two benchmark strategies. The demonstration composes of: i) validation of EMS effect on de-coupling the correlations of household demand; ii) EMS performance under a certain scenario with 50% EMS penetration rate and aggregation scale of 100 customers; iii) EMS performance comparison across different aggregation scales, EMS penetration rate, and weight factors λ (objective to smooth demand), γ (wavelet auto-decoupling regularization term). It is notable that both weight factors λ, γ are set to constant 500 in the first two parts.

5.4.1 Validation of EMS De-coupling Effect on Household Demand Correlations

In order to validate the de-coupling effect of proposed wavelet auto-decoupling regularization term, the demand correlations of all investigated households are demonstrated as follows. Four scenarios are compared in **Table 5-3**: i) original household demand with EMS engagement, ii) new household demand with EMS impact operated by min-cost strategy, iii) new household demand with EMS impact operated by min-variance strategy, and iv) new household demand with EMS impact operated by proposed WARP strategy.

The demand correlations are measured by Pearson product-moment correlation coefficients (PCCs) [78], three statistics, includes expectation, 25% quantile and 75% quantile is presented in **Table 5-3** across 900 investigated residential households.

The result indicates, both min-cost and min-variance strategy of EMS operation will largely increase the correlations of household demand. Compared to original demand without EMS effect, household demand correlations are averagely increased by 45%/57% with min-cost/min-variance strategy. This result indicates classical EMS operation strategy which shifts energy from high price periods to low price periods will cause similar patterns in the demand shape, hence largely increase the demand correlations.

Table 5-3 STATISTICS (EXPECTATION, 25% QUANTILE & 75% QUANTILE) OF DEMAND CORRELATIONS IN PCCS ACROSS 900 RESIDENTIAL HOUSEHOLDS

Time scales	*High level component proportion*	*Average Pair-wise PCCs*	*25% Quantile of Pair-wise PCCs*	*75% Quantile of Pair-wise PCCs*
Without EMS	34.0%	0.150	0.092	0.207
P1 model	21.2%	0.217	0.159	0.284
P2 model	20.4%	0.235	0.155	0.319
WAP model	31.7%	0.166	0.094	0.242

With proposed strategy, the proportion of high decomposition level components will increase around 50% from benchmark strategy, and the correlation will decrease by 24%/30% from min-cost/min-variance strategy. In other words, average demand correlation will only increase by 11% from original household demands, which equals to nearly 75%-80% of the correlation increase brought by classical EMS strategies. This result indicates the proposed wavelet auto-decoupling regularization term can eliminate most of the correlation increase brought by classical EMS strategies.

5.4.2 Performance Evaluation Case: 50% Penetration Rate with 100 Households

In this section, the proposed WARP strategy is firstly evaluated on a certain scenario with 50% penetration rate of residential EMS systems, and the network aggregated on 100 households. The data of 100 households are randomly sampled from the Irish database. The investigated period ranges from 14th Sep 2009 – 14nd May 2010, which contains least missing data and unknown data during the whole period in the database.

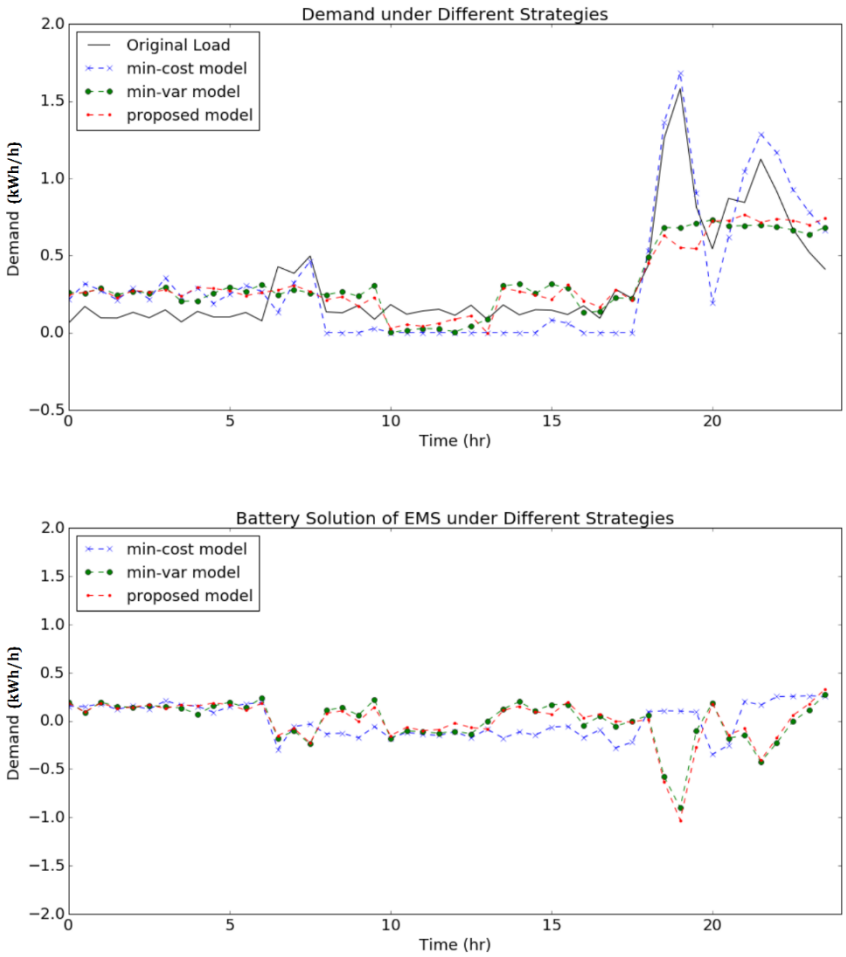


Figure 5-4 Household demand and battery solution with different EMS strategy

Figure 5-4 presents an example of the original residential demand and the new demand scheduled by household EMS system with different operational strategies. According to the result, both min-variance and proposed WARP strategy will not only shift energy from high price periods to the low price but also seek for solutions to decrease the gap between peak and valley. The solution generated by proposed strategy is presented in red dashed line with dot marker, which also has a high-frequency pattern (ups and downs in a short period).

Figure 5-5 shows the system daily load profile changes before and after different EMS strategy engaged in the scenario of 100 households on 12th April 2010. The demonstrated households and date are randomly selected to present how EMS take direct effects on household demand and how different between solutions given by proposed EMS strategy and benchmark strategies. Accordingly, network demand after EMS engaged is presented in quantile **Figure 5-5**. The result indicates when the

household EMS effect propagates to the system, the proposed WARP strategy outperforms the benchmark strategies and witness less amount of uncertainty and lower peak.

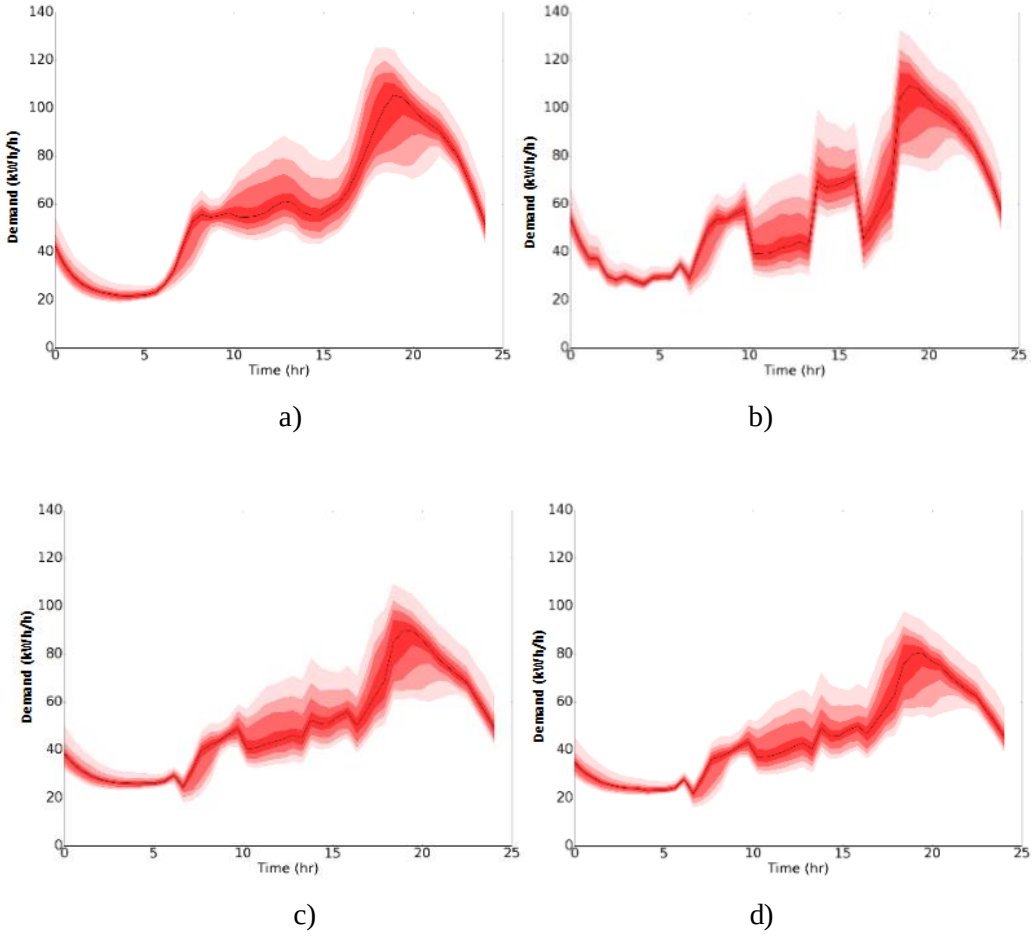


Figure 5-5 Quantile plot of distribution network with different EMS strategies: a) without EMS; b) benchmark P1; c) benchmark P2; d) WARP strategy

Table 5-4 summarizes the result of performance comparison between three EMS strategies. The performance is evaluated by energy bill saving, system peak reduction and network demand uncertainty. The result in **Figure 5-5** implies that: i) min-cost strategy seeks for solution to achieve optimal energy bill savings for individual households (9.59% energy bill saving), while causes 6.9% annual peak increase and 2.44% uncertainty increase; ii) min-variance strategy can achieve peak reduction (12.64% daily and 19.11% annually reduction) and uncertainty reduction (34.83%) at system level with a slight compromise to energy cost (1.56%); iii) the proposed WARP strategy outperforms two benchmark strategy both in peak reduction (18.01% daily and 31.60

annually reduction) and uncertainty reduction (45.28%), with similar energy cost compromise (1.7%).

Table 5-4 EMS Performance Summary on Scenario with 100 Aggregation Scale and 50% EMS Penetration Rate

strategyies	*Original Demand*	*P1 Model*	*P2 Model*	*WAP Strategy*
Annual Electricity Bills $(£/yr)$	708.22	640.27	651.35	652.33
Annual Bill Saving Increase $(£ \cdot yr^{-1}/\%)$	–	67.95 / 9.6%	56.87 / 8.0%	55.89 / 7.9%
Annually Network Demand Peak (kWh)	143.05	152.91	115.71	97.84
Daily Demand Peak Expectation (kWh)	97.12	102.80	84.85	79.63
Daily Demand Peak Standard Deviation (kWh)	16.04	18.76	9.39	7.01
Demand Peak Reduce Rate Annually/Daily (%)	–	−6.9% /−5.8%	19.1% /12.6%	31.6% / 18.0%
Network Demand Uncertainty in STD (kWh)	11.63	11.91	7.76	6.36
Reduction Rate of Uncertainty (%)	–	−2.4%	34.8%	45.3%

5.4.3 Performance Comparison under Multiple Scenarios

In order to provide a comprehensive view of the performance, capability and limitation of the proposed WARP strategy, scenarios with different aggregation scales and EMS penetration rates are investigated. The detail results are presented in **Table 5-4**.

As shown in **Figure 5-6**, benchmark P1 (min-cost strategy) which designed to minimize the energy cost, will have an adverse impact on the system regarding system peak and uncertainty, i.e., increase the system peak and uncertainty, even though the adverse impact is less than 10%. Benchmark P2 (min-variance strategy) can bring peak and uncertainty reduction to the system. In detail, it can damp down 10% to 15% of system peak in different EMS penetration rate from 30% to 80%. In terms of

uncertainty reduction, the min-variance strategy can reduce nearly 35% of uncertainty across different penetration rate.

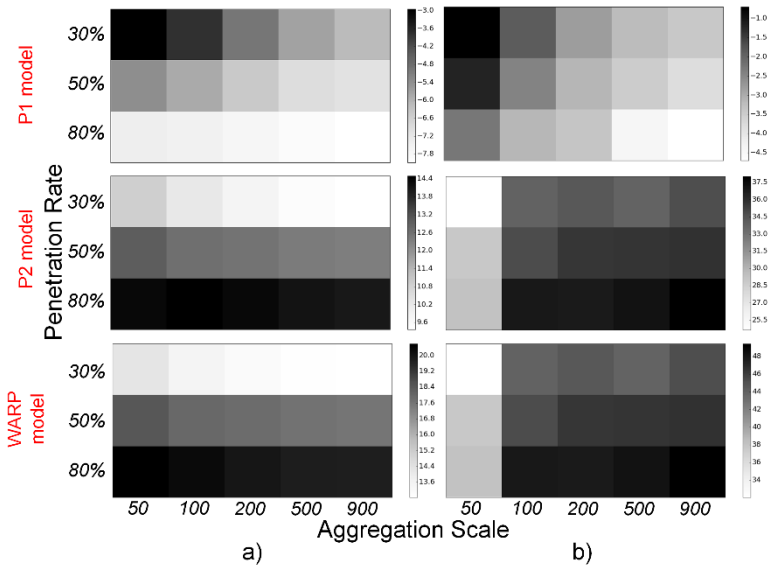


Figure 5-6 EMS performances on a) peak reduction & b) uncertainty reduction across different penetration rate and aggregation scales

The proposed WARP strategy significantly outperforms the two benchmark strategies: i) system peak reduction effect ranges from 13% to 20%, which has nearly 30% of performance improvement from benchmark P2; ii) uncertainty reduction effect ranges from 34% to 49%, which is equivalent to 20% of performance improvement from benchmark P2.

It is notable that inadequate aggregation scale will largely deteriorate the effect on uncertainty reduction, however, it does not have much impact on peak reduction effect. This is because the aggregation of demand at a small aggregation scale will lack smoothness, which may affect the uncertainty reduction through propagation but not affect the peak reduction. **Figure 5-7** further demonstrates the impact of weight factors λ, γ.

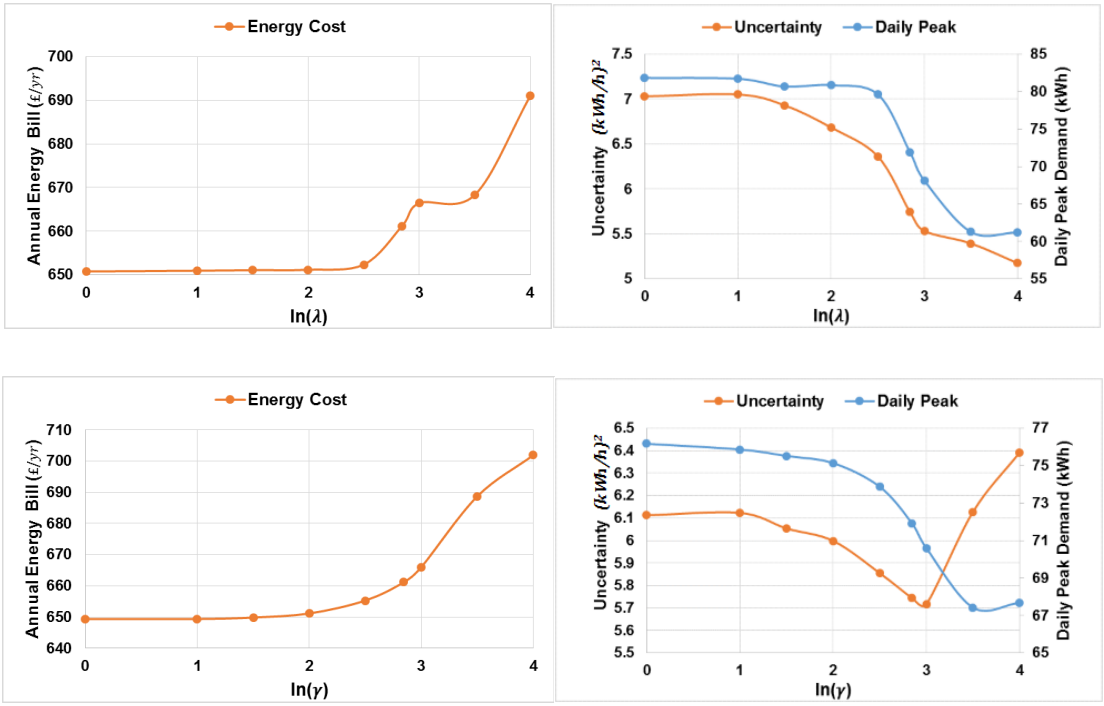


Figure 5-7 Performance of WARP strategy across different weight factors

5.5 Chapter Summary

This thesis for the first time explores the functionality of household EMS system with battery storage to support network uncertainty management in a decentralised fashion. A novel Wavelet Auto-decoupling Regularization Programming (WARP) model is proposed as the control strategy of the EMS operations. The proposed strategy is compared with two benchmark strategies min-cost and min-variance strategies. In detail: i) The classical EMS strategy to minimize energy cost; ii) A multi-objective EMS strategy to minimize the demand variance whilst reduce energy cost.

Performances are validated on four aspects: i) demand correlation, ii) energy cost saving per households, iii) peak reduction on a system level, and iv) uncertainty reduction on a system level. The results indicate:

i) Benchmark strategies will increase the correlations of household demands by 45%/57%, while proposed WARP strategy can de-couple the correlation to alleviate nearly 75% to 80% of correlation increases.

ii) In the demonstrated scenario of 100 households and 50% EMS penetration rate, the proposed EMS achieves 18% of daily peak reduction and 45% peak of uncertainty reduction at a slight compromise on cost reduction (8%). Critically,

the reduction in the annual peak is over 30%, which can contribute to a significant investment deferral.

According to the comparison across scenarios, higher EMS penetration rate and larger aggregation scale can generally lead to larger peak and uncertainty reductions. The only exception is that small aggregation level can guarantee more peak reduction.

In conclusion, the demonstration application witnessed grid level peak and uncertainty reduction with decentralised way. This result validates the EMS can be improved by using the new findings and information from uncertainty research that: uncertainty components can mitigate through uncertainty propagation, and the degree of mitigation is determined by the diversity between uncertainty components.

Chapter 6

Uncertainty Model Applied in Point Load Forecasting

THIS chapter introduces another application of uncertainty modelling, in household point load forecasting. This chapter for the first time directly models epistemic uncertainty components shared by neighbourhoods, to improve forecasting accuracy.

6.1 Introduction

Last chapter has demonstrated the application of uncertainty modelling in the Energy Management System (EMS). This and next chapters will further discuss the application of uncertainty modelling in load forecasting problems. In details, twofold of applications will be presented in this and next chapters: i) how to improve the **point load forecasting** at the domestic level; ii) and how to achieve enhanced **probabilistic load forecasting** with uncertainty modelling techniques at domestic level. This chapter will discuss the first application in point load forecasting.

Due to the high volatility and uncertainty of load profiles. Traditional methods tend to avoid such uncertainty by load aggregation (to offset uncertainties), load clustering (to cluster uncertainties) and load denoising (to filter out uncertainties). This chapter for the first time to improve the accuracy of load forecasting by directly modelling the parts of uncertainty components, which is commonly shared by neighbourhoods.

The following parts are organized as: Section 6.2 states the problem of point load forecasting and probabilistic load forecasting. Section 6.3 discusses the challenge of household point load forecasting considering uncertainty. Section 6.4 discusses the rationales of using Deep Learning techniques for modelling uncertainty. The proposed Deep Learning model architecture is discussed in Section 6.5. A novel training strategy named as Pooling Strategy is introduced in Section 6.6. Implementation and Demonstrations are presented in Section 6.7 & 6.8.

The content of this chapter is cited from the author's published article in IEEE Transactions on Smart Grid. The structure of this chapter is organised in alternative-based format.

6.2 Point Load Forecasting and Probabilistic Load Forecasting

To clarify the contribution and distinguishes between chapter 7 & 8, the terminologies of point load forecasting and probabilistic load forecasting will be discussed first in this part.

The terminology 'Load Forecasting' normally refers to 'Point Load Forecasting' (point LF), which has been widely recognised and researched for decades. The output of point LF is a single forecast for estimating the demand at a future time point. This point forecasting can be regarded as to forecast the mathematic expectation to the uncertain load. However, point forecasting can only provide a point value forecast that cannot contain sufficient information regarding the uncertainty that lies in the electric load. Therefore, over the last decades, the requirements are largely increased to achieve forecasting not only the expectation but also the probability. One alternative to meet this technical requirement is the probabilistic load forecasting. In the status quo, extensive and comprehensive review papers on point load forecasting at the aggregated level already exist [117 - 130]. However, most works are done on aggregated levels [131-139] since the disaggregated demand is extremely uncertain and volatile.

Probabilistic load forecasting (probabilistic LF), literally refers to the forecast outputs are associated with probabilities. The forms of probability expressed by probabilistic LF can be various, such as quantiles, scenarios, or even probability density functions. Compared to point load forecasting, probabilistic LF can provide more informative and intuitive forecasting results that associate uncertainty distribution with the expectation. It can meet the requirement of many smart grid applications, such as stochastic unit commitment, risk minimization in energy management system, and so on. Unlike the extensive literature in point load forecasting, prior works in probabilistic load forecasting are relatively limited. In the energy sector, probabilistic LF is mostly investigated in electric and wind load forecasting problems. For instance, in works [141], probabilistic LF is implemented in the scenario of wind forecasting problem. In addition, Hong, et al [130, 140] also reviews the frameworks and methods of probabilistic load forecasting electric forecasting problems. Most of the prior works are based on Multi Linear Regression and its meta methods.

6.3 Deep Learning for Household Load Forecasting- A Novel Pooling Deep RNN

This section presents the original article published in IEEE Transactions on Smart Grid [142], the indexes, equations, table, figures and titles are numbered independently.

