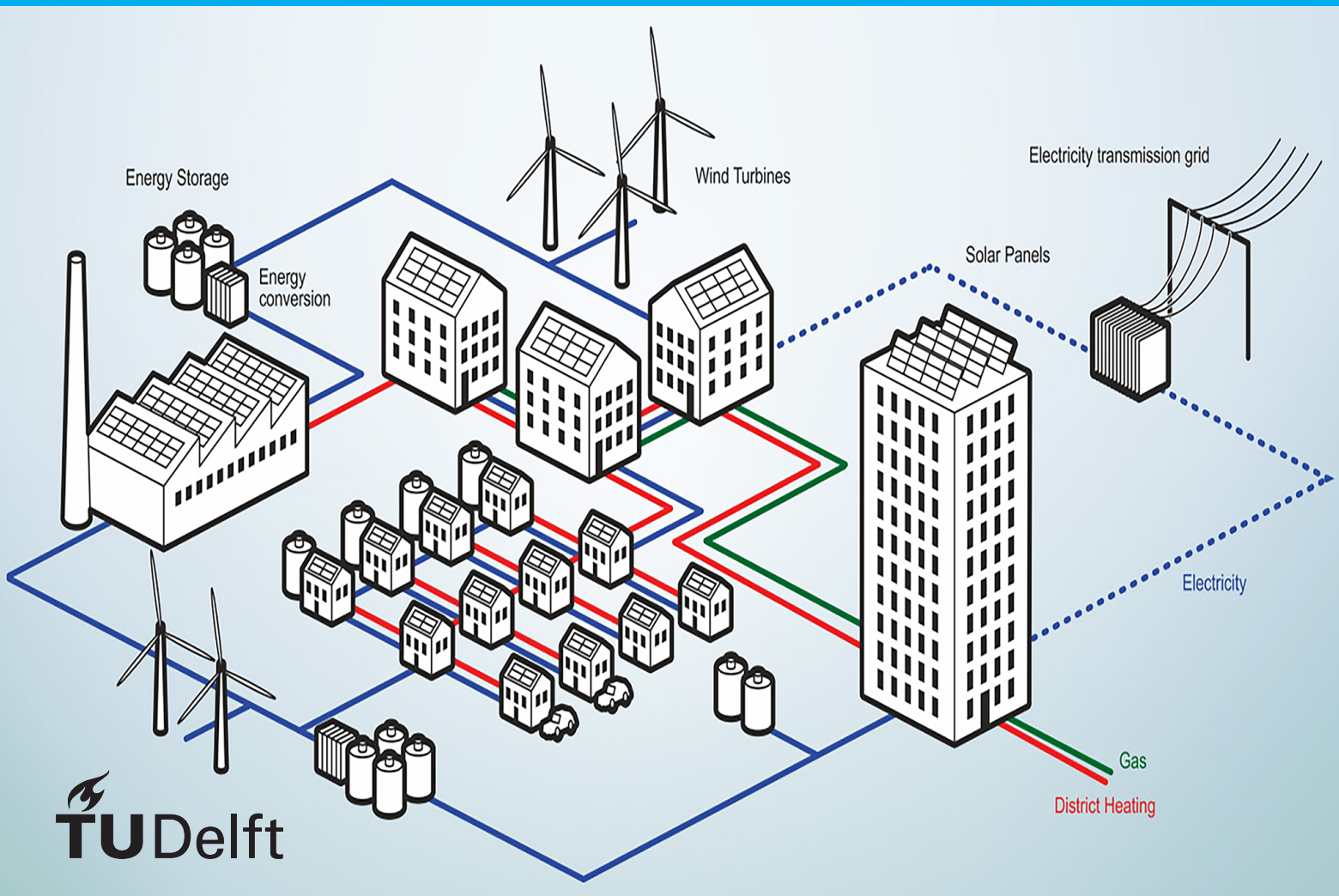
Surrogate Modelling for Multi-carrier Distribution Networks

Master Thesis

Guozheng Zhang

Intelligent Electrical Power Grids (IEPG) Group



Surrogate Modelling for Multi-carrier Distribution Networks

Master Thesis

by

Guozheng Zhang

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday May 26, 2021 at 10:00 AM.