Deep Learning for Household Load Forecasting – A Novel Pooling Deep RNN

Heng Shi, Minghao Xu, Furong Li, *Senior Member, IEEE*, Ran Li, *Member, IEEE*

***Abstract*—The key challenge for household load forecasting lies in the high volatility and uncertainty of load profiles. Traditional methods tend to avoid such uncertainty by load aggregation (to offset uncertainties), customer classification (to cluster uncertainties) and spectral analysis (to filter out uncertainties). This paper, for the first time, aims to directly learn the uncertainty by applying a new breed of machine learning algorithms – deep learning. However simply adding layers in neural networks will cap the forecasting performance due to the occurrence of overfitting. A novel pooling-based deep recurrent neural network (PDRNN) is proposed in this paper which batches a group of customers' load profiles into a pool of inputs. Essentially the model could address the over-fitting issue by increasing data diversity and volume. This work reports the first attempts to develop a bespoke deep learning application for household load forecasting and achieved preliminary success. The developed method is implemented on Tensorflow deep learning platform and tested on 920 smart metered customers from Ireland. Compared with the state-of-art techniques in household load forecasting, the proposed method outperforms ARIMA by 19.5%, SVR by 13.1% and classical deep RNN by 6.5% in terms of RMSE.**

***Index Terms*—big data, deep learning, load forecasting, long short-term memory, machine learning, neural network, smart meter**

I. INTRODUCTION

DEMAND side response (DSR) plays a key component in achieving the political goals set in the UK and EU energy sector [1, 2]. The popularisation of smart meters will make the DSR easier than ever for domestic customers [1]. Various direct and indirect control methods have been proposed to realise DSR [3-5] given that household load can be accurately forecasted.

Extensive and comprehensive review papers on point load forecasting at aggregated level already exist [14-26, 40-42]. However, the literature on individual household load forecasting is actually limited [5-14] as it is widely acknowledged that short-term load forecasting (STLF) at such granular level is extremely challenging due to significant uncertainty and volatility [6-8] underlying the smart metering data. Experiments have been

carried out by [6, 7, 9-13] to benchmark state-of-art methods for STLF at individual household level. Testing methods include time-series analysis (e.g. ARIMA and exponential smoothing) and machine learning approaches (e.g. neural networks and support vector machine). Similar findings are reported in both papers [9, 10] as none of the classical methods could beat linear regression or even simple persistence forecasting (i.e. tomorrow equals today) at individual household level.

1.1 Uncertainty

The complexity of household load forecasting lies in the significant volatility and uncertainty. In the context of STLF, load could be decomposed into three components. As shown in Fig. 1, the original household load profile i) is decomposed into: ii) regular pattern, which reflects the periodical load inherited from historical data; iii) uncertainty, which is the aperiodic part influenced by external factors such as weather, events and customer behaviour and iv) noise, the residue load which cannot be physically explained [14, 15].

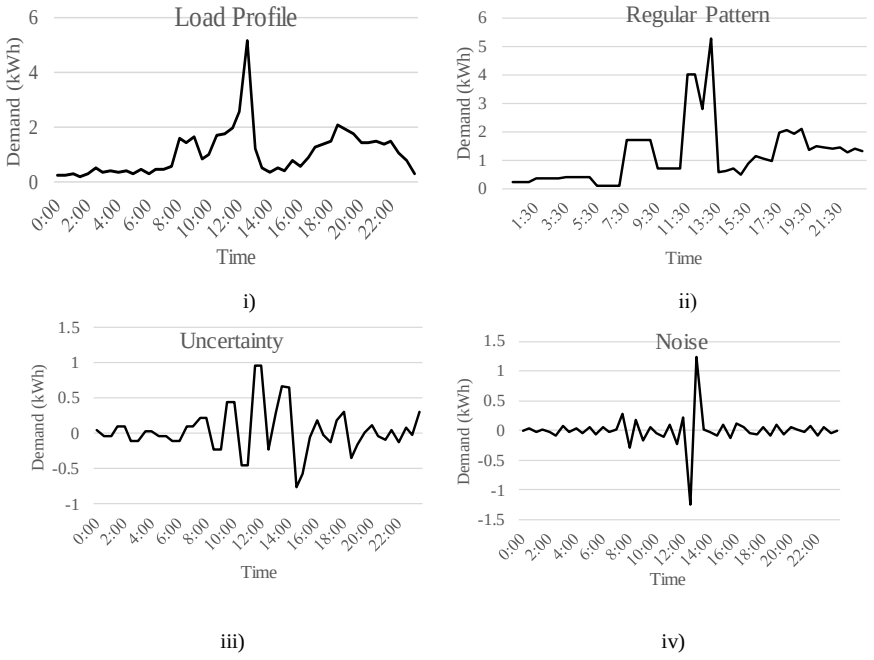


Fig. 1. Sketch of load composition: i) original load, ii) regular pattern, iii) uncertainty and iii) noise[1]

Most forecasting models focus on the regular pattern as it is more predictable and makes up a dominating proportion at the aggregated level. However, household demand is composed of a substantially larger share of uncertainty. At this level, uncertainty is more influenced by customer behaviour, which is too stochastic to predict. Therefore, the nature of the challenge is to forecast load with significant uncertainty. To tackle this problem, three categories of methods have been reported:

1 Uncertainty and noise in figure 1 are equivalent to terminologies epistemic and aleatory uncertainty in former chapters.

1) Using clustering/classification techniques to group similar customers, days or weather [6, 16-19] in the hope of reducing the variance of uncertainty within each cluster. However, the performance is heavily dependent on the data.
2) Using aggregated smart metering data to cancel out the uncertainties [20-23] so that the aggregated load exhibits mostly regular patterns and easier to predict, yet the prediction is obviously only at aggregated level.
3) Using pre-processing techniques, mostly spectral analysis such as wavelet analysis [24], Fourier transforms [25] and empirical mode decomposition (EMD) [26] aiming to separate the regular patterns from the other two components. This method can be ruled out in household load forecasting due to its significantly lower proportion of regular patterns.

To the best of our knowledge, the existing methods towards the problem are indirect, which aim to avoid uncertainty by reducing (clustering) or canceling out (aggregation) or separating (spectral analysis) the uncertainty. This paper aims to explore the possibility of deploying the state-of-art deep learning algorithm to directly learn uncertainties in their raw forms. Deep learning is a branch of machine learning methods relying on 'deep' architectures, which are compositions of multiple processing layers in the neural network, enabling the learning of highly non-linear, complicated relationships and correlations that are beyond the reach of traditional shallow architectures. Deep learning has achieved many breakthroughs in tackling sophisticated problems and becomes the most promising technique in data science community, for example, Google Goggles, Alpha Go [27] and new drugs design [28]. Attempts have been reported in [29, 30] to adopt deep learning for time series forecasting.

1.2 Overfitting

However, direct implementation of deep learning in household load forecasting will not necessarily provide significant improvement. A preliminary test has been carried out by the authors to benchmark the performance of household load forecasting by a neural network with a different number of layers.

The indicative result shown in Fig. 2 demonstrates the occurrence of overfitting when the number of layers reaches 3. As the number of layers increases, the forecasting error decreases before 3 layers. However, further increase of the network depth will see a rebound of error. As acknowledged in most literature [31], the primary drivers are model capacity and training epochs (training iterations). To prevent excessive training iterations, we implemented early stopping technique to find optimal number of training iterations. In

detail, dataset is split into training, validation and test sets. In each of the training iteration, the process will stop if the RMSE on validation set no longer decreases for a certain number of epochs.

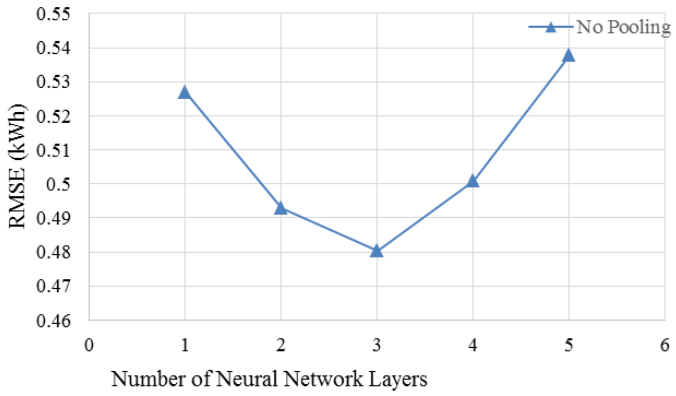


Fig. 2. Household load forecasting performance by neural networks from shallow to deep

Model capacity refers to the ability to fit a wide variety of functions. Model with large capacity tends to suffer from overfitting. To avoid excessive model capacity, one way is to increase the data diversity so that sufficient model capacity is becoming an advantage rather than a burden. Particularly for Deep Learning techniques, whose model capacity is much larger than the rest of models. When increasing the deep neural network layer number, the inherent parameters with the network will grow exponentially and eventually become excessive for the available training data. As a result, the model will begin to capture the noise and fit the training data too well, hence impact the predictive performance in a negative way.

In order to enable the power of deep learning algorithm in our problem, a novel pooling-based deep recurrent neural network (PDRNN) is proposed. The pooling technique will batch customers and input into the deep recurrent network as a whole.

The key contributions of this paper are as follows:

1) New technique: this paper for the first time explores the feasibility of a cutting edge algorithm, deep learning, in the application of load forecasting at individual household level.
2) New problem: although deep learning has received high expectation in forecasting community, our experiment indicates that deep learning is more prone to over-fitting compared with its 1980s' cousin, neural networks. This is possibly due to more parameters and relatively fewer data.

3) New method: we propose a novel pooling method to address the over-fitting issue by introducing a new data dimension: historical load data of neighbours. The idea is to use the interconnected spatial information to compensate insufficient temporal information. The proposed load profile pool allows for the correlations and interactions between neighbouring households. New features can be automatically generated through deep layers hierarchically and thus increases the inputs volume and diversity.

The rest of the paper is organised as follows: Section II briefly introduces the rationale for applying Deep Learning in household STLF tasks and the specific LSTM technique applied in the paper. Section III proposes pooling strategy and pooling-based DRNN method. Section IV explains the implementation process on GPU-based high-performance computing platform, as well as the details of experiment setup. In Section V, results are demonstrated through comparison with previous state-of-the-art methods (ARIMA, SVR), shallow learning (normal RNN), classical deep learning (DRNN) and proposed deep learning (PDRNN). Conclusions are drawn in part VI.

II. DEEP LEARNING

Deep learning is a branch of machine learning methods lying on ‘deep’ network architectures. The concept of ‘deep learning’ has been proposed for decades with the name ‘cybernetics’ in 1943, by McClulloch and Pitts [32]. However, it has been regarded as being more of a fancy concept than an applicable technology, due to three major technical constraints. The three technical constraints are: 1) lack of sufficient data, 2) lack of computing resources for large network size, and 3) lack of efficient training algorithm.

Recently, the constraints are tackled by the digitalization of modern society and the development of high-performance computing. Furthermore, Geoffrey Hinton [33] made a breakthrough in efficient deep neural network training via a strategy called greedy layer-wise pre-training, which enables practical implementations of deep learning.

Deep learning has recently seen phenomenal success in various areas including: 1) Computer Vision (CV) such as Google Goggles, which uses deep learning for object recognition; 2) expert systems such as Alpha Go designed by DeepMind [27] and 3) medical sciences, which employs deep learning to assist pharmaceutical companies in new drugs design [28].

A. Rationale of using deep learning

Deep learning is regarded as one the most promising techniques to this study due to two superior attributes compared with "shallow" architecture:

1) To learn highly non-linear relationships

In the problem of STLF at the household level, the inherent uncertainties are caused by differing known or unknown external factors simultaneously. These factors, ranging from weather conditions, temperatures to property size, photovoltaic generations are correlated to each other, which leave a highly non-linear impact to the household load. For example, temperature and sunshine duration are two of the external factors which are highly correlated to each other, i.e., the increase in sunshine duration can result in higher temperature in the region.

The essence of neural networks and other load forecasting methods is to learn the non-linear relationships between feed-in inputs and outputs by constructing linear or non-linear functions that approximate the real relationships between inputs and outputs. The universe approximation theorem [34] indicates the neural networks can make accurate approximations towards any non-linear functions with sufficient network size. The approximation capability of a shallow network is much lower than that of a deep network even with extra neurons at each layer. The reason is that, in 'shallow' neural networks, hidden neurons are learning the non-linear combinations of inputs as the features. However, 'deep' neural network can learn the non-linear combinations of features in deeper layers of the network, hence naturally learns the highly non-linear correlations.

2) To learn shared uncertainties

The uncertainties are normally coming from external sources which make consistent impacts on differing households. Therefore, these uncertainties can be commonly shared within a group of customers at similar locations and time. However, these uncertainties are not always evenly shared among households. For example, the temperature increase can impact most of the households within a region, while the increase in sunshine duration mainly affects households with PV installed.

In 'deep' architecture, one of the most exciting properties is that it can learn load features hierarchically. Features with different sharing levels will be learned at different layers. Load features learned in higher layers are normally the combination of lower layer features. With respect to former example, the temperature rise features are normally learned at a lower level, since it can be concluded directly from inputs. However, the impact from sunshine hour is influenced by features like temperature, PV installation, and household direction, and hence should be learned at higher network layers. With this property, deep

B. Deep RNN with Long Short-Term Memory unit

Typical architecture designs of deep learning including, Convolutional Deep Neural Networks (CNN), Deep Sparse Autoencoder (DSA), Deep Recurrent Neural Networks (DRNN), Multi-Layer Perceptions (MLP), Deep Restricted Boltzmann Machines (DRBM), etc. [31]. As a state-of-the-art deep learning architecture specifically designed for time-series forecasting, DRNN is employed to perform STLF for households in this paper.

The architecture of deep-RNN is stacking multiple RNN layers together into a 'deep' architecture. Most successful implementation of Deep-RNN is in the area of Speech Recognition [35], which is also one-dimensional time-series data with high uncertainty. In terms of the specific implementation of RNN layers, a state-of-the-art RNN, named Long Short-Term Memory (LSTM) has been employed to approach the best performance of RNN.

In this section, the deep-RNN architecture is introduced firstly, and then the implementation of LSTM units are followed.

1) Deep recurrent neural network (Deep-RNNs)

In deep-RNNs, the sharing states are decomposed into multiple layers in order to gain nice properties from 'deep' architectures. Experimental evidence has been given by [35, 36] to suggest the significant benefit of building RNNs with 'deep' architectures.

The computational graph and its unfolding topological graph is presented in Fig. 3 to demonstrate the working process of a deep-RNN with *N* layers.

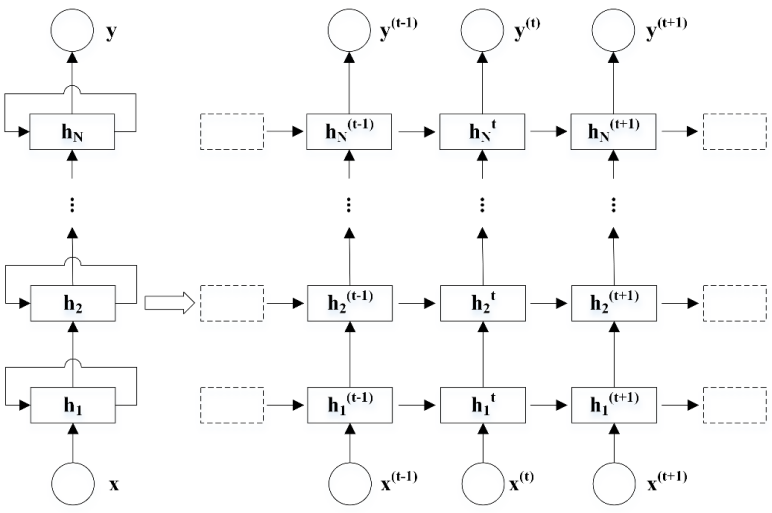


Fig. 3. The computational graph and unfolded topological graph of an *N* layer deep-RNN

In the computational graph, the RNN aims to map the input sequence of x values into corresponding sequential outputs: y. As presented in computational graph, the learning process conducted every single time step from $t = 1$ to $t = \tau$. For time step t, the network neuron parameters at l^{th} layer update its sharing states with following equations [31]:

$$a_1{}^{(t)} = b_1 + W_1 \cdot h_1{}^{(t-1)} + U_1 \cdot x^{(t)} \tag{1}$$

$$h_l{}^{(t)} = activation_{function(a_l{}^{(t)})} \quad for\ l = 1,2,..,N \tag{2}$$

$$a_l{}^{(t)} = b_l + W_l \cdot h_l{}^{(t-1)} + U_l \cdot h_{l-1}{}^{(t)} \quad for\ l = 2,3,\dots,N \tag{3}$$

$$y^{(t)} = b_N + W_N \cdot h_N{}^{(t-1)} + U_N \cdot h_N{}^{(t)} \tag{4}$$

$$L = loss\_function(y^{(t)}, y_{target}{}^{(t)}) \tag{5}$$

Where $x^{(t)}$is the data input at t^{th} time step, $y^{(t)}$ is the corresponding prediction, and $y_{target}{}^{(t)}$ is the true values of output targets. $h_l{}^{(t)}$ is the sharing states of l^{th} network layer at time step t. $a_l{}^{(t)}$ represents the input value of l^{th} layer at time step t, which consists of three components: 1) t^{th} time step input $x^{(t)}$ or sharing state $h_{l-1}{}^{(t)}$ at time t from $l-1^{th}$ layer, 2) bias b, and 3) sharing states $h_l{}^{(t-1)}$ at current network layer l from last time step $t-1$. Due to the sharing properties of RNNs, the algorithm is thus capable to learn uncertainties repeated in previous time steps.

2) Boosting with Long short-term memory (LSTM) unit

Long short-term memory unit refers to a specific architecture of RNNs, which aims to tackle long-term dependencies challenge unsolved in earlier RNN architectures. When learning time-series data, RNNs aim to learn representations of patterns repeatedly occurred in the past, by sharing parameters across all time steps. However, the memory of past learned patterns can fade as time goes on. In the figure, the dependencies of earliest two inputs $x^{(0)}$ and $x^{(1)}$ becomes weak in output $y^{(t)}$ when it is reasonably large.

LSTM is hence designed to tackle this challenge by creating paths where the gradient can

patterns, the computational graph of LSTM is illustrated in following Fig. 4:

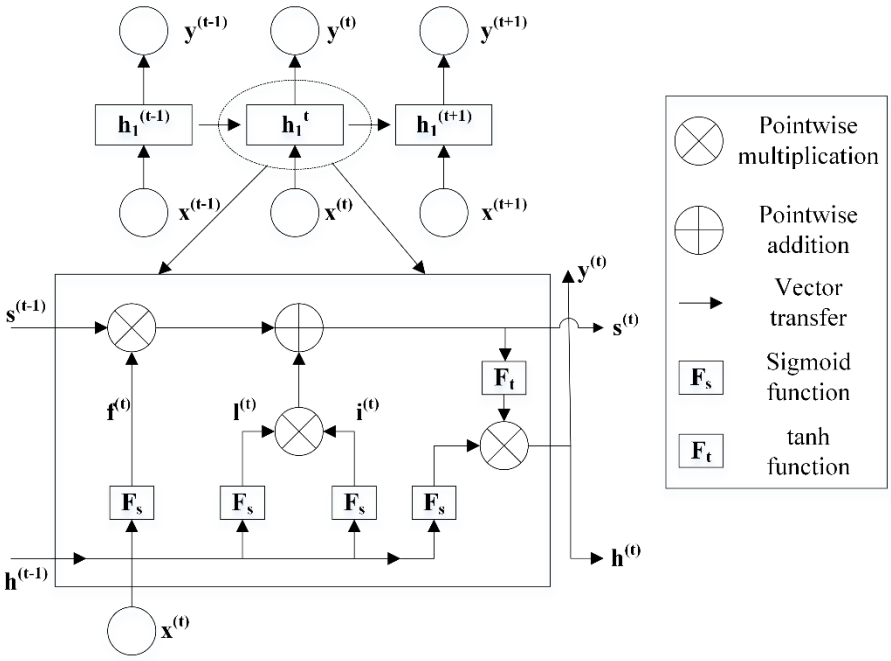


Fig. 4. The computational graph and unfolded topological graph

Fig. 4 presents a typical LSTM cell. Apart from traditional RNN units, LSTM cells have a special sharing parameter vector called memory parameter vector $s^{(t)}$ and are deployed to keep the memorized information. In each of the time steps, the memory parameter has three operations: 1) discard useless information from memory vector $s^{(t)}$; 2) add new information $i^{(t)}$ selected from input the $x^{(t)}$ and previous sharing parameter vector $h^{(t-1)}$into memory vector $s^{(t)}$; 3) decide new sharing parameter vector $h^{(t)}$ from memory vector $s^{(t)}$.

As shown in the LSTM cell, the sharing memory parameters $h^{(t)}$ are passing through differing time steps only with two operations to memorize new information and forget out-of-time memories. Therefore, the sharing memory can keep useful information for a fairly long time and result in RNN performance enhancement.

I. PROPOSED METHODOLOGY

In this section, the proposed PDRNN is presented for STLF at the household level. Details of this methodology are illustrated in Fig. 5:

In general, the proposed method consists of two stages: 1) load profiles pooling, and 2) household STLF with deep-RNN. The detailed rationale and design of each stage are further discussed in the following sub-sections:

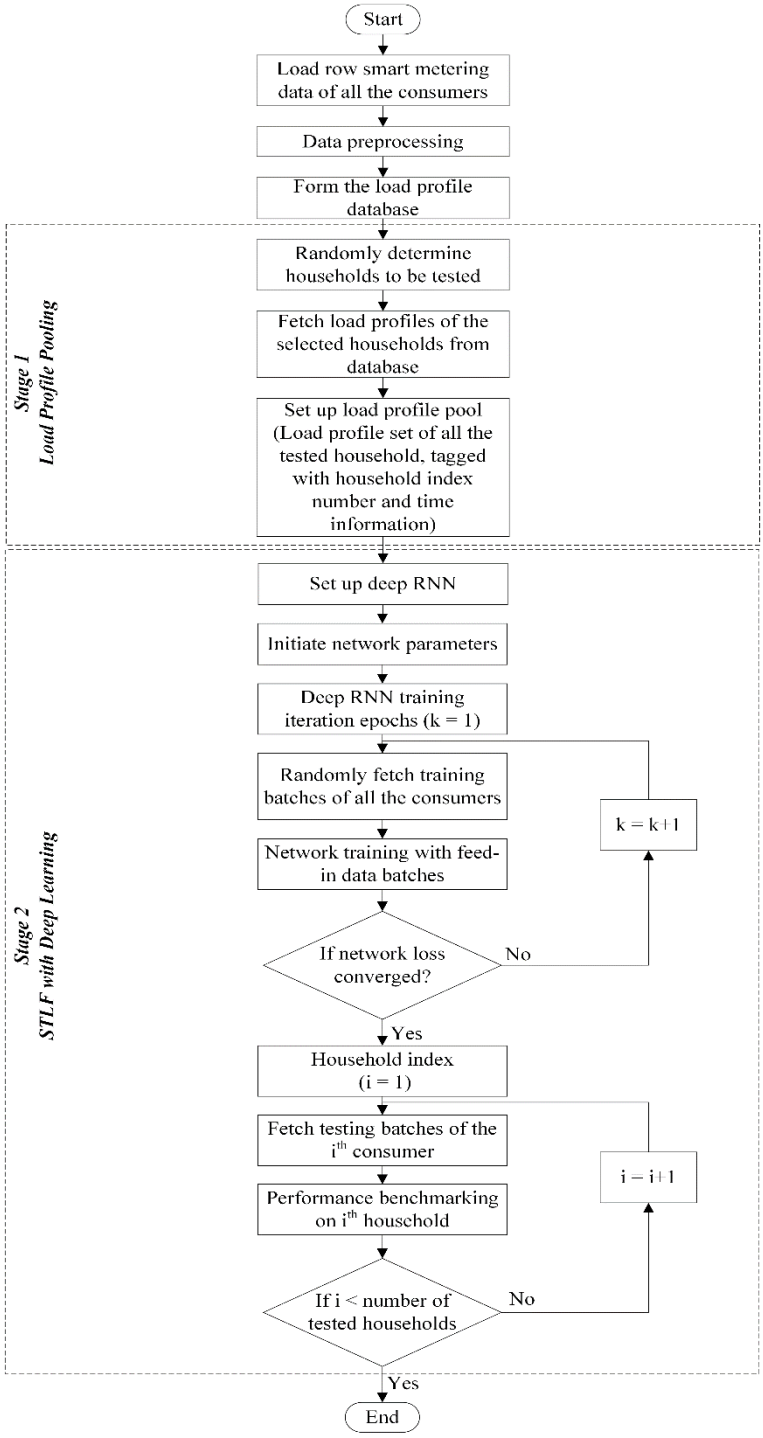


Fig. 5. Flowchart of proposed two-stage STLF methodology

1.3 Stage 1: load profiles pooling

In the 1^{st} stage, households' load profiles are batched into a load profile pool. The pool is fed into the 2^{nd} stage as input so that forecasting is not only based on targeted household's own data, but also load profiles of his neighbours in the pool.

3) Rationale of pooling strategy

The pooling strategy is designed to tackle the two major challenges of STLF at the household level, i.e., the overfitting issue and the inherent high uncertainties in household load profiles:

The overfitting issue is one of the technical constraints when applying deep learning in load forecasting. Because of the inherent large number of neural layers in deep learning networks, the available historical load profile data in households are normally insufficient, which even can cause grave overfitting with a fairly small amount of network layers. The pooling stage can increase the data volume for load forecasting, which hence delays the presence of overfitting.

Because of the contingency of the load data, the inherent load uncertainties are extremely hard to be learned or modeled. However, some of the uncertainties are caused by common external factors, such as weather conditions, the day of the week, etc. Their effects are normally sharing across many customers. According to the information theory, the data diversity represents the amount of information contained. Therefore, sufficient diversity in customer load is the prerequisite for learning these common sharing uncertainties. In this stage, pooling customers' load profiles together is basically to increase the diversity in load dataset, hence increases the information related to common sharing uncertainties. Consequently, it enables the deep recurrent network to perform more accurate load forecasting by understanding these common sharing uncertainties.

1) Design of pooling strategy

In this paper, the household load profiles are captured from smart meters half-hourly. Therefore, the daily load profiles are of the form of 48-data-point values. Due to time connectivity of household load between continuous dates, the load samplings on d^{th} date are continuous with $(d-1)^{th}$ and $(d+1)^{th}$ dates. In order to keep this property in data, the load profile pool uses a long vector sequence, consisting of concatenated load profiles of multiple continuous dates starting from the first date of historical data. The denotation is:

$$\overrightarrow{X^{(c)}} = (\overrightarrow{X^{(c)}}_1, \overrightarrow{X^{(c,2)}}_2, \dots, \overrightarrow{X^{(c,L)}}_L) \tag{6}$$

where L represents the total length of the demand sequence data for c^{th} $customers$. The load profile pool is then generated through 3 steps: 1) add customer id label in the form of dummy variables, 2) split data into training and test sets, 3) merge all training data to

in Fig. 6:

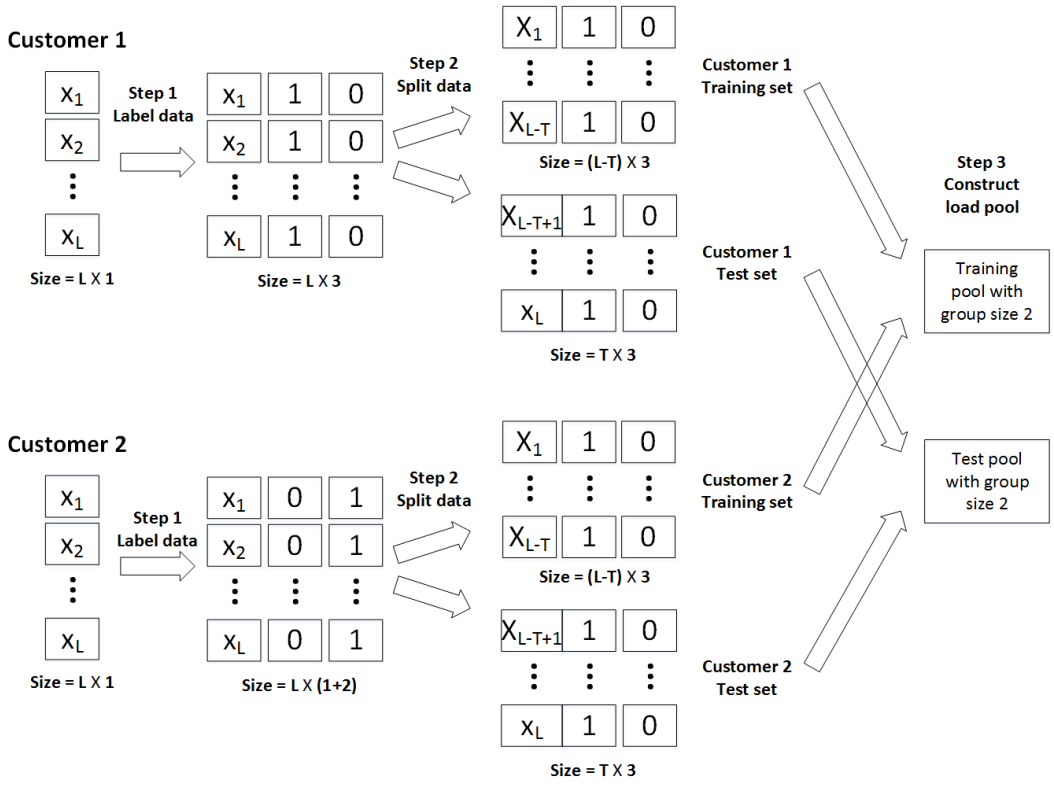


Fig. 6. Example of load pool construction with 2 customers group

As illustrated above, 1^{st} and 2^{nd} customers' demand are noted as two data sequences with size $L \times 1$ and $L \times 1$. In the first step, the demand data will be labelled with dummy variables to identify its customer id. In the example, the demand data are expanded with size $L \times 3$ and$L \times 3$. The number of expanded columns is equivalent to customer group size. In the second step, demand data of each customer will be split into training sets and test sets. The training sets of each customer are finally batched together as the training pool. Same procedure is taken to form the test pool.

1.4 Stage 2: pooling-based load forecasting using deep recurrent neural network

This stage of the proposed method consists of training and testing of pooling-based load forecasting: 1) In the training part, the deep recurrent neural network is trained by load profile batches randomly fetched from the load profile pool, so that deep-RNN are not only learning individual load patterns but also common sharing load features and uncertainties. 2) In the testing part, test load profiles are fed forward into the well-trained deep-RNN network.

Assuming the cleaned load profile database is Ψ_1, and the N testing households are listed in set $C = \{c_1, c_2, \dots, c_N\}$. The deep RNN configuration parameters are specified with L and H, which represent the network depth (number of layers) and amount of hidden units. With these parameters, the training and testing process can be conducted in following steps:

4) Initiation of deep recurrent neural network

At the beginning, the deep recurrent neural network is built with network configuration parameters, i.e., the network depth L, amount of hidden units H, batch size B, input sequence size I, and output sequence size O.

5) Network training iterations

After network initiation, the program is then running training iteration epochs until the network is well-trained with converged network prediction loss in the form of root mean squared error (RMSE).

$$Loss(y_{predict}, y_{target}) = \sqrt[2]{\frac{1}{B} \cdot \frac{1}{O} \cdot \sum_{i=1}^{B} \sum_{j=1}^{O} (y_{predict} - y_{target})^2} \tag{7}$$

In each of its training epochs, the training batch is firstly fetched from the load profile pool, then fed into the deep recurrent neural network. Each training batch is two matrices with fixed size, i.e., input matrix with size $B \times I$ and output matrix with size $B \times O$

The time-cost and iteration epochs of training process highly depend on feed-in data sequence size I, the choice of optimizer, network size (L, H) and training batch size B. In order to strike a well balance between training efficiency and efficacy, the training batch size B is variant during training: 1) at early epochs, B is set as a small number in order to approach the optimum point rapidly. 2) Then B is gradually increasing towards better training performance but sacrifices time cost.

6) Testing iteration and performance benchmarking

The well-trained deep recurrent neural network is then tested on individual households by performing as a feed-forward prediction neural network. In the testing process, load forecasting is conducted on testing households one by one, to identify whether the proposed methods can achieve a performance improvement of load forecasting individually. In each of the iterations, a performance comparison is made with other load forecasting methods, including ARIMA, SVR, RNN and deep RNN which are trained with load profile data

III. IMPLEMENTATION

This section introduces the implementation of the proposed methodology, including hardware, software platforms, and the program design.

A. Data Description

The data used in this paper are from the Smart Metering Electricity Customer Behaviour Trials (CBTs) initiated by Commission for Energy Regulation (CER) in Ireland. The trials took place during 1st July 2009 and 31st December 2010 with over 5000 Irish residential consumers and small and medium enterprises (SMEs) participating. The full anonymized data sets are publicly available online and comprise three parts: 1) half-hourly sampled electricity consumption (kWh) from each participant; 2) questionnaires and corresponding answers from surveys; 3) customer type, tariff and stimulus description, which specifies customer types, allocation of tariff scheme and Demand Side Management (DSM) stimuli [37].

In this trial, there were 929 1-E-E type consumers, meaning that they are all residential (1) customers with the controlled stimulus (E) and controlled tariff (E). To put it into perspective, these consumers were billed on existing flat rate without any DSM stimuli, which are most representative since the majorities of consumers outside the trial are of the type. In this paper, 920 1-E-E consumers were randomly selected as the testing customers. With group size 10, 920 consumers were split into 92 groups randomly.

Data with missing intervals are encountered and hence are not continuous. Different households may have different missing intervals and need to be pre-processed individually. Hardware and Software platforms.

B. Hardware and Software platforms

The program is implemented on a high-performance Dell workstation equipped with Ubuntu 14.04 operating system and a computable GPU unit. The deep learning code is programmed based on an open-sourced deep learning framework named as Tensorflow [38], which is developed by one of the leading industry in the deep learning community, Google. Superior features of it include: 1) it is designed for the most popular programming language in data science, i.e., Python; 2) it supports GPU-based high-performance parallel computing towards big data tasks; 3) it employs symbolic programming mechanism and enables computing graph optimization feature, which is the most cutting-edge technique in deep learning community.

1.5 Program Implementation

The deep learning program is designed with multiple stages: 1) data pre-processing and cleaning; 2) data pooling; 3) data sampling and 4) network training and 5) benchmark evaluation. The program design is demonstrated with pseudo code in Program 1:

Program 1: Deep learning program for STLF

Load dataset Ψ_0 of household demand from smart meters.
Clean and pre-process demand data in dataset Ψ_1.
Generate tuple set $< L, H, C >$ of testing parameters: deep-RNN layer number $l \subseteq L$, deep-RNN hidden unit number $h \subseteq H$, and testing households set C.
For parameters $< l, h, c >$ in tuple set $< L, H, C >$**:**
According to household set C, get generate load profile pool $\Psi \subseteq \Psi_1$.
Divide Ψ into training set Ψ_{tr} and test set Ψ_{ts}.
Build deep-RNN $\aleph$ with network size (l, h) on Tensorflow.
Repeat
At k^{th} epochs **Do**:
Train deep-RNN $\aleph$ with randomly fetched data batch $\Phi \subseteq \Psi_{tr}$
Evaluate performance by mean squared error Λ_k on cross-validation samples.
Update a performance queue:

$$\Omega = [\Lambda_{k-v}, \Lambda_{k-v+1}, \dots, \Lambda_{k-1}]$$

By pop out Λ_{k-v} from Ω, then push in Λ_k
End
Until $var(\Omega) \leq \varepsilon$, where ε is a convergence threshold.
End
For household c in set C**:**
Fetch test samples φ_c of household c from dataset Ψ_{ts}
Evaluate performance of $\aleph$ on test samples φ_c, with multiple performance benchmarks Θ
Compare load forecasting performance with other methods on household c.
End
Terminate

1.6 Experiment Setup

This part presents the details for setting up the experiments, including data pre-process, algorithm configuration.

Regarding the data pre-process, raw data from Irish dataset is manipulated into input data sets through three steps: 1) split all customers into sub-groups; 2) construct load profile pool for each customer group; 3) split each pool into training, validation and test sets. The

test set consist of data points during the last 30 days of available dataset (720 hours, 1440 data points), validation set is randomly selected from the rest of the data.

In order to reach optimal performance of each algorithm (SVR, ARIMA, RNN, DRNN, Pooling-based DRNN), we prepared multiple algorithm settings for each algorithm. However, not all results are reported in the result section, the comparison is made with the optimal settings of each algorithm.

In summary, all the experiment settings and parameters are presented as follows:

$customer\ group\ size \in \{10\}$
$customer\ group\ amount \in \{92\}$
$test\ set\ size \in \{1440\}\ points\ of\ data$
$validation\ set\ size \in \{2880\}\ points\ of\ data$
$RNN\ network\ layer\ number \in \{1\}\ DRNN\ network\ layer\ number \in \{2,3,4,5\}$
$PDRNN\ network\ layer\ number \in \{2,3,4,5\}$
$training\ batch\ size \in \{96, 240, 480\}\ points\ of\ data$
$validation\ batch\ size \in \{240, 480, 960\}\ points\ of\ data$
$hidden\ neuron\ number \in \{5,10,20,30,50,100\}$
$input\ sequence\ length \in \{48, 96,336\}$
$training\ method \in \{AdamOptimizer\}$
$neuron\ cell\ unit \in \{LSTM\}$
$learning\ rate \in \{0.001, 0.002, 0.005, 0.01\}$
$training\ stop\ strategy \in \{early\ stopping\}$
$loss\ function \in \{RMSE\}$

IV. RESULT AND DISCUSSION

In this section, the proposed method is validated on realistic smart metering load data from Irish load profile database [37]. The data selection and pre-processing are exploited in the data description section. To assess the performance of proposed method in conducting STLF for residential households, three widely used metrics are employed, including root mean squared error (RMSE), normalised root mean squared error, and mean absolute error [8].

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(\widehat{y_t} - y_t)^2}{N}} \tag{8}$$

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{9}$$

$$MAE = \frac{\sum_{t=1}^{N}|\widehat{y_t} - y_t|}{N} \tag{10}$$

Where $\widehat{y_t}$ is the forecasted value, y_t is the actual value, y_{max} and y_{min} is the maximum and minimum value among the test set. N refers to the test set size.

This assessment consists of three parts: 1) the performance of proposed method are compared to 3 methods and typical deep-RNN method to validate the efficacy; 2) the effect of network depth increase are illustrated to reveal the performance impact from 'shallow' to 'deep' architectures, to indicate the potential of deep learning for load forecasting and reveal the challenge of overfitting; and 3) the effect of pooling strategy are revealed by comparing proposed PDRNN typical with deep-RNN algorithm without pooling strategy, specifically to indicate the effect of pooling strategy to defer the overfitting issue.

1.7 Benchmarking of STLF methods in households

To validate the efficiency of the proposed PDRNN, three load forecasting methods, including autoregressive integrated moving average (ARIMA), support vector machine (SVR), and a 3-layer deep-RNN method are taken as a comparison and assessed under preceding mentioned benchmarks (RMSE, NRMSE, and MAE). The performance comparison across all testing residential households (920 households) is presented in Fig. 7 to Fig. 9 in form of heat map.

It is notable that the other 4 methods (RNN, SVR, DRNN, PDRNN) receive better average performance compared to ARIMA in the experiments. Therefore, we presents the performance improvement of 4 methods with respect to ARIMA method in the heat map. In the heat map, y axis refers to 4 methods (method 1: RNN, method 2: SVR, method 3: DRNN, method 4: PDRNN). x axis refers to 920 testing households. Lighter colour in the figure refers to better performance.

The results in Fig. 7 to Fig. 9 indicate that:

i) In terms of Average performance of three benchmarks, RNN and SVR achieve even performance, however, SVR performs slightly better than RNN in terms of RMSE and NRMSE. DRNN can receive a considerable improvement from RNN and SVR in all three benchmarks. The proposed PDRNN outperforms the other three methods, and can observe a clear reduction on all benchmarks.

ii) Regarding the results of different customers, the improvements of three benchmarks are not with same pattern. The improvements of RMSE among differing customer are largely diverse. While some customers receive 0.2 RMSE reductions, the other customers

more consistent.

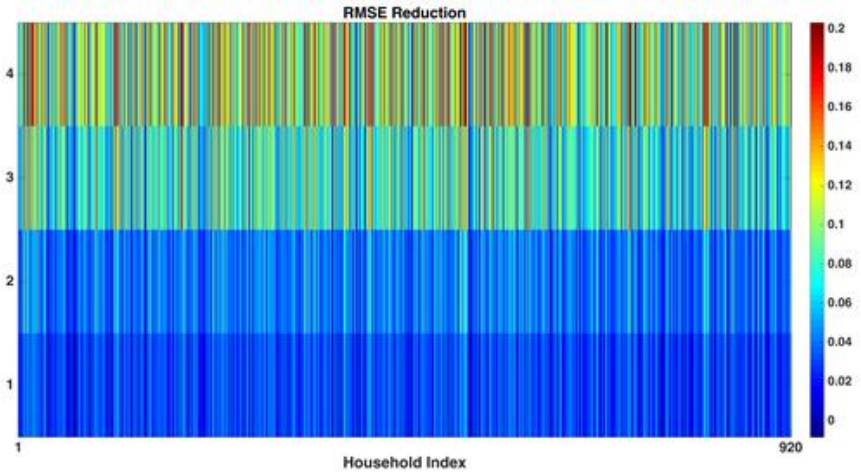


Fig. 7. RMSE reduction of 4 methods compared to ARIMA: 1) RNN, 2) SVR, 3) DRNN, 4) PDRNN

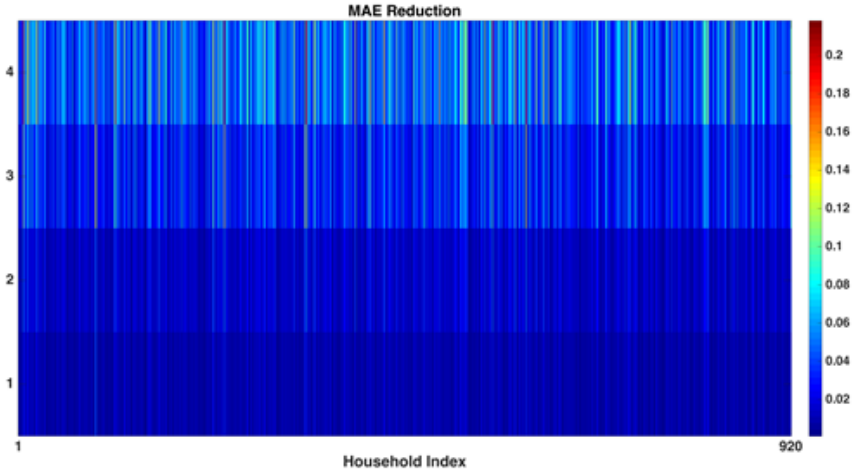


Fig. 8. MAE reduction of 4 methods compared to ARIMA: 1) RNN, 2) SVR, 3) DRNN, 4) PDRNN

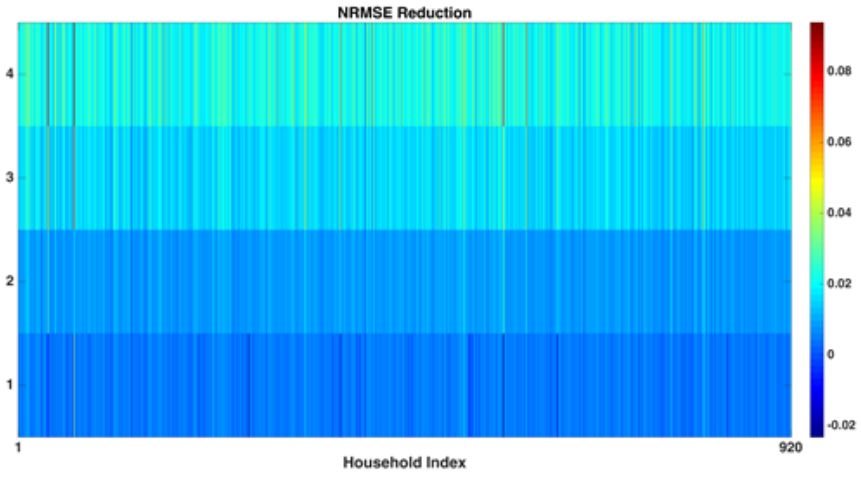


Fig. 9. NRMSE reduction of 4 methods compared to ARIMA: 1) RNN, 2) SVR, 3) DRNN, 4) PDRNN

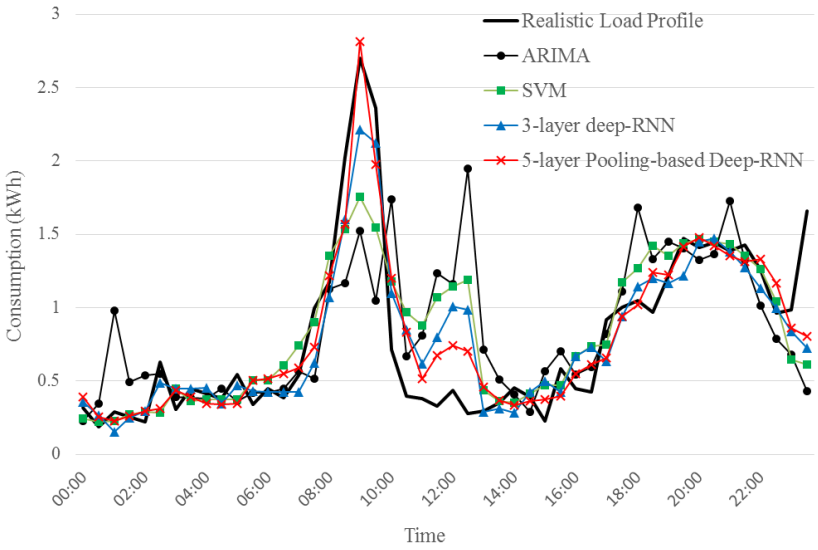


Fig. 10. The computational graph and unfolded topological graph

Furthermore, Fig. 10 demonstrates the real load and forecasted load by different methods on a random day 20 Jan. 2010, household 1059. The proposed method can deliver substantially improved performance at spikes and troughs. As shown in the figure, the morning peak during 8:00 a.m. and 10:00 a.m. is accurately captured by the proposed method. In addition, ARIMA, SVR, and 3-layer deep-RNN followed the inertia and predict a peak between 10:00 a.m. and noon while the proposed method successfully avoids overestimating.