Student number: 4792149

Project duration:	February, 2020 – May, 2021	
Thesis committee:	Prof. dr. Peter Palensky,	IEPG (TU Delft)
	Dr. ir. Milos Cvetkovic,	IEPG (TU Delft)
	Dr. Jianning Dong,	DCE&S (TU Delft)
	Ir. Digvijay Gusain	IEPG (TU Delft)

TU Delft

Abstract

The climate crisis is being paid more attention owing to its notable effects on the environment globally. For the sake of facing the challenge, the European Union (EU) has developed different strategies. One of which is the “2030 Climate Target Plan” published by the European Commission in Brussel in September 2020. It attempts to achieve a greenhouse gas (GHG) emissions reduction target by at least 55% by 2030.

Based on the plan, a list of ideas and definitions centred around integrated energy systems has been proposed. To assess energy system integration strategy in a Dutch context, a generic model of the energy system is necessary. To this end, a comprehensive hybrid energy system, consisting of a Dutch electricity distribution network model and a heating distribution network model, is developed. The electrical network is established by pandapower with three areas (net-zero energy building, rural, urban) owned by main Dutch Distribution System Operators (DSOs). Especially, the rural area is expanded by adding photovoltaic (PV) modules as a representative to evaluate the impact of integrating renewable energy sources (RES). The heating network is also built on the basis of the rural area, and it is simulated in OpenModelica with taking into account the supply system, the pipe system, the heater & storage system as well as the demand system. Moreover, controllers are introduced for the sake of tackling the excess PV power issue existing in the above electrical network. They are built in OpenModelica by designing operational strategies based on the charging state and the discharging state of the heater & storage system in the heating network.

However, the simulation of the integrated energy system model is computationally expensive due to the inter-dependencies between various energy sectors, dynamic operation of components within individual energy domain, etc. To overcome the computational burden of the detailed models, surrogate models are proposed to enhance the computational efficiency by replacing the expensive parts of the models in the simulation. In the thesis work, eight machine learning algorithms are covered to create the respective surrogate models in the electrical network and the heating network. Furthermore, their performances are compared in terms of different indicators. In consequence, the best model is selected in each network, which offers a useful recommendation for the integrated energy system modelling with a high computation efficiency.

The simulation results show the linear regression (LR) model can be used as a surrogate model to represent the electrical network based on the rural area of the integrated energy system. The long short-term memory (LSTM) model can be taken as a surrogate model to represent the heating network based on the rural area of the integrated energy system.

Preface

Throughout the whole period of my master's study including the writing of this thesis, I have received a great deal of support, assistance and encouragement.

I would first like to express my great appreciation to my daily supervisor, PhD student Digvijay Gusain, for his valuable and constructive suggestions during my research work. His patient guidance pushed me to sharpen my thinking and brought my work to a higher level. Furthermore, he gave me plenty of active encouragement in the daily research, which made me feel confident about conquering all the difficulties when facing different challenges!

I would like to express my deep gratitude to my supervisors Dr.Milos Cvetkovic and Professor Peter Palensky for their insightful guidance, enthusiastic encouragement and helpful suggestions of this research work. Dr. Milos offered me this interesting topic and kept my progress on schedule, and his guidance gave me a better understanding of my work! Moreover, his lecture related to the co-simulation in the energy system made me feel interesting in the field and would like to explore deeper along the direction in the thesis. Prof. Palensky put forward a list of practical suggestions in the work, and I really seized great opportunities when taking parting in the discussion every time! After a deeper thinking from his guidance, my research progress has obvious enhancement by considering more key details.

I would like to thank the committee members especially to Dr. Jianning for taking interest in this research and for the willingness to be part of the committee. Furthermore, I would like to acknowledge Prof. Jose and Dr. Simon in the Intelligent Electrical Power Grids (IEPG) group. They are always open to arrange a schedule and discussion, and hence I feel very lucky that I can join the group to carry out my research work. I would also like to extend my thanks to my university TU Delft which made me get a deeper understanding of my major knowledge and offered my opportunity to expand my horizons when exchanging in KU Leuven and ETH Zurich.