1.8 Effect from 'shallow' to 'deep'

A sensitivity analysis is conducted to investigate the effect of network depth on load forecasting performance, in terms of neural network based load forecasting methods. To make a fair assessment, recurrent neural networks with differing depth are all: 1) enhanced with LSTM units, 2) subjected to same input size, output size, network configuration parameters, and 3) implemented on Tensorflow with Python. The results are presented in Fig. 11.

In Fig. 11, deep RNN witnesses the best performance with 2 to 3 layers, with around 0.485 in $RMSE$, 0.27 in R, and 0.1 in $NRMSE$. Further increase in network depth will lead to severe overfitting issue. With 5 layers, deep-RNN gives even worse result than 1-layer RNN.

In general, the sensitivity analysis on network depth indicates that increasing network depth into 'deep' can only enhance the accuracy up to a limit number of layers, which reflects the occurrence of overfitting, due to the lack of data diversity and network parameter redundancy [39].

1.9 Effect of proposed pooling strategy

The proposed pooling strategy attempts to tackle the occurrence of overfitting. The performance is investigated by comparing the load forecasting performance between deep-RNN methods with and without pooling at different depths. The corresponding results are demonstrated in Fig. 11:

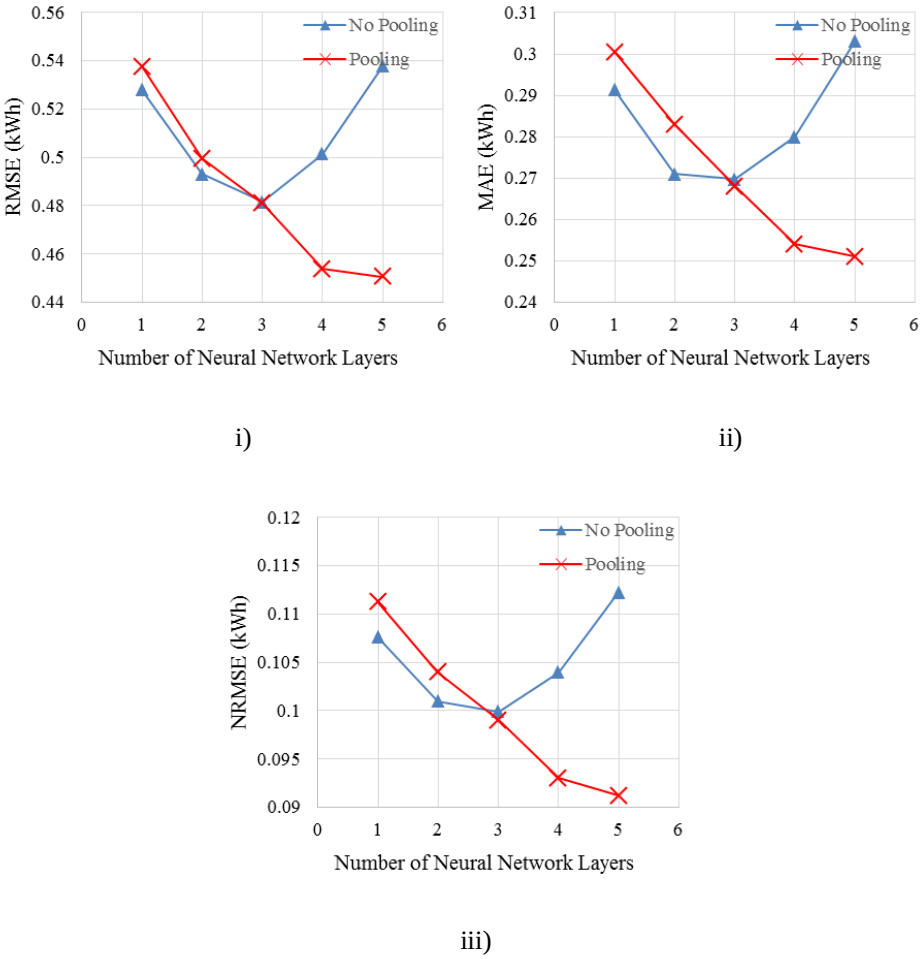


Fig. 11. Household load forecasting benchmarks from shallow to deep: i) root mean squared error (RMSE), ii) mean absolute error (MAE), iii) normalised root mean squared error (NRMSE).

In Fig 11, the proposed PDRNN (red line marked with cross) are compared with classical deep RNN method (blue line marked with triangle). In terms of RMSE, MAE, NRMSE classical deep RNN's performance stops improve after 3 layers due to overfitting while the proposed method keeps improving as the number of layers increases till as deep as we tested.

TABLE I

PERFORMANCE COMPARISON

Network Architecture	*RMSE (kWh)*	*NRMSE (kWh)*	*MAE (kWh)*
ARIMA	0.5593	0.1132	0.2998
RNN	0.5280	0.1076	0.2913
SVR	0.5180	0.1048	0.2855
DRNN	0.4815	0.0974	0.2698
PDRNN	0.4505	0.0912	0.2510
Improvement from DRNN to PDRNN	6.45%		6.96%
Improvement from ARIMA to PDRNN	19.46%		16.28%

Table I compares the performance of the proposed PDRNN in terms of RMSE, NRMSE, and MAE with four other techniques, i.e., classical DRNN, SVR, shallow RNN and ARIMA. All the presented metrics in the table take the averaged values across all the tested households. As illustrated, DRNN outperforms SVR, shallow RNN and ARIMA in all metrics used. With the introduction of the proposed pooling strategy, the new PDRNN with the same network settings (5 layers, with 30 hidden units in each layer), could further improve the performance. Specifically, compared with classical DRNN, the proposed PDRNN brings 6.45 % reduction in RMSE and NRMSE, 6.96 % reduction of MAE. Compared with ARIMA, the reduction in RMSE and MAE brought by PDRNN is even more significant, reaching 19.46% and 16.28% respectively.

V. CONCLUSION

This paper for the first time explores the potential of employing the state-of-art deep learning technique for household STLF under high uncertainty and volatility. A novel PDRNN is proposed to successfully address the overfitting challenges brought by the naive deep network. This paper proposes method enables learning of spatial information shared between interconnected customers and hence allowing more learning layers before the occurrence of overfitting.

The result indicates the proposed method can deliver significant improvement for household load forecasting. Compared with state-of-the-art, the proposed method outperforms ARIMA by 19.5%, SVR by 13.1% and classical deep RNN by 6.5% in terms of RMSE and similar performance under other metrics.

Although quantitative comparison has been conducted, we would like to emphasize that we do not draw an arbitrary conclusion of the superiority of deep learning model. The key findings are the overfitting problem identified in direct applying deep learning models and the novel pooling methodology developed to overcome the limitation. The paper aims to report the preliminary attempt and provide learnings for wider researchers who aim to tap into this state-of-the-art technique. Future work includes:

i) To exploit the overfitting point by further extending the network size.

ii) To exploit optimal pooling strategy by pooling customers with differing features, such as similar geographic locations, similar social status.

iii) To further exploit the potential of proposed method by considering more external factors, for instance, weather information.

References

[1] "Creating the right environment for demandside response," Ofgem, London30 April 2013 2013.

[2] "Demand Side Response Policy Paper," European network of Transmission System Operators for Electricity, Belgium, 15 September 20142014.

[3] A. Garulli, S. Paoletti, and A. Vicino, "Models and Techniques for Electric Load Forecasting in the Presence of Demand Response," *IEEE Transactions on Control Systems Technology*, vol. 23, pp. 1087-1097, 2015.

[4] P. Du and N. Lu, "Appliance Commitment for Household Load Scheduling," *IEEE Transactions on Smart Grid*, vol. 2, pp. 411-419, 2011.

[5] X. Liu, L. Ivanescu, R. Kang, and M. Maier, "Real-time household load priority scheduling algorithm based on prediction of renewable source availability," *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 318-326, 2012.

[6] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating Practice Theory in Sub-Profile Models for Short Term Aggregated Residential Load Forecasting."

[7] M. Chaouch, "Clustering-Based Improvement of Nonparametric Functional Time Series Forecasting: Application to Intra-Day Household-Level Load Curves," *IEEE Transactions on Smart Grid*, vol. 5, pp. 411-419, 2014.

[8] C. J. Willmott, S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, *et al.*, "Statistics for the evaluation and comparison of models," 1985.

[9] A. Marinescu, C. Harris, I. Dusparic, S. Clarke, and V. Cahill, "Residential electrical demand forecasting in very small scale: An evaluation of forecasting methods," in *Software Engineering Challenges for the Smart Grid (SE4SG), 2013 2nd International Workshop on*, 2013, pp. 25-32.

[10] S. Humeau, T. K. Wijaya, M. Vasirani, and K. Aberer, "Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households," in *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*, 2013, pp. 1-6.

[11] A. Veit, C. Goebel, R. Tidke, C. Doblander, and H.-A. Jacobsen, "Household electricity demand forecasting: benchmarking state-of-the-art methods," in *Proceedings of the 5th international conference on Future energy systems*, 2014, pp. 233-234.

[12] S. Haben, J. Ward, D. V. Greetham, C. Singleton, and P. Grindrod, "A new error measure for forecasts of household-level, high resolution electrical energy consumption," *International Journal of Forecasting*, vol. 30, pp. 246-256, 2014.

[13] Y.-H. Hsiao, "Household electricity demand forecast based on context information and user daily schedule analysis from meter data," *IEEE Transactions on Industrial Informatics*, vol. 11, pp. 33-43, 2015.

[14] Y. Wang, Q. Xia, and C. Kang, "Secondary Forecasting Based on Deviation Analysis for Short-Term Load Forecasting," *IEEE Transactions on Power Systems*, vol. 26, pp. 500-507, 2011.

[15] Tao Hong and D. A. Dickey. (2013, 01 August 2016). *Electric load forecasting: fundamentals and best practices*. Available: https://www.otexts.org/elf

[16] F. L. Quilumba, W. J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities," *IEEE Transactions on Smart Grid*, vol. 6, pp. 911-918, 2015.

[17] H. Tao, W. Pu, A. Pahwa, G. Min, and S. M. Hsiang, "Cost of temperature history data uncertainties in short term electric load forecasting," in *Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on*, 2010, pp. 212-217.

[18] T. Hong, P. Wang, and L. White, "Weather station selection for electric load forecasting," *International Journal of Forecasting*, vol. 31, pp. 286-295, 4// 2015.

[19] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *International Journal of Forecasting*, vol. 32, pp. 585-597, 7// 2016.

[20] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. S. Venkata, *et al.*, "An Efficient Approach to Short-Term Load Forecasting at the Distribution Level," *IEEE Transactions on Power Systems*, vol. 31, pp. 2526-2537, 2016.

[21] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—Part II: Peak load estimation by clusterwise regression," *IEEE Transactions on Power Systems*, vol. 30, pp. 3045-3052, 2015.

[22] A. Espasa and I. Mayo-Burgos, "Forecasting aggregates and disaggregates with common features," *International Journal of Forecasting*, vol. 29, pp. 718-732, 10// 2013.

[23] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol. 98, pp. 40-49, 2016.

[24] Y. Chen, P. B. Luh, C. Guan, Y. Zhao, L. D. Michel, M. A. Coolbeth, *et al.*, "Short-Term Load Forecasting: Similar Day-Based Wavelet Neural Networks," *IEEE Transactions on Power Systems*, vol. 25, pp. 322-330, 2010.

[25] R. Al-Otaibi, N. Jin, T. Wilcox, and P. Flach, "Feature Construction and Calibration for Clustering Daily Load Curves from Smart-Meter Data," *IEEE Transactions on Industrial Informatics*, vol. 12, pp. 645-654, 2016.

[26] D. Shi, R. Li, R. Shi, and F. Li, "Analysis of the relationship between load profile and weather condition," in *2014 IEEE PES General Meeting | Conference & Exposition*, 2014, pp. 1-5.

[27] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484-489, 01/28/print 2016.

[28] A. Lusci, G. Pollastri, and P. Baldi, "Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules," *Journal of chemical information and modeling*, vol. 53, pp. 1563-1575, 2013.

[29] C. Y. Zhang, C. L. P. Chen, M. Gan, and L. Chen, "Predictive Deep Boltzmann Machine for Multiperiod Wind Speed Forecasting," *IEEE Transactions on Sustainable Energy*, vol. 6, pp. 1416-1425, 2015.

[30] E. Busseti, I. Osband, and S. Wong, "Deep learning for time series modeling," Technical report, Stanford University2012.

[31] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning," *An MIT Press book in preparation. Draft chapters available at http://www. iro. umontreal. ca/∼ bengioy/dlbook*, 2015.

[32] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, pp. 115-133, 1943.

[33] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, pp. 1527-1554, 2006.

[34] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, pp. 359-366, 1989.

[35] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*, 2013, pp. 6645-6649.

[36] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *arXiv preprint arXiv:1312.6026*, 2013.

[37] "Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report," The Commission for Energy Regulation, Dublin, Information Paper2011.

[38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[39] D. M. Hawkins, "The problem of overfitting," Journal of chemical information and computer sciences, vol. 44, pp. 1-12, 2004.

[40] Tao Hong, Pierre Pinson, Shu Fan, Global Energy Forecasting Competition 2012, International Journal of Forecasting, Volume 30, Issue 2, April–June 2014, Pages 357-363, ISSN 0169-2070,.

[41] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, Rob J. Hyndman, Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond, International Journal of Forecasting, Volume 32, Issue 3, July–September 2016, Pages 896-913, ISSN 0169-2070,

[42] Tao Hong, Shu Fan, Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting, Volume 32, Issue 3, July–September 2016, Pages 914-938, ISSN 0169-2070, http://dx.doi.org/10.1016/j.ijforecast.2015.11.011.

6.4 Chapter Summary

This chapter for the first time explores to improve forecasting accuracy by directly modelling the epistemic uncertainty components shared by neighbourhoods. A pooling-based strategy is then proposed and integrated into the Deep Learning forecasting framework, i.e., the Deep Recurrent Neural Network.

The result indicates the proposed pooling strategy can deliver 6.5% improvement in the measure of RMSE compared to vanilla DRNN network without pooling strategy. This result indicates the validity of the proposed idea: to improve forecasting accuracy by modelling uncertainty. This idea will be further extended in Chapter VII, to completely model the uncertainty components, and connect these components with observable uncertainty sources.

Additionally, interesting visualization results can be found from Fig. 7 to Fig. 9. Although the performance enhancements of load forecasting can be easily identified by comparing the rows of heatmap Fig. 7 to Fig. 9. Forecasting performances over different households are largely diversified. For some households, the performance by proposed PDRNN method even witnesses worse performance than traditional methods. According to the dataset, this phenomenon is caused by the diverse scenarios of households. For instance, residents in some households are in holiday or not living in the property for a long period, the electric load are consistently low. In this scenario, proposed PDRNN method may deteriorate the forecasting accuracy, since it tend to make more aggressive prediction by learning from its neighbours, rather than conservative predictions as traditional methods.

Chapter 7

Uncertainty Model Applied in Probabilistic Load Forecasting

THIS chapter further extends the idea of improving forecasting accuracy by directly modelling uncertainty components. An advanced Probabilistic Deep Learning model is developed for probabilistic load forecasting is presented.

7.1 Introduction

Chapter 7 has demonstrated how to improve the performance of point load forecasting with uncertainty modelling techniques. This chapter will further discuss the application in probabilistic load forecasting, to seek how to generate probabilities as outputs in the forecasting scheme, and how to integrate the uncertainty modelling into the typical load forecasting processes.

As discussed in chapter 6, probabilistic load forecasting not only predicts a single value approaching the expectation of future loads, but also aims to associate probabilities with the prediction. To achieve this aim, the methodologies proposed in Chapter 7 cannot be introduced directly to PLF problems. The rationales are twofold:

1) Neural Networks (including the model proposed in chapter 7) normally model and learn the uncertainty as latent variables or parameters to be stored in the networks, and finally approximating the expectation of uncertainty through averaging among all data samples. However, in PLF problem, the result should also express the uncertainty as quantiles, intervals or distributions. Therefore, this calls for advanced methods that are not only learning and storing the latent variables as model parameters but also can export these uncertainties to outputs.

2) In order to model the epistemic uncertainty that is caused by systematic effects with consistent mechanism behind, the model proposed in chapter 7 only learns the correlation of uncertainty components that are shared by neighbourhoods but ignore the specific epistemic uncertainty components caused by other factors and also the aleatory uncertainty. This configuration cannot explain all the uncertainty properties lies in the load profiles. Therefore, this chapter seeks for methods that can distinguish different uncertainty components and allocate the relationships between the input factors and these uncertainty components.

Therefore, this chapter further improves the deep learning techniques for achieving PLF tasks by using the uncertainty modelling technique proposed in chapter 5. In details, the deep learning model proposed in this chapter integrates two techniques in former chapters: i) the network architecture is established on Recurrent Neural Networks, which can model the temporal dependency between different resolutions of time

duration. This functionality is achieved by Long Short-Term Memory [143, 144] units and also proved in chapter 5; ii) Dropout Stochastic Regularization units are integrated into the DRNN model to bring random properties, which can enable DRNN model to generate probabilistic forecasting results.

7.2 Literature of prior PLF methods

As discussed in chapter 6, probabilistic load forecasting not only predicts a single value approaching the expectation of future load, but also aims to associate probabilities with the prediction. To meet the requirement from industries, the forms could be a probability distribution, intervals, quantiles and other relevant formats. Prior works that attempting to achieve PLF problems are normally to generalize existing methods that designed for point load forecasting problems by adding probabilistic attributes in model inputs, outputs or directly in the model architecture. These base methods are widely recognised and deeply researched by thousands of researches, and can be classified into three categories:

i) Regression models: These methods attempt to model the structure and patterns of variables through statistical approaches. Typical methods include Support Vector Regression (SVR) [145], Multi Linear Regression (MLR) [146, 147, 148], Autoregressive Moving Average (ARMA) [149], etc. In terms of the family of regression methods, prior systematic knowledge is the key component that is required to assist the modelling of the underlying physical system behind the data uncertainty. For instance, ARMA assumes the data follows the periodical pattern under different resolutions, therefore, it models the dependency between different resolutions of time spans, to achieve the forecasting on the periodical patterns of original time-series data. Another example is the MLR, which assumes the observed data is influenced by a set of factors, and the relationships between observed data and causing factors can be approximated by linear functions. Based on this assumption, the variations of data can be estimated by formulating the coefficients of these linear functions.

ii) Artificial Neural Nets: Other than regression methods, The family of Artificial Neural Networks (ANN) are also extensively deployed for load forecasting since the early 1990s [140]. Compared to regression models that are designed

under a certain amount of assumptions and theorems, ANN can be more generally used in different datasets for forecasting tasks since it does not need to specify and describe the underlying physical system explicitly. Despite this advantage in generality, the network architecture configuration and parameters setting also impact the performance of forecasting heavily. Therefore, most works regarding ANN for forecasting problems usually fell into threefold: i) to design better network architecture; ii) to estimate optimal network parameters; iii) to manipulate data and makes it more easily to forecast with ANN. For instance, many works segment the load data into the base load and residual/variation by decomposition methods [150] and forecast them individually.

In general, above two types of methods are originally designed for forecasting deterministic time-series data, specifically for point forecasting. Based on these methods, many works have been done to achieve PLF tasks by bringing random properties into the deterministic model. As the forecasting procedure can be intuitively demonstrated as a model with inputs (historical data and features), model functions (algorithms) and outputs (forecasted results. The existing approaches for producing probabilistic forecasts is generally following three frameworks, to bring randomness at each stage of forecasting procedure accordingly (inputs, model functions, and outputs).

- The first framework type is to introduce uncertain quantities into the model inputs whilst maintain the deterministic design of the model function. Therefore, by sampling realizations of inputs from the probability distribution of input uncertainty, we can obtain a collection of input realizations. For each input realization, a forecasting output can be generated accordingly and deterministically. By conducting this step over the collection of input realizations, a collection of forecasting output realizations can be obtained. Taking the example of PLF considering temperature associated with demand. Many works have proved the temperature information can assist the PLF to achieve higher accuracy, however, the future temperature is also estimated by the weather station and heavily affected by the underlying uncertainty. Some works of PLF represent temperature uncertainty as a set of scenarios with deterministic temperature estimated. In these methods, the uncertain

temperature is represented as a collection of possible temperature scenarios, such as similar day method that collect all the similar day temperature at past. Hereafter, a set of possible forecasts could be predicted.

- The second type of framework is to generate probability at the output stage. In general, the input and model functions are deterministic configurations. At first, this model will generate a deterministic output as well, which is the estimate of the expected uncertain load. By assuming the residual between forecasting value and the true value could follow a certain form of probability distributions, such as Gaussian distribution. A group of output realizations can be obtained by invoking sampling method to a prior known Gaussian distribution and added to the forecasted value. Following this setting, many works have attempted different principles to introduce uncertainties to the output residual, and hence witnessed different performances in probabilistic LF problem.

- The third type of framework is purely based on historical data and regards the data samples as realizations from its uncertainty distribution. It uses the specific design of model functions to achieve the purpose of PLF. For instance, if the expression of output uncertainty is the probability quantiles, the PLF model can be achieved by simply replacing optimization objective of the model function, i.e., the L_2 loss between the true value and forecasted value, into pinball loss [104, 105, 106]. The deterministic forecasting output will become the investigated quantiles rather than the expectation of uncertain load.

Although these techniques for PLF problems have witnessed successful applications in academia and industries. There still exist two limitations to the status quo, i.e., the lack of generality and the loss of probability information.

i) Lack of Generality:

In terms of existing regression methods and ANN methods, once the required format of probability is fixed, the complexity for producing this forecasting results will be simplified with the ad-hoc design of methodologies that only aims at the desired information of uncertainty rather than the whole probability distribution. For instance, using temperature scenarios collected as the model

inputs can generate a set of output realizations that roughly follows the probability distribution of uncertain load. However, the practicality and performance heavily rely on the availability of temperature data and similarity between the actual future temperature and historical temperatures that are considered as the scenarios. These ad-hoc methods can perform well in some cases, whilst cannot achieve fair accuracy in other cases.

ii) Loss of Probability Information:

For some prior methods for PLF, the expression of uncertainty may be fixed as quantiles, intervals, and so on. However, given a fixed form of uncertainty expression can largely reduce the complexity and computational costs, but may result in information loss due to the limitation of expression form of uncertainty. For instance, the quantiles of uncertain load can be estimated by replacing the original objective with pinball loss between the true value and forecasting value. With this model, the uncertainty information is stored in the form of quantiles, rather than the entire probability distribution.

According to the literature, none of the prior methods can tackle the two challenges simultaneously.

7.3 Modelling Load Uncertainty with Weather, Temporal and Social Features

As discussed in Chapter II, load uncertainty is actually the combination of multiple uncertain components. These components are caused by different uncertainty sources. Classified by its properties, some components are driven by systematic effects with external causes such as temperature, humidity, the day of the week, and so on. These components are named as epistemic uncertainty. While the rest components of uncertainty may be caused by unknown procedures or pure randomness, which is named as aleatory uncertainty. One difference between the two categories of uncertainty components is that epistemic uncertainty is capable to be modelled with sufficient prior knowledge and proper approaches, while aleatory uncertainty can only be quantified with probability distribution functions.

Following this concept, a fundamental idea for improving probabilistic load forecasting in this thesis is to model the epistemic uncertainty components, and find the relationship between the observations of model inputs and outputs. As discussed in chapter V, these outcomes can be achieved by inverse UQ process. Therefore, this thesis attempt to integrate the Deep Learning tool designed for inverse UQ problems into the framework of forecasting, to ultimately achieve superior performance on probabilistic load forecasting.

Considering weather conditions, temporal information and social information, a collection of historical electric load data at the household level can be assembled from open-sourced datasets and websites. A topology graph is presented in **Figure 7-1** to indicate how these external factors affect the load uncertainty.

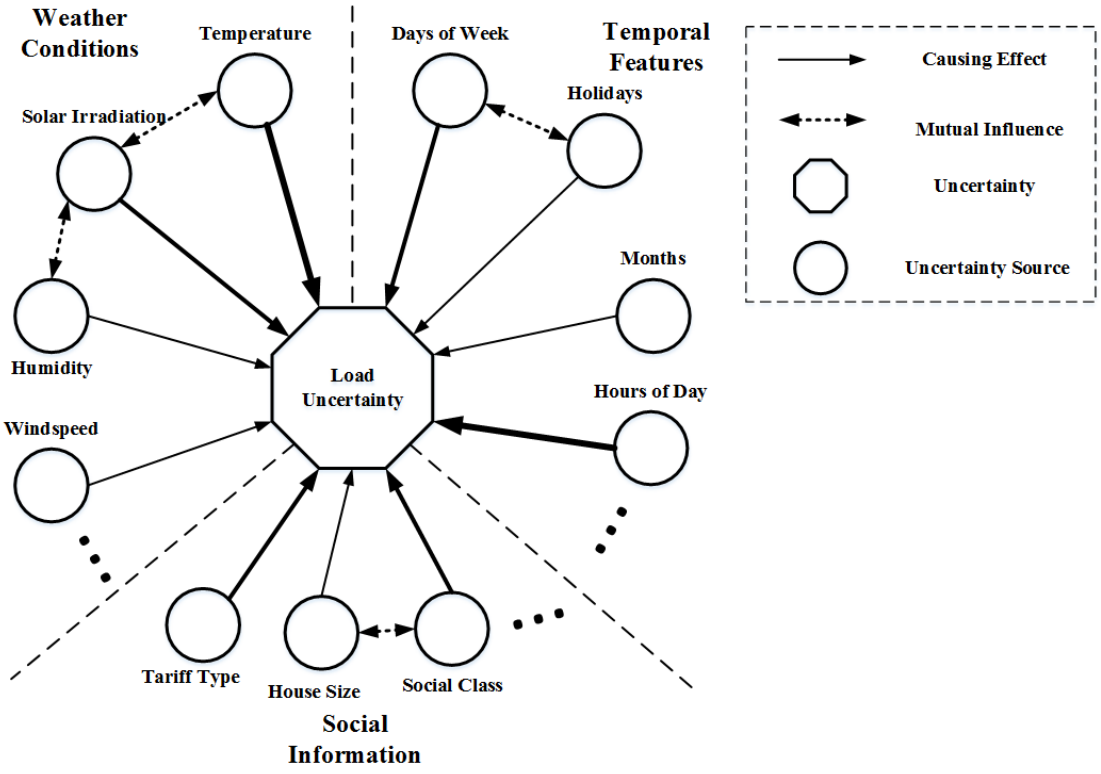


Figure 7-1 Demonstration of How Uncertainty Sources Generate Load Uncertainty

As shown in the figure, many external sources influence the load uncertainty, while mutual impact also exists between these sources. Therefore, it poses huge difficulty to model the effect of uncertainty source without discriminating the mutual effect between various sources. For example, temperature and solar radiation may all be highly correlated with PV outputs installed in households. However, solar radiation directly

affects the output power from PV panel, while the temperature is indirectly related to PV outputs due to its high correlation to the solar radiation. Hence, it is crucial to understand the mechanism between these external uncertainty sources.

Temperature is believed to be a major source of load uncertainty in the prior literature. Therefore, many works [13, 14, 15, 120] attempt to model the correlation between electric load profiles and temperature by different statistical methods, such as significance analysis, ANOVA, and so on. Literally, temperature mainly influences load uncertainty by changing the customers' behaviours. For instance, customers tend to use air conditioners if the temperature is higher than the thresholds and electric heaters may be used if the temperature is too low. Furthermore, temperature even can determine the time points for daily events, such as the starting time point for PV generates power, the time for making dinner, and so on. In the result demonstrated in Chapter V where the domestic demand in London area is taken as an example, higher temperature is related to lower electric demand. This phenomenon is driven by a figure of fact that the proportion of electric heater is much higher than the demand of air conditioner across the whole UK. As air condition may increase demand as temperature increase, the electric heater will obviously decrease demand with temperature increase.

Rainfall is another crucial weather condition factor that may affect the customer behaviours. As imagination, on rainy days, customers tend to stay at home, in particular at weekends. In addition, it may incentivize the usages of home appliances, such as wash machine, air dryer, and electric heater. Similarly, humidity is a factor that highly correlated with rainfall factors, and is expected to witness higher electric demand when humidity/rainfall increases.

Solar radiation is a key factor directly related to the generation of PV energy, where nearly 12700 MW capacity of PV panels installed across the whole UK by January of 2018 [151]. As expected, solar radiation will affect the wholesale energy price due to the cost of energy generating changes. However, due to the dynamic pricing of PV energy price is still developing, this factor does not affect the domestic demand much at present. However, it may become a crucial factor that affects the load uncertainty if dynamic PV pricing applied in the future SG. Similar to solar radiation, wind speed is another factor that also affects the load uncertainty indirectly through influence on

electricity generation costs. Therefore, the results shown in chapter V indicate the wind speed does not have significant influences on demand uncertainty.

Temporal features, including days of the week, hours of day and months of the year are the key factors that determine the periodical patterns of time-series data across different resolutions. Obviously, the daily, weekly, monthly and annually periodicity is proven in the past literature of LF [118-141]. Notably, in the UK, Friday is a special day when a huge peak in the noontime appears, due to the tradition that all schools and many homes will cook fish and chips. Alongside the widely known periodical patterns of domestic demand. According to the researches, the demand patterns in the special days are extremely different from the normal days due to the different customer behaviours.

In this thesis, these features are considered as the input feature and the historical data associated with demand samples are collected from the open-sourced dataset. For instance, to acquire the temperature feature of one household at London on 10th of Dec 2015, the temperature reading recorded on the same date by Hampstead weather station at north London, UK [102] is associated, since it is the nearest weather station to the customer location. After similar data pre-processing and manipulation. The formatting and representation of dataset are introduced in the experiment setting part.

7.4 Producing Probabilistic Forecasts with Inverse UQ Method

As discussed in former chapters, this thesis for the first time proposes to perform probabilistic load forecasting by modelling demand uncertainty with inverse UQ. This section will discuss the feasibility and validity of this proposed idea formally by following the mathematical notation used in Chapter V.

Assuming the uncertain electric load of interest is denoted as $y(t)$, and the associated input features (temperature, holiday, weekday, etc.) are manipulated into a N-dimensional input vector $x(t) = \{x_1(t), x_2(t), \dots x_N(t)\}$. In Chapter V, a tilde accent is added to the notation to replace the deterministic variables with random variables to clarify the investigated variables are associated with uncertainty. In this chapter, for mathematical simplicity, we omit the tilde accent and the time index t for input uncertainty source vector x and output demand uncertainty y. The rest of content will

prove how demand uncertainty can be learnt with inverse UQ. An illustration is given in **Figure 7-2**, to demonstration the comparison between the underlying physical model of demand uncertainty with the proposed inverse UQ (in Chapter V) that implemented as a Deep Learning model.

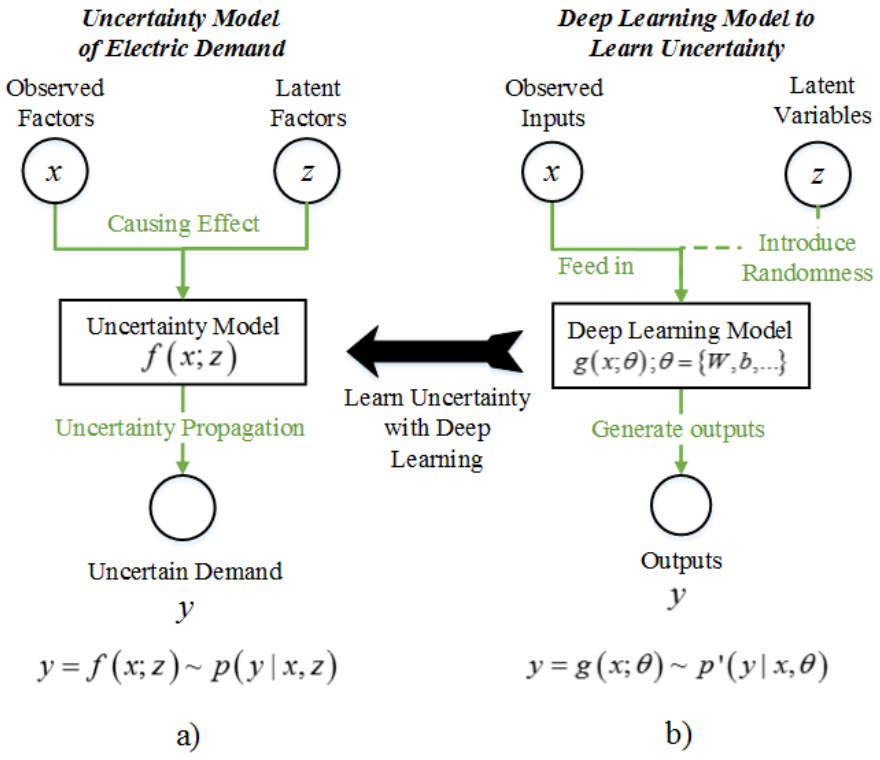


Figure 7-2 Comparison of Computational graphs of: a) Underlying Uncertainty Model, and b) Deep Learning Model, on Load Uncertainty

Although in the practical implementation, the input features could consider as much information as possible if the data is available. However, the demand uncertainty is produced through fair complicated procedures with contributions from many sources that can far beyond our expectation. Therefore, some unobservable uncertainty sources are not included in the input feature x. A latent uncertainty source vector $z = \{z_1, z_2, \dots z_N\}$ is introduced to represent the effect from unknown sources and pure randomness. The uncertainty model of demand uncertainty can be formulated as: $y = f(x; z)$ where the function or functional $f(*)$ can be used to represents the underlying physical model. The demand observation will follow probability distribution $y = f(x; z) \sim p(y|x; z)$.

However, when these observations are attempted to be estimated with uncertainty model featured with model parameter set θ, the mechanism of this uncertainty model can be presented as $y = g(x;\theta)$, where the synthesized function $g(*)$ represents the function of the uncertainty model. Specifically, for deep learning model, the model parameters are determined by weight matrices, bias matrices and other specific parameters associated with the certain network architecture. i.e., $\theta = \{W, b, ...\}$. Load uncertainty can be then rewritten as $y = g(x;\theta) \sim p'(y|x;\theta)$.

As a fact, the underlying physical system behind electric demand is believed to be a complex system without closed form representation. Although the function $f(*)$ is too complex to formulate, but we can use the uncertainty model $g(*)$ to approximate it by following the equation $\lim_{\theta \mapsto \theta^*} p'(y|x;\theta) = p(y|x;z)$, where θ^* is the well-trained parameter set of the uncertainty model.

Assuming the historical datasets contains the data samples of x and y, which is noted as X, Y. These data samples are used as a training dataset for the deep learning model. As proven in Chapter V and Appendix A. The posterior distribution of model parameter $p'(\theta|X,Y)$ can be approximated as variational functional $q(\theta)$, which is inferred by Variational Inference [47, 48, 87] method. Through optimization that maximizes the Evidence Lower Bound (ELBO) [47, 48, 91] term, the solution of deep learning model $q^*(\theta)$ can result in minimal KL divergence between $q^*(\theta)$ and the underlying uncertainty model $p'(\theta)$.

For any future input feature x^*, the posterior distribution of output forecasts y^* can then be formulated as:

$$p(y^*|x^*;z,X,Y) = p'(y^*|x^*;\theta,X,Y) = \int p'(y^*|x^*,\theta) q^*(\theta)\, d\theta \approx q^*(y^*|x^*,\theta) \qquad (7\text{-}1)$$

where proves that the realization of demand uncertainty y^* will follows the probability distribution of the deep learning model output. i.e., $y^* \sim q^*(y^*|x^*,\theta)$. Since the solution of variational functional q^* corresponds to the output distribution of fine-tuned deep learning model. In other words, by perform feed-forward operations on the fine-tuned deep learning model, we can generate a collection of observations on the target demand

uncertainty y^*. Upon this step, we can produce any form of forecasting results, including probability distribution, quantiles, intervals and so on, with sufficient volume of realization size.

7.5 Implementing Probabilistic Load Forecasting with Deep Learning Model

As introduced in Chapter V, a dropout Stochastic Regularization Technique (SRTs) [87, 93, 96] is deployed to introduce random properties to the model parameter θ. The validity and efficacy of this approach is already discussed and proved in Chapter V and Appendix B, hence the relevant content will be skipped in this section. Instead, this section will more focus on how to integrate proposed PDDGN model into the forecasting framework. Firstly, the flowchart of the proposed PLF framework is illustrated in following **Figure 7-3**:

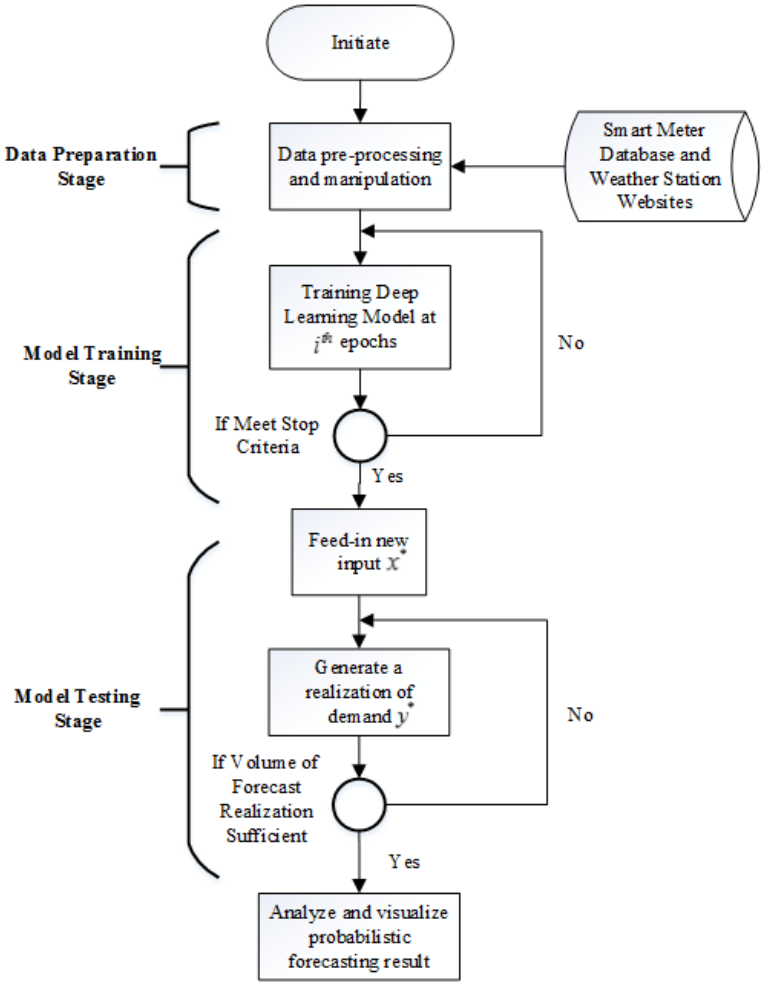


Figure 7-3 Flowchart of Proposed Deep Learning Model for Probabilistic Load Forecasting

As shown in the flowchart, the procedure of PLF framework consists of three stages sequentially: i.e., data preparation stage, model training stage and model testing stage.

7.5.1 Data preparation stage: pre-processing and manipulation

Through the data pre-processing and manipulation stage, data frames from all diverse sources will be arranged together in the form of model input and output pairs. In other words, each pair consists of a demand uncertainty observation and its associated observation vector on uncertainty sources exactly at the same time point. Notably, the weather information is collected from weather station websites, and the actual forecasting error on future weather is not considered in this work. The rationale is this work focuses on how to improve PLF accuracy by modelling uncertainty components composition, while how to mitigate forecasting error lies in the input information is a different but crucial research direction and could be investigated in the future work.

7.5.2 Model Training Stage

In the model training stage, the deep learning model will sample training input and output pairs from the dataset formed in the previous stage. As shown in **Figure 7-4**, for i^{th} training epochs, a training data batch $< X_i, Y_i >$ with batch size B will be sampled from input output pair dataset. The training input batch X_i will be transformed to feed-in into the deep learning model, while the training output batch Y_i will also be transformed to corresponds to the training input. For same index k within the batch, the input data frame $X_i(k)$ and output data frame $Y_i(k)$ should be exactly match.

As shown in the, at the beginning of i^{th} epochs, the first step is to realize the deep learning random units into deterministic values. Assuming the model parameter set θ consists of deterministic parameters W and probabilistic parameters ω. Through this step, the probabilistic parameters ω will be replaced by its realization $\hat{\omega}$. Hence the deep learning model will become a deterministic deep neural network within i^{th} epochs. The second step at i^{th} epochs is to perform a feed-forward operation by feed-in the training input batch X_i. Through the deterministic model, a batch of model output O_i can be generated accordingly. Last but not least, next step is to perform back-propagation based on the minimization of the loss between model output batch O_i and

the training output batch Y_i. Normally, this loss is formulated with $L_2 - norm$ between O_i and Y_i. In the back-propagation, the deterministic parameters W will be updated.

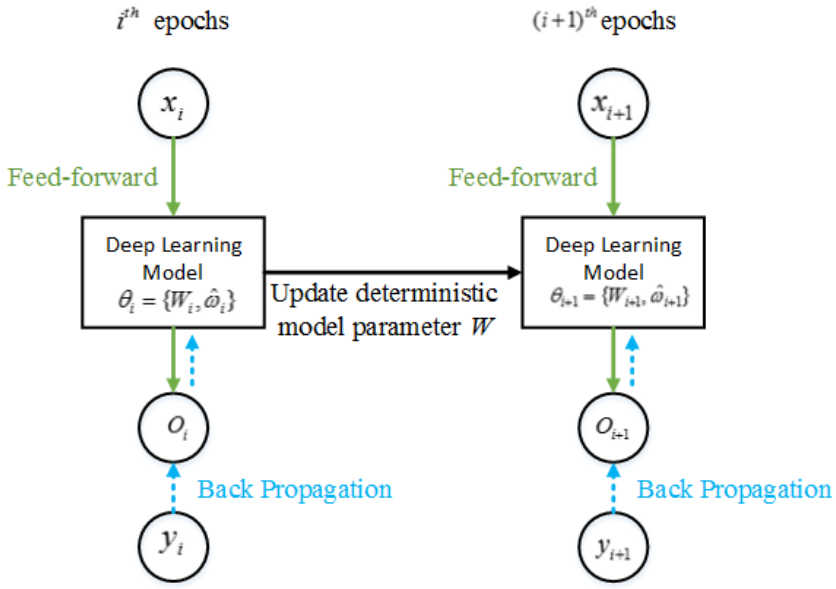


Figure 7-4 The computational graph and unfolded topological graph in two continuous epochs at the training stage

7.5.3 Model Testing Stage

After fine-tuning the deep learning model, the synthesized function of this model $q^*(y^*|x^*,\theta)$ is equivalent to the posterior distribution of uncertainty model $p(y^*|x^*;z,X,Y)$, with acceptable tolerance. Hence, this model can now be used to forecasting new uncertainty demand of interest y^* given estimation on new input feature vector x^*. The procedure of testing stage is demonstrated in **Figure 7-5**:

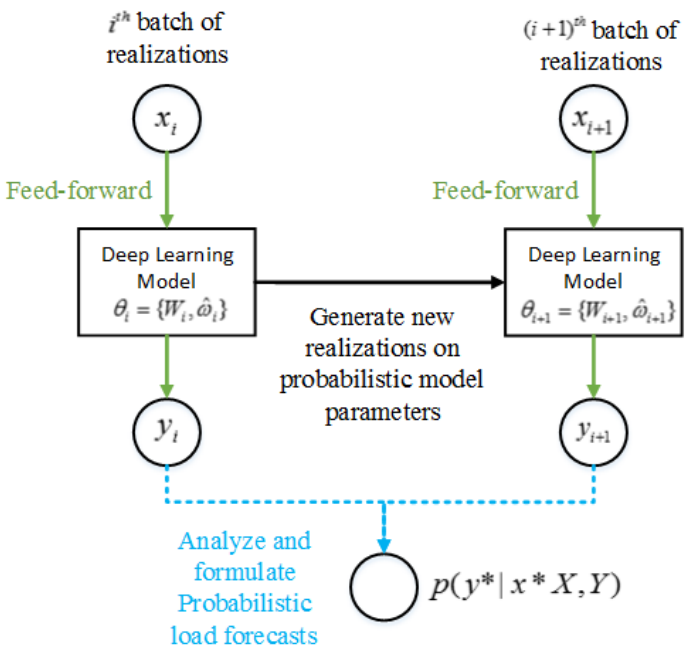


Figure 7-5 The computational graph and unfolded topological graph in two continuous epochs at the testing stage

As shown in the figure, testing step also consists of many epochs, since the probabilistic parameters ω of the deep learning model are required to generation sufficient realizations to maintain its probabilistic property. For i^{th} epochs, the first stage is to generate a model realization by replacing the probabilistic parameters ω with its realization $\hat{\omega}$. The second stage is to feed-in the inputs vector ${x^*}_i$ and perform a feed-forward operation to get model output O_i. Equivalently, model output O_i can be regarded as a realization of target uncertain demand y^*, i.e., $O_i = \widehat{{y^*}_i}$. By collection demand realization over sufficient epochs, the probabilistic forecasting on y^* can be obtained in any form of uncertainty representation. For instance, if we only interested in expectation of y^*, the expected $\overline{y^*}$ can be estimated with MC estimator $\frac{1}{N}\sum_{i=1}^{N} O_i$. The expected outcomes include intervals, quantiles, and even a complete probability distribution of y^*.

7.6 Experiment Settings

7.6.1 Data Description

In this work, three categories of data are collected to form the demonstration dataset. For demand data, the realistic smart metering data are collected from the whole London

area from Low Carbon London (LCL) project. This dataset commissioned over 5000 local domestic households across the inner London, during the period from November 2011 to February 2014. Notably, these customers are all unrestricted domestic customers, which means none smart grid techniques and incentives are involved with the customer and affect the energy consumption behaviours of the customers. In other words, this smart metering dataset purely maintains the load profiles patterns of domestic users that can reflect the features of domestic customers across the whole UK.

Weather information in this work is collected from the website of Hampstead weather station [102]. This weather station is a meteorological observation site located near Hampstead, in North London, UK. The site was established with an automatic personal weather station since July 2010 without break. The updating frame is as quick as one minute. In terms of the historical weather data, hourly recordings are available since it established and open-sourced to the academia.

Temporal information is auto-generated with Python default package and functionalities. The holiday information, in particular, the Bank Holiday [152] information is collected from the dataset published by UK government, available at website https://www.gov.uk/bank-holidays.

7.6.2 Parameter Settings

The parameters and hyper-parameters used in training and testing processes are concluded in **Table 7-1**:

Table 7-1 Parameters Settings in Training and Testing Stage

Parameters	***Unit***
Net Layer Number	4 (Layers)
Training Epochs	100k
Testing Epochs	20k
Early Stopping Threshold	5.0E-5
Batch Size	1024
Input Dimension	48
Output Dimension	1
Hidden Size Each Layer	{128,128,128,64}
Dropout Rate	0.83
Optimization Algorithm	AdamOptimizer (Adaptive Moment Estimation)
Learning Rate	1.0E-3
Adam Hyper-Parameter Beta1	0.8
Adam Hyper-Parameter Beta2	0.7

7.6.3 Benchmarks

In order to validate proposed method, a vanilla Multiple Linear Quantile Regression (MLQR) is introduced as the benchmark methodology. MLR is a well-developed approach used in prior PLF researches. T Hong, et al, in their works [140], deployed MLQR to perform long-term load forecasting in monthly scale for zonal aggregated energy consumption. Work [37, 105, 153], uses MLQR to approximate the quantiles of load profile uncertainty to generate interval forecasts. It integrates Multiple Linear Regression (MLR) with Quantile Regression (QR) to perform probabilistic load forecasting with the theoretical basis of MLR method. MLR and its meta methods have witnessed great success in the Global Energy Forecasting Competition 2014 [130], and took all of the Top 6 places in the long-term energy forecasting and price forecasting problem.

The Average Pinball Loss [104, 105, 106] is employed in this work to assess the accuracy of PLF. Assume the quantile of interest is represented as a fraction within the interval [0,1]. Denote the target quantile as τ, y is the aimed real value to be forecasted, and the forecasting output of the model is o. The pinball loss L can be formulated as:

$$(y, o) = \begin{cases} \tau(y - o) & if\ y > o \\ (1 - \tau)(o - y) & if\ o > y \end{cases} \quad (7\text{-}2)$$

If the pinball loss can be minimized to approach zero, the forecasting output o is expected to be exactly the target quantile.

In this thesis, The average quantile error with respect to 9 different quantiles $\{0.1, 0.2, 0.3, \ldots, 0.9\}$ is used for the evaluation. i.e.,

$$L_{target} = \frac{1}{9}\sum_{n=1}^{N}\sum_{\tau\in\{0.1,0.2,\ldots,0.9\}} L(y_n, o_n, \tau) \quad (7\text{-}3)$$

7.7 Demonstration

In this section, a performance evaluation is presented on proposed deep learning model for probabilistic load forecasting. Two parts of the content are presented in the demonstration: the results of probabilistic load forecasting are visualized and benchmarked at first. A discussion is followed to explain why the proposed method can achieve better performance than prior state-of-the-art for domestic load forecasting.

7.7.1 Result Visualization and Benchmarking

In order to clearly demonstrate and visualize the forecasting result, cases of one household during a continuous week period is present. The comparison is made between the vanilla MLQR method and the proposed PDDGN forecasting model. The result is presented in figure 7-6 and 7-7.

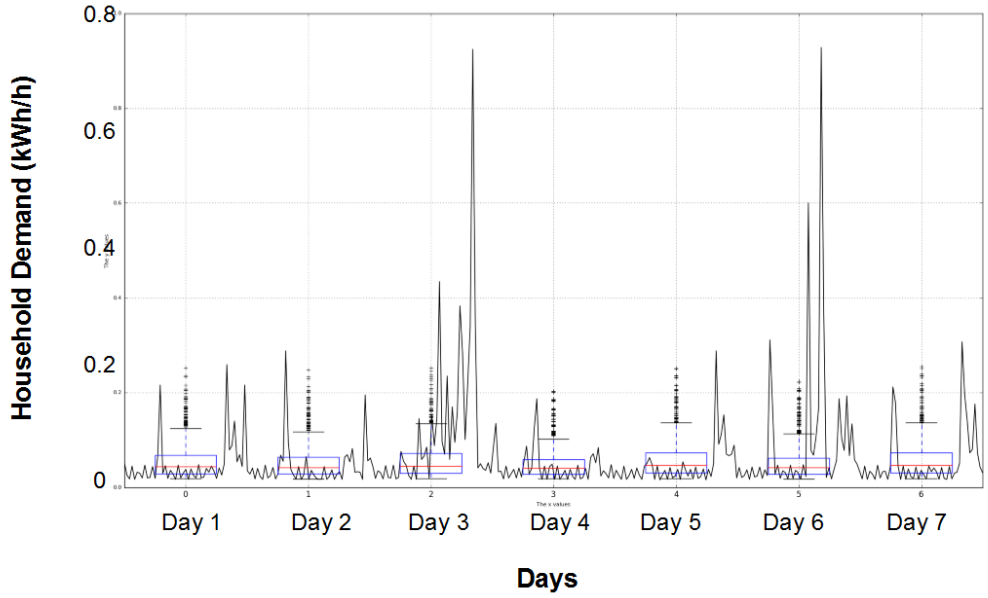


Figure 7-6 Weekly quantile forecasts with vanilla MLQR method

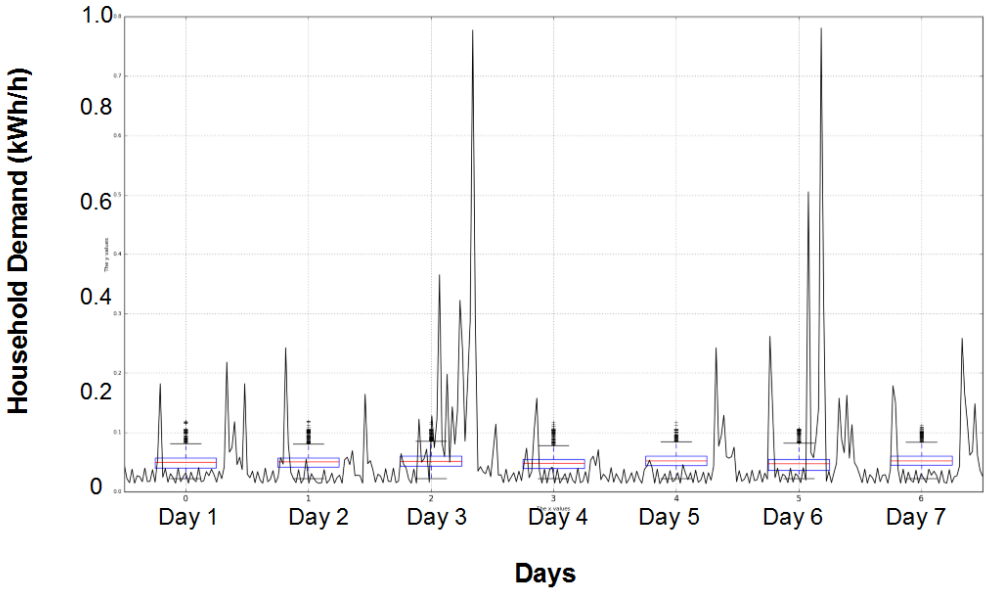


Figure 7-7 Weekly quantile forecasts with proposed PDDGN method

As shown in the figures, the vanilla MLQR method and proposed method are all capable to forecast the uncertain demand in the form of uncertainty quantiles. Comparably, vanilla MLQR method tends to forecast more conservatively, while the proposed method aggressively converges to a narrow band within same quantile ranges. In addition, the proposed method has a very short tail on the probability distribution while vanilla MLQR method has longer tail towards the upper bound of demand uncertainty.

Measured with Pinball loss across 1000 households, the average pinball loss between vanilla MLQR method and proposed PDDGN method is presented in **Table 7-2**.

Table 7-2 Comparison of Average Pinball Loss between Benchmarking method and Proposed Method Across 5000 Customers

Quantiles(%)	*Vanilla MLQR method*	*Proposed PDDGN method*
10	7.564	10.124
20	14.878	15.458
30	21.233	19.958
40	26.513	23.743
50	30.924	26.707
60	34.325	28.828
70	36.834	30.064
80	37.077	30.315
90	31.165	29.258
Average	**26.724**	**23.828**

As shown in the table, the proposed method can achieve 12% lower forecasting error in terms of pinball loss across all customers. Notably, for quantile larger than 20%, the proposed method can achieve better performance than vanilla MLQR stably. However, for 10% and 20% quantile, the proposed method performs worse.

7.7.2 Result Discussion

This discussion will include three parts of ideas: i) why proposed method can make more aggressive forecasting; ii) why proposed method deteriorated at lower quantile; and iii) why proposed method can achieve better performance averagely than prior state-of-the-art.

At first, the essential difference between vanilla MLQR and the proposed method is that the state-of-the-art assumes linear relationships between uncertainty and its causing

sources, while proposed method respects the fact of its non-linearity and mutual influence. In this sense, the state-of-the-art will overestimate the combined contribution to the uncertainty through uncertainty propagation. To match the historical demand uncertainty, it will moderately adjust its coefficients, since the variations of coefficient will over impact the model outputs. Ultimately, the forecasting results tend to be conservative. In vice versa, the proposed method tries its best to model the complex non-linearity and mutual influence, hence it performs relatively aggressive in predicting future demand.

Secondly, the proposed method outperforms the vanilla MLQR method for most quantiles except for 10%. Due to the specific features of electric load that the demand often decreases to around 0 when residents leave the property. Therefore, the 10% quantiles are close to 0. In terms of the proposed PDDGN method, it learns the uncertainty and thus can make aggressive predictions, whilst vanilla MLQR method is more conservative comparably. For 10% quantiles, the conservative predictions by vanilla MLQR method are broad enough to contain the very low demand, whilst PDDGN method is too aggressive. Therefore, PDDGN method outperforms vanilla MLQR with respect of average quantiles from 10% to 90% in overall, but witnessed worse prediction results specifically for 10% quantiles.

During the demonstrated week, the demand valley is lower than the average. Since vanilla MLQR method performs more conservative, the lower quantiles can cover the true values better. Hence, the benchmarking method outperforms the proposed method at lower quantiles during this period of data.

Last but not least, the vanilla MLQR overestimates the uncertainty, and cannot forecast narrow quantile band since its modelling capability is constrained within the linear space. However, the proposed method can model the latent influence caused by non-linearity and coupling effect of uncertainty sources, hence can make better probabilistic forecasting result as expected.

7.8 Chapter Summary

Extending the idea of Chapter VI that to improve forecasting accuracy by modelling epistemic uncertainty components, this chapter attempt to further improve the

performance by considering the uncertainty sources and model the relationship between uncertainty components and corresponding sources. An application is demonstrated in probabilistic load forecasting.

By developing advanced probabilistic deep learning model based on Bayesian Inference, the epistemic uncertainty components caused by weather conditions and temporal features are hence modelled and quantified. The efficacy of proposed forecasting method is tested on domestic load profiles. 12% of performance improvement in Pinball loss are reported from the results.

According to the results of both Chapter VI and this Chapter, an evidence is provided that load forecasting can be further improved if the load uncertainty epistemic components can be modelled correctly.

Chapter 8

Conclusions and Future Works

THIS chapter summarises the conclusion of former chapters and further discusses the future work of this thesis.

8.1 Conclusions

Nowadays, the increasing uncertainty has become a major challenge to the existing applications and techniques to the smart homes and smart grids. However, prior works in quantifying and dealing with the uncertainties are only focused on forward uncertainty quantification that provides the numerical quantification of the uncertainty of interest. In practice, smart grid application may need more detail information such as the temporal, spatial characteristics, the uncertainty component composition, and causing sources that contribute to uncertainty components. It calls for researches that can provide the deeper insights for understanding the uncertainty. Fortunately, inverse uncertainty quantification problem, which is widely recognised by many engineering fields, is the natural tool for the desired purpose. This thesis attempts to present breakthrough both to the fundamental science and the applications at Energy Management System (EMS) and Load Forecasting (LF).

To the fundamental sciences, two contributions are proposed by this thesis:

i) For forward UQ problem, this thesis further extends the forward UQ approach into a multi-time-scale version. The proposed method can not only provide efficient numerical quantification that already acquired by traditional forward UQ approaches, but also can characterise the temporal features of the uncertainty of interest across multiple time scales.

ii) To fill the research gaps in inverse UQ problem in the power community, this thesis proposes a new breed of deep learning model to implement the inverse UQ functionality. To ensure the generality and efficacy of proposed approach, the methodology is designed purely in a data-driven fashion, which does not rely on the prior knowledge to the system.

Along with the contributions to the fundamental science, this thesis also demonstrated examples of smart homes and smart grid applications, to indicate how to further improve the existing techniques and applications with understandings of uncertainty properties and features. In details, two applications are taken as examples:

i) For Energy Management System, this thesis presented an application that attempts to achieve grid level goal through purely decentralised control. The benefits of decentralised design lie in the savings of economic and operational cost compared to centralised or distributed fashion. The challenge is how to coordinate between individual households to achieve grid level goal without any information exchange for coordination. Hence, this application takes advantage of a property of uncertainty propagation model, i.e., lower coupling degree between input uncertainties will result in less propagated uncertainty.

ii) For Load Forecasting, two applications are presented for point load forecasting and probabilistic load forecasting accordingly. For point load forecasting, the idea is to learn the uncertainties that are commonly shared by neighbours to assist with the forecasting task. For probabilistic load forecasting, the fundamental idea is to integrate the UQ method proposed for inverse UQ problem to model the uncertainty components and associate the underlying variation with external information such as weather, time of the day, etc. Both approaches have witnessed the superior performance in dealing with uncertain household electric load compared to traditional methods.

The results and outcomes of this thesis can be concluded in four aspects accordingly:

8.1.1 Multi-time-scale Forward Uncertainty Quantification for Household Load

In order to characterize the temporal features of household load uncertainty, this thesis further extends the forward UQ into multiple time scales. The breakthroughs in techniques are:

- A generalised multi-time-scale forward UQ method is proposed by integrating decomposition methods into the process of forward UQ framework to achieve individually quantification to each uncertainty components allocated at differing time scales.

By demonstrating on the realistic smart metering data collected from Irish domestic household nationwide, the following results are indicated:

- According to the result, some temporal characteristics of demand uncertainties can be summarized as: i) demand uncertainties is roughly proportional to the average demand consumption; ii) demand uncertainty has two peaks, a higher one in the evening between 16:00 p.m. - 19:30 p.m., and the daytime peak is between 9:00 a.m. - 14:00 p.m.; iii) uncertainty peak emerges on any of Wednesdays, Thursdays and Winter season. Analysis of cross-channels indicates: i) inter-days uncertainty have stronger long-term dependencies than short-term dependencies; ii) intra-day uncertainty have stronger dependency at hourly time-scales.

- In terms of the uncertainty propagation from disaggregated level to aggregated levels, a mathematic model is proposed based on CLT theorems to explain the uncertainty mitigation phenomenon. The proposed model is inconsistent to the demonstration results: i) disaggregated uncertainties can mitigate at the aggregated level, the mitigated rate can approach a limit with sufficient aggregation scale. Specifically, 200 households are sufficient for enabling 9o% of the uncertainty mitigation effect; ii) the potential UPF limit is positively related to the average PCCs. Specifically, half-hourly and daily time scales have the most UPF limits at around 90%.

In summary, this part of research demonstrates the underlying temporal features that lie in different time scales of the original data are quite different from the feature of the complete uncertainty. In the practical application, treating these uncertainty components at different time scales can better mitigate the adverse impact of the uncertainty.

8.1.2 Inverse Uncertainty Modelling for Household Load based on Deep Learning Techniques

In order to fill the research gaps, this thesis for the first time investigates and implements inverse uncertainty quantification for the power system problems. The

research work investigates the uncertainty of electric load at the household level, which is contributed by known external sources, such as weather, measurement error, time of day, the day of weeks, etc. and other unknown factors and unpredictable pure randomness. The purpose of this research is to model the uncertainty components caused by known external sources with relevant information collected from open-sourced dataset and websites of weather stations. The breakthroughs in techniques are:

- First time to tackle the inverse UQ problem, which is complementary analysis to the forward UQ. Specifically, it investigates the underlying properties of uncertainty that are not presented in forward UQ problems. The key theoretical basis for supporting the feasibility for performing inverse UQ in the power system is the Bayesian Inference Theory, that aims at the posterior distribution of underlying parameters of the uncertainty model on condition of historical data samples. It provides the mathematical tool to formulate the underlying physical model by training on the historical data.

- An advanced Deep Learning model, named as Probabilistic Deep Dropout Generative Nets (PDDGN) is implemented for inverse UQ quantification. Compared to ad-hoc methods that are specifically designed for certain problems, this deep learning model can be generalised to any inverse UQ problem in the power system once feed in the required data. In addition, by feed-in input features (information of external sources), the model can also generate probabilistic realizations of the uncertain load, which can be regarded as the sampling from load uncertainty distribution. This attribute makes this model a natural tool for performing probabilistic load forecasting.

The proposed approach is demonstrated on the smart metering dataset of Irish domestic households as well. The result indicates the proposed approach can: i) segment the mutual influence between different uncertainty sources by constructing the union posterior distribution of demand uncertainty with respect to multiple causing sources; ii) model the impact from any given uncertainty sources as input features, by integrating to quantify the marginal posterior distribution of the uncertainty with respect to the source of interest. The presented research also provides pilot experiment results to

indicate the future research of inverse uncertainty quantification in the power system.

8.1.3 Decentralised Energy Management System to Achieve Grid Level goal by Decoupling Uncertainty Correlations

By learning the understandings from former research works in this thesis, an advanced home Energy Management System is developed to achieve grid level goals. Since the grid level goals, such as peak reduction and uncertainty reduction is the synthesized effect by coordinating across many individual households. In traditional works, this aim requires centralised optimization scheme and complex programming model to coordinate a large number of customers involved and suffers from the disaster of dimensionality and complexity in information exchange framework. In addition, the practical implementation will further consider the investment in communication devices. Therefore, the realization for grid-level goals through home EMS is expensive both in economic and operational cost. To tackle these challenges, the breakthroughs on techniques in this thesis are:

- A natural underlying property of uncertainty is introduced in the work as the theoretical basis for achieving grid level goal in a decentralised fashion. In other words, lower coupling degree of inputs uncertainties will result in less propagated output uncertainty in a typical uncertainty propagation model of forward UQ problem.

- An advanced home EMS system is developed to decouple correlations between uncertain load at individual households purely without any information exchanges for coordination. A novel technique named as Wavelet Auto-decoupling Regularization (WAR) is proposed for the purpose, which integrates the wavelet decomposition into the EMS programming scheme, and achieves the decoupling requirement by using Wavelet Decomposition.

The proposed technique demonstrated on hundreds of smart homes with battery storages installed and inherent electric load monitored by smart meters. The demonstrated results indicate the proposed method can achieve: i) 8% of energy cost

savings for individual households, ii) 18% daily peak reductions at grids, iii) 30% annual peak reduction at grids, iv) and 45% uncertainty reduction at grids.

8.1.4 Point and Probabilistic Load Forecasting by Modelling Load Uncertainty with Deep Learning

The major challenge of load forecasting at the household level is the significant uncertainty and volatility caused by a variety of factors, such as weather, customer behaviours, and inherent variations of appliances. As discussed in former chapters, these factors will contribute to either epistemic uncertainty or aleatory uncertainty in the load pattern. While aleatory uncertainty components are unpredictable, epistemic uncertainty can be mitigated or predicted with proper ways, which gives the potential to improve household load forecasting.

According to the literature of Traditional LF methods, three ways are deployed to improve the performance against uncertainty, i.e.: i) clustering techniques to group similar customer, ii) predict aggregated load profiles; and iii) filter out uncertainty with data pre-processing techniques. These methods, attempts to present better performances by performing load forecasting on less uncertain dataset rather than directly model the uncertainty of load. However, these datasets are not the original data to be predicted, hence these methods cannot improve the load forecasting performance in the original dataset of interest.

However, this thesis aims to improve the performance of load forecasting from a novel idea: to learn the epistemic uncertainty. The technical breakthroughs are:

- This thesis for the first time integrates inverse UQ technique with load forecasting techniques, which aims to improve load forecasting accuracy directly by modelling the epistemic uncertainty. By deploying the inverse UQ modelling method implemented in Deep Learning models, the well-trained Deep Learning model can generate collections of forecasts realizations sampled from its uncertainty distribution.
- Two Deep Learning models are implemented for point forecasting and probabilistic

forecasting accordingly. It integrates promising techniques in Deep Learning community such as: Deep Recurrent Neural Nets (DRNN), Long Short-term Memory units (LSTM), Dropout Stochastic Regularization Techniques (DSRT), and Bayesian Variational Inference. In details, DRNN is an advanced deep learning architecture designed to extract the temporal features of time-series data; LSTM is prototype of neuron units in deep learning that can simulate the memory feeding away progress as human does, to help the simulation of time-series process; DSRT is a well-developed technique to introduce random attributes to the deep learning model, originally to prevent the ill-posed solution in overfitting problems [87, 96, 97], but is used to achieve Bayesian Variation Inference in the deep learning training framework. Last but not least, the Bayesian Variational Inference is the theoretical basis to perform inverse UQ with machine learning models. This thesis successfully integrates these cutting-edge techniques to largely increase the forecasting accuracy by 20% for household load point forecasting and 12% for household load probabilistic forecasting compared to relevant state-of-the-art techniques in each area.

8.2 Future Works

8.2.1 Extending the Fundamental Research on Uncertainty Quantification

Currently, two parts of works are delivered regarding the fundamental science of uncertainty quantification. On the one hand, the typical forward uncertainty quantification framework is generalised by this thesis into multi-time-scale version, which can provide detail information on the temporal characteristics of uncertainty across different time scale channels. On the other hand, the research gap in inverse uncertainty quantification is firstly filled by this thesis. A deep learning model hence developed to implement data-driven inverse uncertainty quantification based on Bayesian Inference, dropout Stochastic Regularization Techniques (SRT), and Deep Recurrent Neural Nets (DRNN).

In terms of the research work on multi-time-scale forward uncertainty quantification, future works can be further extended and improved in following directions:

i) **Spatial Features of Uncertainty**: the current approach is developed to analyse the temporal features and decompose the uncertainty into multiple time scale channels. This approach can be further improved to extract the spatial features of uncertainty, to understand how the uncertainty of electric load is affected by geographical information, especially the common features shared by neighbourhoods under the same substation. The result will be constructive to the planning and operation of substations in dealing with the inherent demand uncertainty with flexible resources.

ii) **Demonstration of other uncertainty components of interest in the power system**: In this thesis, the proposed multi-time-scale uncertainty quantification framework is demonstrated in the case of electric loads at domestic households. Future works can further apply the proposed method to analyse uncertainties of other crucial but uncertain components in the power system. For instance, Electric Vehicle (EV) as a novel low carbon and flexible resources, is of immense significance to many smart grid applications. However, its demand uncertainty model is not only heavily relied on temporal correlations but also largely influenced by spatial correlations, such as the correlation between the home address and working places. Therefore, the proposed approach can provide a natural tool to segment dependencies between different geographical locations and distances.

iii) **How to generalise the results with limited data and visibility**: As discussed in the thesis, the proposed approaches and investigated applications are assumed in the Big Data Era with sufficient data volumes and visibility. Currently, however, most distribution networks are still suffering from the lack of visibility. This issue may further last for decades. Therefore, this poses huge challenges in directly employing well-developed approaches from other disciplines. This thesis considers how to maximise the value of smart metering data available in some pilot projects, for both

fundamental science and potential smart grid applications. However, the data volume used in the analysis is still beyond the status quo. How to further improve and develop approaches to adapt to the lack of visibility and data volume is an urgent need.

As to works on inverse uncertainty quantification with deep learning model, future works can be conducted in following aspects:

i) **Modelling the relationship between social profiles and demand uncertainty**: Continue with the work presented the thesis, which performs inverse uncertainty quantification to model the relationship between the demand uncertainty and some external sources, especially temperature information, the future work can further complete the scenario by introduced social profiles as new input features. The result can be deeply explained together with the clustering result of social classes. The expected outcomes are: i) to understand how the uncertainty of load profiles varies along with different social groups, to help the tariff setting for network operators; ii) and to further improve the accuracy for uncertainty components modelling, and ultimately improve related smart grid applications, such as household load forecasting.

ii) **Exploring and Assessing Other Deep Learning Architecture and alternative approaches**: On the one hand, the proposed method is integrated into a simple deep fully-connected neural network and deep recurrent neural network. Yet, more architecture can be investigated to explore the possibility and potential of deep learning in dealing with uncertainty quantification. On the other hand, alongside with the Deep Learning techniques that are already well-developed for training with stochastic steps, alternative methods could be explored to achieve inverse uncertainty quantification.

8.2.2 Improving the Smart Grid Applications

In terms of the applications in smart homes and smart grids, the examples given in the thesis are covered by two research scenarios, i.e., energy management system and load forecasting. However, the significance of uncertainty research is not limited to the example. Besides on the energy management system and load forecasting, the potential benefits by making use of uncertainty properties can be extended to the smart grids applications both in depth and width. Some potential research ideas in the future works are discussed as follows:

i) **EMS improved with Robust Optimization considering Extreme Value Distribution:** Considering the forecasting error in the input data of realistic energy management system, such as the forecasted demand, and forecasted electricity price, etc., robust optimization is a widely recognized approach for operating the system whilst avoid excessive financial risks caused by uncertainty. Traditional robust optimization either form up possible scenarios from historical data samples as uncertainty set, or directly formulate the quantiles from data samples. However, estimating the boundaries, or quantiles from the finite volume of data samples are not an easy work, since it cannot directly reflect the actual and unbiased boundary and quantiles of the underlying uncertainty distribution.

Extreme Value Distribution theory indicates the extreme value of a set of data samples that sampled from arbitrary distribution functions, will strictly follow three general extreme value distribution types. In this sense, the distribution of the extreme value of a set of data samples can be estimated accurately and rapidly with higher convergence rate.

ii) **Further improve Load Forecasting performance consider more input features and under different forecasting scenarios:** According to the investigation in Chapter VI and VII, load forecasting performance can be improved if the load epistemic uncertainties are modelled accurately. Following this hint, future works can be done from two angles: i) further enlarge the uncertainty model by considering more features that may have

significant impact on load uncertainty; ii) currently, the evidence and investigation is not sufficient, since the demonstration is only implemented on the household level, and for short-term forecasting. More forecasting scenarios could be investigated.

iii) **Shares of stakeholders for uncertainty information in future networks:** The significances and potentials of uncertainty information and its analysis are extensively discussed in former content. A considerable problem is how to determine the shares of stakeholders of available uncertainty informations and benefits by integrating uncertainty modelling in existing applications. Whether these informations are holded by network operators centrally or these informations are owned distributely by individual participants in future energy markets that deregulates the ownership of network assets andd information, such as Peer-to-peer energy market? As expected by the author, applications and information of uncertainty will be centrally owned by network operators or distributly owned by third-parties such as aggregators, individual customers may not be involved due to the circumstance of modern energy markets at this stage. However, in the future, the stakeholder share should be more decentralised to enable the initiative of market participants. Potentially, Block Chain may be the key technique to the implementation of this future work.

Appendix. A

Discrete Wavelet Transform with Haar Expansions

Wavelet Transforms (WT) is one of the most popular techniques in time-frequency transformations during the last decades. Haar wavelets transforms (HWT) are one of the typical WT, which uses constant step function basis. Compared to other smooth wavelets, Haar wavelets have the nature of discontinuity. This nature makes Haar wavelets the most appropriate method to approximate the electric demand, which contains considerable variance, spikes and steep changes. Assuming the unit indicator function, which can be denoted as:

$$I_{[0,1)}(t) = \begin{cases} 1, & 0 \le t < 1, \\ 0 & otherwise, \end{cases}$$

The scaling functions of this indicator functions are:

$$\phi_k^{(j)}(t) = 2^{\frac{j}{2}}\phi\left(2^j t - k\right), j \in \{\mathbb{z} \ge 0\}, \}, k \in [0, 2^j - 1] \cap \mathbb{z}$$

Where the L_2 norm $\left\|\phi_k^{(j)}\right\|$ is equivalent to 1. The inner product of two scaling functions are:

$$\langle \phi_k^{(j)}(t), \phi_m^{(l)}(t) \rangle = \int_{-\infty}^{\infty} \phi_k^{(j)}(t)\phi_m^{(l)}(t)dt = \delta_{j,l}\delta_{k,m}$$

Accordingly, the Haar mother wavelets is defined as:

$$\Psi(\mathrm{t}) = \begin{cases} 1, & 0 \le t < 1/2, \\ -1, & 1/2 \le t < 1, \\ 0 & otherwise \end{cases}$$

The Haar mother wavelet can be translated and dilated by scale factor j and translate factor k:

$$\Psi_k^{(j)} = \sqrt{2^j}\Psi\left(2^j t - k\right),\ \ j,k \in \mathbb{Z}$$

Since the Haar mother wavelet maintains the property that $\left\|\Psi_k^{(j)}\right\| = 1$, and $\langle \Psi_k^{(j)}, \Psi_m^{(l)} \rangle = \delta_{j,l}$. Haar wavelets can form up an complete space as orthonormal basis. A continuous function $x(t)$ can be expanded into Haar wavelet series as:

$$x(t) = \sum_{j,k} c_k^{(j)} \Psi_k^{(j)}(t) = c_0 \Psi_0^{(0)}(t) + \sum_{m=0}^{j-1} \sum_{k=0}^{2^m - 1} c_k^{(m)} \Psi_k^{(m)}(t)$$

And corresponding coefficients can be evaluated as:

$$c_k^{(j)} = \frac{\langle x(t), \Psi_k^{(j)}(t) \rangle}{\langle \Psi_k^{(j)}(t), \Psi_k^{(j)}(t) \rangle} = \frac{1}{\left\|\Psi_k^{(j)}\right\|} \int_{-\infty}^{\infty} x(t)\, \Psi_k^{(j)}(t) dt$$

Due to the discretization of smart meter data of this thesis, HWT is implemented in discretized version, i.e., Discrete Wavelet Transforms (DWT). DWT refers to wavelet transforms (WT) where the wavelets are discretely sampled and represented and are popular for analysis on discrete sampled digital data. Deriving from the discretization of standard WT, discretized time series data $x(n)$ with length$N$ can be decomposed into detail coefficients $d_k^{(j)}$approximation coefficients $a_k^{(j_0)}$ at level j:

$$d_k^{(j)} = \frac{1}{\sqrt{2^j}} \sum_{n=0}^{N} x(n)\, \Psi_k^{(j)}(\mathrm{n}) = \frac{1}{\sqrt{2^j}} \sum_{n=0}^{N} x(n)\, \Psi\left(\frac{n - 2^j \cdot k}{2^j}\right)$$

$$a_k^{(j)} = \frac{1}{\sqrt{2^j}} \sum_{n=0}^{N} x(n)\, \phi_k^{(j)}(\mathrm{n}) = \frac{1}{\sqrt{2^j}} \sum_{n=0}^{N} x(n)\, \phi\left(\frac{n - 2^j \cdot k}{2^j}\right)$$

Hereafter, the decomposed components of time series $\hat{x}(n)$ can be expanded with Haar series by reconstructing from coefficients:

$$x(n) = \sum_k a_k^{(j_0)}\, \phi_k^{(j_0)}(\mathrm{n}) + \sum_{j=1}^{j_0} \sum_k d_k^{(j)} \Psi_k^{(j)}(\mathrm{n})$$

Appendix. B

Lemma Proofs of Bayesian Inference in Chapter IV

Lemma I: Given that $f(x^*;\theta)\sim p(y^*|x^*;\theta)$, the first moment $\mathbb{E}_{p(y^*|x^*)}[y^*]$ can be estimated with an unbiased estimator $\frac{1}{S}\sum_{s=1}^{S} f(x^*;\hat{\theta}_s)$.

Proof.

$$\mathbb{E}_{p(y^*|x^*)}[y^*] = \int y^*\, p(y^*|x^*)dy^* = \int y^* \int p(y^*|x^*;\theta)p(\theta)d\theta\, dy^*$$
$$= \int \left[\int p(y^*|x^*;\theta)\, y^* dy^*\right] p(\theta)d\theta = \int f(x^*;\theta)\, p(\theta)d\theta \tag{B-1}$$

Given a large enough sampling size S, we have:

$$\frac{1}{S}\sum_{s=1}^{S} f(x^*;\hat{\theta}_s) \to \int f(x^*;\theta)\, p(\theta)d\theta \tag{B-2}$$

Lemma II: Given that $f(x^*;\theta)\sim p(y^*|x^*;\theta)$ and $p(y^*|x^*;\theta) = \mathcal{N}(y^*: \mathbb{E}_\theta[f(x^*;\theta)], \sigma I)$, the second moment $\mathbb{E}_{p(y^*|x^*)}[(y^*)^T y^*]$ can be estimated with an unbiased estimator $\frac{1}{S}\sum_{s=1}^{S}\left[f\left(x^*;\hat{\theta}_s\right)\right]^T f\left(x^*;\hat{\theta}_s\right) + \sigma^2 I$.

Proof.

$$\begin{aligned}
\mathbb{E}_{p(y^*|x^*)}[(y^*)^T y^*] &= \int (y^*)^T y^* \, p(y^*|x^*) dy^* \\
&= \int (y^*)^T y^* \int p(y^*|x^*;\theta) p(\theta) d\theta \, dy^* \\
&= \int \left[\int p(y^*|x^*;\theta) \, (y^*)^T y^* dy^* \right] p(\theta) d\theta \\
&= \int \left[\mathbb{Cov}_{p(y^*|x^*)}[y^*] + \left(\mathbb{E}_{p(y^*|x^*)}[y^*] \right)^T \mathbb{E}_{p(y^*|x^*)}[y^*] \right] p(\theta) d\theta \\
&= \int \{ \sigma^2 I + [f(x^*;\theta)]^T f(x^*;\theta) \} \, p(\theta) d\theta \\
&= \sigma^2 I + \int \{ [f(x^*;\theta)]^T f(x^*;\theta) \} \, p(\theta) d\theta
\end{aligned} \tag{B-0-1}$$

Given a large enough sampling size S, we have:

$$\frac{1}{S} \sum_{s=1}^{S} \left[f\left(x^*;\hat{\theta}_s\right) \right]^T f\left(x^*;\hat{\theta}_s\right) + \sigma^2 I \to \sigma^2 I + \int \{ [f(x^*;\theta)]^T f(x^*;\theta) \} \, p(\theta) d\theta \tag{B-4}$$

Publications

Journal Publications

[1] Shi, Heng, Minghao Xu, and Ran Li. "Deep Learning for Household Load Forecasting – A Novel Pooling Deep RNN," IEEE Transactions on Smart Grid (2017).

[2] Shi, Heng, et al. "Data-driven Uncertainty Quantification and Characterization for Household Energy Demand Across Multiple Time-scales." IEEE Transactions on Smart Grid (2018).

[3] Shi H, Li F. Understand Demand Uncertainty with Deep Learning-A Dropout Stochastic Regularization Technique[J]. IEEE Transactions on Smart Grid, 2018. under review.

[4] Shi, Heng, Zhong Zhang and Furong Li, Decentralised Home Energy Management System to Reduce System Peak and Uncertainty by Wavelet Auto-decoupling Optimization, IEEE Transactions on Smart Grid, 2018. under review.

[5] Heng Shi, Furong Li, “Household Probabilistic Load Forecasting with Deep Learning using Dropout Generative Nets” [J]. IEEE Transactions on Smart Grid, 2018. under review.

[6] Zhong Zhang, Heng Shi, et al., “Volatility pricing and cost sharing for electricity load”, IEEE Transactions on Smart Grid, 2018, submitted.

[7] I. Hernando-Gil, H. Shi, F. Li, S. Djokic and M. Lehtonen, "Evaluation of Fault Levels and Power Supply Network Impedances in 230/400 V 50 Hz Generic Distribution Systems," in IEEE Transactions on Power Delivery, vol. 32, no. 2, pp. 768-777, April 2017.

[8] Chi Zhang, Heng Shi and Ran Li, “Household load forecasting with Incomplete Data”, *IEEE Transactions on Smart Grid*, 2018. Under Review.

Conference Publications

[9] Chen Zhao, Heng Shi, Ran Li and Furong Li, "Demand side response performance assessment: An impact analysis of load profile accuracy on DSR performances," 2015 IEEE Power & Energy Society General Meeting, Denver, CO, 2015, pp. 1-5. **(Best Paper Award in PES GM 2015**)

[10] Heng Shi, Zhong Zhang, Furong Li, "Decentralised control for combined heat and power system in energy community, " *CIRED 2018-25th International Conference on Electricity Distribution*. 2017.

[11] Shi, Heng, et al. "A Whole System Assessment of Novel Deep Learning Approach on Short-Term Load Forecasting." *Energy Procedia* 142 (2017): 2791-2796.

[11] Xinhe Yan, Heng Shi, et al. "Probabilistic Network Pricing Considering Demand Uncertainty in Distribution Systems." Accepted by 2018 IEEE Power & Energy Society General Meeting, Portland, OR, 2018.

[12] Chi Zhang, Ran Li, Heng Shi, et al, "Day-Ahead Electricity Price Forecasting Based on Deep Learning, " Special issue on CSEE JPES, to be submitted.

Reference

[1] "UK Greenhouse Gas Emission Trading Scheme Regulations," The Stationery OfficeDec, 2012.

[2] J. H. Williams, A. DeBenedictis, R. Ghanadan, A. Mahone, J. Moore, W. R. Morrow, *et al.*, "The technology path to deep greenhouse gas emissions cuts by 2050: the pivotal role of electricity," *science,* vol. 335, pp. 53-59, 2012.

[3] T. J. Foxon, "Transition pathways for a UK low carbon electricity future," *Energy Policy,* vol. 52, pp. 10-24, 2013.

[4] Hajian, Mahdi, William D. Rosehart, and Hamidreza Zareipour. "Probabilistic power flow by Monte Carlo simulation with Latin Hypercube Sampling." *IEEE Transactions on Power Systems* 28.2 (2013): 1550-1559.

[5] Jorgensen, P., J. S. Christensen, and J. O. Tande. "Probabilistic load flow calculation using Monte Carlo techniques for distribution network with wind turbines." *Harmonics and Quality of Power Proceedings, 1998. Proceedings. 8th International Conference On*. Vol. 2. IEEE, 1998.

[6] Morales, Juan M., et al. "Probabilistic power flow with correlated wind sources." *IET generation, transmission & distribution* 4.5 (2010): 641-651.

[7] Cai, Y. P., et al. "Identification of optimal strategies for energy management systems planning under multiple uncertainties." *Applied Energy* 86.4 (2009): 480-495.

[8] Cai, Y. P., et al. "An optimization-model-based interactive decision support system for regional energy management systems planning under uncertainty." *Expert Systems with Applications* 36.2 (2009): 3470-3482.

[9] Dong, C., et al. "Robust planning of energy management systems with environmental and constraint-conservative considerations under multiple uncertainties." *Energy Conversion and Management* 65 (2013): 471-486.

[10] Zhang, Yu, Nikolaos Gatsis, and Georgios B. Giannakis. "Robust energy management for microgrids with high-penetration renewables." *IEEE Transactions on Sustainable Energy* 4.4 (2013): 944-953.

[11] Y. Liu, Y. Li, H. B. Gooi, J. Ye, H. Xin, X. Jiang, J. Pan. "Distributed robust energy management of a multi-microgrid system in the real-time energy market," IEEE Transactions on Sustainable Energy, 2017, early access.

[12] Rahman, Saifur, and Rahul Bhatnagar. "An expert system based algorithm for short-term load forecast." *IEEE Transactions on Power Systems* 3.2 (1988): 392-399.

[13] Henley, Andrew, and John Peirson. "Non-Linearities in Electricity Demand and Temperature: Parametric Versus Non-Parametric Methods." *Oxford Bulletin of Economics and Statistics* 59.1 (1997): 149-162.

[14] Thornton, H. E., Brian J. Hoskins, and A. A. Scaife. "The role of temperature in the variability and extremes of electricity and gas demand in Great Britain." *Environmental Research Letters* 11.11 (2016): 114015.

[15] D. Shi, R. Li, R. Shi, and F. Li, "Analysis of the relationship between load profile and weather condition," in *2014 IEEE PES General Meeting | Conference & Exposition*, 2014, pp. 1-5.

[16] "Weather stations." *metoffice.gov.uk*. *Met Office, 14 Apr 2016.* https://www.metoffice.gov.uk/learning/making-a-forecast/first-steps/observations/weather-stations. 23 Feb 2016.

[17] S. T. C. THEREFORE, M. AS, and A. S. S. U. APPROVED, "Guide to the Expression of Uncertainty in Measurement," 1995.

[18] Der Kiureghian, Armen, and Ove Ditlevsen. "Aleatory or epistemic? Does it matter?." *Structural Safety* 31.2 (2009): 105-112.

[19] Bae, Ha-Rok, Ramana V. Grandhi, and Robert A. Canfield. "Epistemic uncertainty quantification techniques including evidence theory for large-scale structures." *Computers & Structures* 82.13-14 (2004): 1101-1112.

[20] Helton, Jon C., and William L. Oberkampf. "Alternative representations of epistemic uncertainty." (2004): 1-10.

[21] Ferson, Scott, et al. "Summary from the epistemic uncertainty workshop: consensus amid diversity." *Reliability Engineering & System Safety* 85.1-3 (2004): 355-369.

[22] Helton, J. C., et al. "A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory." *Computer Methods in Applied Mechanics and Engineering* 196.37-40 (2007): 3980-3998.

[23] Rao, K. Durga, et al. "Quantification of epistemic and aleatory uncertainties in level-1 probabilistic safety assessment studies." *Reliability Engineering & System Safety* 92.7 (2007): 947-956.

[24] Du, Xiaoping. "Unified uncertainty analysis by the first order reliability method." *Journal of mechanical design* 130.9 (2008): 091401.

[25] Mathieu, Johanna L., Duncan S. Callaway, and Sila Kiliccote. "Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing." *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*. IEEE, 2011.

[26] Jones, Lawrence E. *Renewable energy integration: practical management of variability, uncertainty, and flexibility in power grids*. Academic Press, 2017.

[27] Tabone, Michaelangelo D., and Duncan S. Callaway. "Modeling variability and uncertainty of photovoltaic generation: A hidden state spatial statistical approach." *IEEE Transactions on Power Systems* 30.6 (2015): 2965-2973.

[28] Saxena, Mr Rajesh, and Surendra Kumar Srivastava. "The Required Condition of Hamburger moment problem for operators." *International Journal of Current Trends in Science and Technology* 8.01 (2018): 20559-20563.

[29] Rohatgi, Vijay K., and AK Md Ehsanes Saleh. *An introduction to probability and statistics*. John Wiley & Sons, 2015.

[30] Feller, William. "A limit theorem for random variables with infinite moments." *Selected Papers I*. Springer, Cham, 2015. 721-726.

[31] Dertwinkel-Kalt, Markus, and Mats Köster. *Local thinking and skewness preferences*. No. 248. DICE Discussion Paper, 2017.

[32] Barato, Andre C., and Udo Seifert. "Skewness and kurtosis in statistical kinetics." *Physical review letters* 115.18 (2015): 188103.

[33] Komsta, Lukasz, and Frederick Novomestky. "Moments, cumulants, skewness, kurtosis and related tests." *R package version 0.14* (2015).

[34] Snyder, John M. *Generative modeling for computer graphics and CAD: symbolic shape design using interval analysis*. Academic press, 2014.

[35] Veroniki, Areti Angeliki, et al. "Methods to estimate the between-study variance and its uncertainty in meta-analysis." *Research synthesis methods* 7.1 (2016): 55-79.

[36] Bellini, Fabio, et al. "Generalized quantiles as risk measures." *Insurance: Mathematics and Economics* 54 (2014): 41-48.

[37] Parente, Paulo MDC, and João MC Santos Silva. "Quantile regression with clustered data." *Journal of Econometric Methods* 5.1 (2016): 1-15.

[38] Postek, Krzysztof, and Dick den Hertog. "Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set." *INFORMS Journal on Computing* 28.3 (2016): 553-574.

[39] Skarmeas, Dionysis, Constantinos N. Leonidou, and Charalampos Saridakis. "Examining the role of CSR skepticism using fuzzy-set qualitative comparative analysis." *Journal of business research* 67.9 (2014): 1796-1805.

[40] Deli, Irfan, and Naim Çağman. "Intuitionistic fuzzy parameterized soft set theory and its decision making." *Applied Soft Computing* 28 (2015): 109-113.

[41] Ferchichi, Ahlem, Wadii Boulila, and Imed Riadh Farah. "Towards an uncertainty reduction framework for land-cover change prediction using possibility theory." *Vietnam Journal of Computer Science* 4.3 (2017): 195-209.

[42] Kanal, Laveen N., and John F. Lemmer, eds. *Uncertainty in artificial intelligence*. Vol. 4. Elsevier, 2014.

[43] Smith, Ralph C. *Uncertainty quantification: theory, implementation, and applications*. Vol. 12. Siam, 2013.

[44] Hauseux, Paul, Jack Hale, and Stéphane Bordas. "Uncertainty Quantification-Sensitivity Analysis/Biomechanics." (2017).

[45] Constantinescu, Emil M., et al. "A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation." *IEEE Transactions on Power Systems* 26.1 (2011): 431-441.

[46] Eldred, Michael, and John Burkardt. "Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification." *47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition*. 2009.

[46] Nasrabadi, Nasser M. "Pattern recognition and machine learning." *Journal of electronic imaging* 16.4 (2007): 049901.

[47] Ghahramani, Zoubin, and Matthew J. Beal. "Variational inference for Bayesian mixtures of factor analysers." *Advances in neural information processing systems*. 2000.

[48] Ranganath, Rajesh, Sean Gerrish, and David Blei. "Black box variational inference." *Artificial Intelligence and Statistics*. 2014.

[49] Pousinho, Hugo Miguel Inácio, Víctor Manuel Fernandes Mendes, and João Paulo da Silva Catalão. "A risk-averse optimization model for trading wind energy in a market environment under uncertainty." *Energy* 36.8 (2011): 4935-4942.

[50] Li, Gengfeng, et al. "Risk analysis for distribution systems in the northeast US under wind storms." *IEEE Transactions on Power Systems* 29.2 (2014): 889-898.

[51] Shu, Zhen, et al. "Accelerated state evaluation and latin hypercube sequential sampling for composite system reliability assessment." *IEEE Transactions on Power Systems*29.4 (2014): 1692-1700.

[52] Allan, R. N., B. Borkowska, and C. H. Grigg. "Probabilistic analysis of power flows." *Proceedings of the Institution of Electrical Engineers*. Vol. 121. No. 12. IET Digital Library, 1974.

[53] Allan, R. N., and M. R. G. Al-Shakarchi. "Probabilistic ac load flow." *Proceedings of the Institution of Electrical Engineers*. Vol. 123. No. 6. IET Digital Library, 1976.

[54] Allan, R. N., AM Leite Da Silva, and R. C. Burchett. "Evaluation methods and accuracy in probabilistic load flow solutions." *IEEE Transactions on Power Apparatus and Systems* 5 (1981): 2539-2546.

[55] Schellenberg, Antony, William Rosehart, and José A. Aguado. "Cumulant-based stochastic nonlinear programming for variance constrained voltage stability analysis of power systems." *IEEE Transactions on power systems* 21.2 (2006): 579-585.

[56] Dadkhah, Maryam, and Bala Venkatesh. "Cumulant based stochastic reactive power planning method for distribution systems with wind generators." *IEEE Transactions on Power Systems* 27.4 (2012): 2351-2359.

[57] Ahsan, Q., and K. F. Schenk. "Sensitivity study of the cumulant method for evaluating reliability measures of two interconnected systems." *IEEE transactions on reliability* 36.4 (1987): 429-432.

[58] Stremel, John P. "Sensitivity study of the cumulant method of calculating generation system reliability." *IEEE Transactions on Power Apparatus and Systems* 2 (1981): 771-778.

[59] Billinton, Roy. "A sequential simulation method for the generating capacity adequacy evaluation of small stand-alone wind energy conversion systems." *Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on*. Vol. 1. IEEE, 2002.

[60] Bickel, P. J., F. Götze, and W. R. Van Zwet. "The Edgeworth expansion for U-statistics of degree two." *The Annals of Statistics* (1986): 1463-1484.

[61] Callaert, Herman, Paul Janssen, and Noel Veraverbeke. "An Edgeworth expansion for U-statistics." *The Annals of Statistics*(1980): 299-312.

[62] El-Ela, Adel Ali Abou. "Fast and accurate technique for power system state estimation." *IEE Proceedings C (Generation, Transmission and Distribution)*. Vol. 139. No. 1. IET Digital Library, 1992.

[63] Dhople, Sairaj V., and Alejandro D. Dominguez-Garcia. "A parametric uncertainty analysis method for Markov reliability and reward models." *IEEE Transactions on Reliability* 61.3 (2012): 634-648.

[64] Verbic, Gregor, and Claudio A. Canizares. "Probabilistic optimal power flow in electricity markets based on a two-point estimate method." *IEEE transactions on Power Systems* 21.4 (2006): 1883-1893.

[65] Morales, Juan M., and Juan Perez-Ruiz. "Point estimate schemes to solve the probabilistic power flow." *IEEE Transactions on power systems* 22.4 (2007): 1594-1601.

[66] Saunders, Christopher Scott. "Point estimate method addressing correlated wind power for probabilistic optimal power flow." *IEEE Transactions on Power Systems* 29.3 (2014): 1045-1054.

[67] Chen, Bokan, et al. "Robust optimization for transmission expansion planning: Minimax cost vs. minimax regret." *IEEE Transactions on Power Systems* 29.6 (2014): 3069-3077.

[68] Jiang, Ruiwei, et al. "Two-stage minimax regret robust unit commitment." *IEEE Transactions on Power Systems* 28.3 (2013): 2271-2282.

[69] Margellos, Kostas, Paul Goulart, and John Lygeros. "On the road between robust optimization and the scenario approach for chance constrained optimization problems." *IEEE Transactions on Automatic Control* 59.8 (2014): 2258-2263.

[70] Beyer, Hans-Georg, and Bernhard Sendhoff. "Robust optimization–a comprehensive survey." *Computer methods in applied mechanics and engineering* 196.33-34 (2007): 3190-3218.

[71] Gabrel, Virginie, Cécile Murat, and Aurélie Thiele. "Recent advances in robust optimization: An overview." *European journal of operational research* 235.3 (2014): 471-483.

[72] Ben-Tal, Aharon, and Arkadi Nemirovski. "Robust optimization–methodology and applications." *Mathematical Programming* 92.3 (2002): 453-480.

[73] Ben-Tal, Aharon, and Arkadi Nemirovski. "Robust optimization–methodology and applications." *Mathematical Programming* 92.3 (2002): 453-480.

[74] Bertsimas, Dimitris, et al. "Adaptive robust optimization for the security constrained unit commitment problem." *IEEE Transactions on Power Systems* 28.1 (2013): 52-63.

[75] Ren, Zhouyang, et al. "Probabilistic power flow analysis based on the stochastic response surface method." *IEEE Transactions on Power Systems* 31.3 (2016): 2307-2315.

[76] Wu, Hao, et al. "Probabilistic load flow based on generalized polynomial chaos." *IEEE Transactions on Power Systems*32.1 (2017): 820-821.

[77] H. Shi, Q. Ma, N. Smith and F. Li, "Data-driven Uncertainty Quantification and Characterization for Household Energy Demand Across Multiple Time-scales," in *IEEE Transactions on Smart Grid*. doi: 10.1109/TSG.2018.2817567

[78] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Quaestiones geographicae*, vol. 30, p. 87, 2011.

[79] J. Hoffmann-Jørgensen and G. Pisier, "The Law of Large Numbers and the Central Limit Theorem in Banach Spaces," *The Annals of Probability,* vol. 4, pp. 587-599, 1976.

[80] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks,* vol. 5, pp. 157-166, 1994.

[81] Zhang, Jing, et al. "A novel data-driven stock price trend prediction system." *Expert Systems with Applications* 97 (2018): 60-69.

[82] Atkinson, Andrew D., et al. "Wavelet ANOVA bisection method for identifying simulation model bias." *Simulation Modelling Practice and Theory* 80 (2018): 66-74.

[83] Yao, Jiangchao, et al. "Joint Latent Dirichlet Allocation for Social Tags." *IEEE Transactions on Multimedia* 20.1 (2018): 224-237.

[84] Abas, Faridah, Intan Safinar Ismail, and Nordin H. Lajis. "Generalized Likelihood Uncertainty Estimation (GLUE) methodology for optimization of extraction in natural products." *Food Chemistry* (2018).

[85] Claveria, Oscar, Enric Monte, and Salvador Torra. "Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection." *arXiv preprint arXiv:1805.00878* (2018).

[86] Sanroma, Gerard, et al. "Learning non-linear patch embeddings with neural networks for label fusion." *Medical image analysis* 44 (2018): 143-155.

[87] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning," *An MIT Press book in preparation. Draft chapters available at http://www. iro. umontreal. ca/~ bengioy/dlbook,* 2015.

[88] Kendall, Alex, and Yarin Gal. "What uncertainties do we need in bayesian deep learning for computer vision?." *Advances in Neural Information Processing Systems*. 2017.

[89] Box, George EP, and George C. Tiao. *Bayesian inference in statistical analysis*. Vol. 40. John Wiley & Sons, 2011.

[90] Dempster, Arthur P. "A generalization of Bayesian inference." *Classic works of the dempster-shafer theory of belief functions*. Springer, Berlin, Heidelberg, 2008. 73-104.

[91] Yarin Gal. Uncertainty in Deep Learning. PhD thesis, University of Cambridge, 2016.

[92] C. M. Carlo, "Markov chain monte carlo and gibbs sampling," *Notes,(April). 連結*, 2004.

[93] Gal, Yarin, and Zoubin Ghahramani. "A theoretically grounded application of dropout in recurrent neural networks." *Advances in neural information processing systems*. 2016.

[94] Ghahramani, Zoubin, and Matthew J. Beal. "Variational inference for Bayesian mixtures of factor analysers." *Advances in neural information processing systems*. 2000.

[95] Mnih, Andriy, and Karol Gregor. "Neural variational inference and learning in belief networks." *arXiv preprint arXiv:1402.0030* (2014).

[96] Srivastava, Nitish, et al. "Dropout: A simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research* 15.1 (2014): 1929-1958.

[97] D. M. Hawkins, "The problem of overfitting," Journal of chemical information and computer sciences, vol. 44, pp. 1-12, 2004.

[98] Carmesin, H-O. "Multilinear perceptron convergence theorem." *Physical Review E* 50.1 (1994): 622.

[99] Heckerman, David, and Christopher Meek. "Models and selection criteria for regression and classification." *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1997.

[100] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467,* 2016.

[101] UK Power Networks, "SmartMeter Energy Consumption Data in London Households." *data.london.gov.uk. London Datastore, 2015.* https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households. 01 July 2017.

[102] Ben Lee-Rodgers, "Hampstead nw3, London - Current Weather." *nw3weather.co.uk. NW3 Weather, 2010.* http://nw3weather.co.uk/wxdataday.php 12 July 2017.

[103] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*(2014).

[104] Xu, Yitian, Zhiji Yang, and Xianli Pan. "A novel twin support-vector machine with pinball loss." *IEEE transactions on neural networks and learning systems* 28.2 (2017): 359-370.

[105] Steinwart, Ingo, and Andreas Christmann. "Estimating conditional quantiles with the help of the pinball loss." *Bernoulli* 17.1 (2011): 211-225.

[106] Huang, Xiaolin, Lei Shi, and Johan AK Suykens. "Sequential minimal optimization for SVM with pinball loss." *Neurocomputing* 149 (2015): 1596-1603.

[107] Firth, Steven, et al. "Identifying trends in the use of domestic appliances from household electricity consumption measurements." *Energy and Buildings* 40.5 (2008): 926-936.

[108] JP Zimmermann, M Evans, et al. "Household Electricity Survey A study of domestic electrical product usage. " Intertek, May 2012, https://www.gov.uk/government/publications/household-electricity-survey--2.

[109] G. Carpinelli, F. Mottola, D. Proto, and A. Russo, "A multi-objective approach for microgrid scheduling," IEEE Transactions on Smart Grid, 2016.

[110] S. S. Reddy, "Multi-objective based congestion management using generation rescheduling and load shedding," IEEE Transactions on Power Systems, vol. 32, pp. 852-863, 2017.

[111] C. S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to wavelets and wavelet transforms: a primer," 1997.

[112] C. E. Heil and D. F. Walnut, "Continuous and discrete wavelet transforms," SIAM review, vol. 31, pp. 628-666, 1989.

[113] Z. Zhang, J. Wang, T. Ding, and X. Wang, "A two-layer model for microgrid real-time dispatch based on energy storage system charging/discharging hidden costs," IEEE Transactions on Sustainable Energy, vol. 8, pp. 33-42, 2017.

[114] C. Zhao, S. Dong, C. Gu, F. Li, Y. Song, and N. P. Padhy, "New Problem Formulation for Optimal Demand Side Response in Hybrid AC/DC Systems," IEEE Transactions on Smart Grid, 2016.

[115] "SoLa Bristol SDRC 9.8 Final Report," WPD, Bristol. SOM2016.

[116] C. Zhao, H. Shi, R. Li, and F. Li, "Demand side response performance assessment: An impact analysis of load profile accuracy on DSR performances," in Power & Energy Society General Meeting, 2015 IEEE, 2015, pp. 1-5.

[117] Y. Wang, Q. Xia, and C. Kang, "Secondary Forecasting Based on Deviation Analysis for Short-Term Load Forecasting," *IEEE Transactions on Power Systems*, vol. 26, pp. 500-507, 2011.

[118] Tao Hong and D. A. Dickey. (2013, 01 August 2016). *Electric load forecasting: fundamentals and best practices*. Available: https://www.otexts.org/elf

[119] F. L. Quilumba, W. J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities," *IEEE Transactions on Smart Grid*, vol. 6, pp. 911-918, 2015.

[120] H. Tao, W. Pu, A. Pahwa, G. Min, and S. M. Hsiang, "Cost of temperature history data uncertainties in short term electric load forecasting," in

Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on, 2010, pp. 212-217.

[121] T. Hong, P. Wang, and L. White, "Weather station selection for electric load forecasting," *International Journal of Forecasting,* vol. 31, pp. 286-295, 4// 2015.

[122] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *International Journal of Forecasting,* vol. 32, pp. 585-597, 7// 2016.

[123] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. S. Venkata, *et al.*, "An Efficient Approach to Short-Term Load Forecasting at the Distribution Level," *IEEE Transactions on Power Systems,* vol. 31, pp. 2526-2537, 2016.

[124] A. Espasa and I. Mayo-Burgos, "Forecasting aggregates and disaggregates with common features," *International Journal of Forecasting,* vol. 29, pp. 718-732, 10// 2013.

[125] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy,* vol. 98, pp. 40-49, 2016.

[126] Y. Chen, P. B. Luh, C. Guan, Y. Zhao, L. D. Michel, M. A. Coolbeth, *et al.*, "Short-Term Load Forecasting: Similar Day-Based Wavelet Neural Networks," *IEEE Transactions on Power Systems,* vol. 25, pp. 322-330, 2010.

[127] R. Al-Otaibi, N. Jin, T. Wilcox, and P. Flach, "Feature Construction and Calibration for Clustering Daily Load Curves from Smart-Meter Data," *IEEE Transactions on Industrial Informatics,* vol. 12, pp. 645-654, 2016.

[128] D. Shi, R. Li, R. Shi, and F. Li, "Analysis of the relationship between load profile and weather condition," in *2014 IEEE PES General Meeting | Conference & Exposition*, 2014, pp. 1-5.

[129] Tao Hong, Pierre Pinson, Shu Fan, Global Energy Forecasting Competition 2012, International Journal of Forecasting, Volume 30, Issue 2, April–June 2014, Pages 357-363, ISSN 0169-2070,.

[130] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, Rob J. Hyndman, Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond, International Journal of Forecasting, Volume 32, Issue 3, July–September 2016, Pages 896-913, ISSN 0169-2070,

[131] X. Liu, L. Ivanescu, R. Kang, and M. Maier, "Real-time household load priority scheduling algorithm based on prediction of renewable source availability," *IEEE Transactions on Consumer Electronics,* vol. 58, pp. 318-326, 2012.

[132] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating Practice Theory in Sub-Profile Models for Short Term Aggregated Residential Load Forecasting."

[133] M. Chaouch, "Clustering-Based Improvement of Nonparametric Functional Time Series Forecasting: Application to Intra-Day Household-Level Load Curves," *IEEE Transactions on Smart Grid,* vol. 5, pp. 411-419, 2014.

[134] C. J. Willmott, S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, *et al.*, "Statistics for the evaluation and comparison of models," 1985.

[135] A. Marinescu, C. Harris, I. Dusparic, S. Clarke, and V. Cahill, "Residential electrical demand forecasting in very small scale: An evaluation of forecasting methods," in *Software Engineering Challenges for the Smart Grid (SE4SG), 2013 2nd International Workshop on*, 2013, pp. 25-32.

[136] S. Humeau, T. K. Wijaya, M. Vasirani, and K. Aberer, "Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households," in *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*, 2013, pp. 1-6.

[137] A. Veit, C. Goebel, R. Tidke, C. Doblander, and H.-A. Jacobsen, "Household electricity demand forecasting: benchmarking state-of-the-art methods," in

Proceedings of the 5th international conference on Future energy systems, 2014, pp. 233-234.

[138] S. Haben, J. Ward, D. V. Greetham, C. Singleton, and P. Grindrod, "A new error measure for forecasts of household-level, high resolution electrical energy consumption," *International Journal of Forecasting*, vol. 30, pp. 246-256, 2014.

[139] Y.-H. Hsiao, "Household electricity demand forecast based on context information and user daily schedule analysis from meter data," *IEEE Transactions on Industrial Informatics*, vol. 11, pp. 33-43, 2015.

[140] Hong, Tao, and Shu Fan. "Probabilistic electric load forecasting: A tutorial review." *International Journal of Forecasting* 32.3 (2016): 914-938.

[141] Wan, Can, et al. "Probabilistic forecasting of wind power generation using extreme learning machine." *IEEE Transactions on Power Systems* 29.3 (2014): 1033-1044.

[142] Shi, Heng, Minghao Xu, and Ran Li. "Deep learning for household load forecasting–A novel pooling deep RNN." *IEEE Transactions on Smart Grid* (2017).

[143] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

[144] Sak, Haşim, Andrew Senior, and Françoise Beaufays. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling." *Fifteenth annual conference of the international speech communication association*. 2014.

[145] Hearst, Marti A., et al. "Support vector machines." *IEEE Intelligent Systems and their applications* 13.4 (1998): 18-28.

[146] Aiken, Leona S., Stephen G. West, and Steven C. Pitts. "Multiple linear regression." *Handbook of psychology* (2003): 481-507.

[147] Andrews, David F. "A robust method for multiple linear regression." *Technometrics* 16.4 (1974): 523-531.

[148] Sousa, S. I. V., et al. "Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations." *Environmental Modelling & Software*22.1 (2007): 97-103.

[149] Jenkins, G. M. "Autoregressive–Moving Average (ARMA) Models." *Encyclopedia of Statistical Sciences* (1982).

[150] Adomian, George. "A review of the decomposition method in applied mathematics." *Journal of mathematical analysis and applications* 135.2 (1988): 501-544.

[151] Kannan, Ramachandran, and Neil Strachan. "Modelling the UK residential energy sector under long-term decarbonisation scenarios: Comparison between energy systems and sectoral modelling approaches." *Applied Energy* 86.4 (2009): 416-428.

[152] "UK bank holidays." *gov.uk. GOV.UK, 2012.* https://www.gov.uk/bank-holidays. 13 July 2017.

[153] Hallin, Marc, et al. "MULTIVARIATE QUANTILES AND MULTIPLE-OUTPUT REGRESSION QUANTILES: FROM L_1 OPTIMIZATION TO HALFSPACE DEPTH [with Discussion and Rejoinder]." *The Annals of Statistics* (2010): 635-703.

[154] Le Maıtre, O. P., et al. "Uncertainty propagation using Wiener–Haar expansions." *Journal of computational Physics* 197.1 (2004): 28-57.

[155] Paley, R. E. A. C. "A remarkable series of orthogonal functions (I)." *Proceedings of the London Mathematical Society* 2.1 (1932): 241-264.

[156] Shensa, Mark J. "The discrete wavelet transform: wedding the a trous and Mallat algorithms." *IEEE Transactions on signal processing* 40.10 (1992): 2464-2482.

[157] Chaovalit, Pimwadee, et al. "Discrete wavelet transform-based time series analysis and mining." *ACM Computing Surveys (CSUR)* 43.2 (2011): 6.

[158] Knowles, Gregory P. "Device and method for data compression/decompression using a discrete wavelet transform." U.S. Patent No. 6,118,902. 12 Sep. 2000.

[159] Pérez-Rendón, Antonio F., and Rafael Robles. "The convolution theorem for the continuous wavelet tranform." *Signal processing* 84.1 (2004): 55-67.

[160] Qureshi, Faisal Zubair. "Image compression using wavelet transform." *Computational Vision Project* (2008).

[161] Wikipedia, “Skewness.” *en.wikipedia.org*. *Wikipedia, 15 Mar 2017.* https://en.wikipedia.org/wiki/Skewness. 28 Feb 2018.