In addition, I would like to take this opportunity to say thank you to my families especially my parents for their love, counsel, support and encouragement. You are always there whenever I meet any problem, and your considerable supports make me believe myself all the time! I would like to sincerely thank all of my friends and my seniors, especially Haydn, Yi, Jiaqi, Yaqi, Smit, Calvin and Dr. Che-hao. Your enthusiasm and friendliness made me not feel lonely when studying abroad! I will not forget those golden memories from the bottom of my heart! Last but not least, I am indebted to my roommates Yue, Xinyue, Shuxin, Jiarui, Zitao and Huaiyang for your company in the special lockdown period. You make me realise we are a family!

Guozheng Zhang
Delft, May 2021

Contents

List of Figures xi

List of Tables xiv

List of Algorithms xv

List of Acronyms xviii

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Problem Definition . . . 3
- 1.3 Objective & Research Questions . . . 4
- 1.4 Research Approach . . . 5
- 1.5 Outline . . . 6

2 Literature Review 7
- 2.1 Establishment Methods of Electrical Distribution Networks . . . 7
- 2.2 Multi-carrier Energy Systems . . . 8
  - 2.2.1 Modelling and Simulation Methods . . . 9
  - 2.2.2 Computationally Expensive Problems . . . 10
- 2.3 Surrogate Models . . . 11
  - 2.3.1 General Approaches for Creating Surrogate Models . . . 11
  - 2.3.2 Applications in Energy Systems . . . 12
- 2.4 Machine Learning Algorithms . . . 12
  - 2.4.1 Basic Definitions . . . 12
  - 2.4.2 Data Splitting Approaches . . . 13
  - 2.4.3 Research Directions in Energy Systems . . . 14

3 The Electrical Network Model 17
- 3.1 The Background of System Operators . . . 17
  - 3.1.1 TSOs in Europe . . . 17
  - 3.1.2 DSOs in Europe . . . 18
- 3.2 The Modelling of the LV Distribution Network in the Netherlands . . . 19
  - 3.2.1 The Distribution of Main Dutch DSOs . . . 19
  - 3.2.2 The Simulation Tool & Open Sources . . . 20
  - 3.2.3 Network Structures & Simulation Parameters . . . 21
  - 3.2.4 Model Validation by Reference Values . . . 25
  - 3.2.5 RES Integration in the Network . . . 26

4 The Multi-carrier Energy System Model 29
- 4.1 Simulation Tools . . . 29
- 4.2 The Modelling of the Heating Network . . . 29
- 4.3 The Modelling of Controllers . . . 33
- 4.4 The Modelling of the Multi-carrier Energy System . . . 34
  - 4.4.1 The Methodology of Co-simulation . . . 35
  - 4.4.2 The Establishment of the Multi-carrier Energy System . . . 35

5 Surrogate Models for the Multi-carrier Energy System 39
- 5.1 The Background of Surrogate Models . . . 39
- 5.2 The Mathematical Description of Surrogate Models . . . 40
- 5.3 Detailed Method Illustrations of Creating Surrogate Models . . . 40
  - 5.3.1 Linear Regression . . . 41
  - 5.3.2 Regressor Chain . . . 42

5.3.3 Linear Support Vector Machine . . . 43
5.3.4 Decision Tree . . . 44
5.3.5 Random Forest . . . 44
5.3.6 k-nearest Neighbours . . . 46
5.3.7 Multilayer Perceptron . . . 46
5.3.8 Long Short-term Memory . . . 48
5.4 The Modelling of Surrogate Models . . . 50
5.4.1 Simulation Platforms . . . 50
5.4.2 The Establishment of the Surrogate Models in the Energy System . . . 51
5.4.3 The Description of the Modelling Approach . . . 54
5.4.4 Performance Indicators . . . 55

6 Results & Discussions **57**
6.1 Metric 1: Speed-up Factor . . . 57
6.1.1 Simulation Time of the Original Model . . . 57
6.1.2 Simulation Time of Surrogate Models . . . 58
6.1.3 Performance Comparisons Based on Metric 1 . . . 58
6.2 Metric 2: Root Mean Square Error . . . 60
6.2.1 Simulation Results of the Original Model . . . 60
6.2.2 Simulation Results of Surrogate Models . . . 61
6.2.3 Performance Comparisons Based on Metric 2 . . . 64
6.3 Another surrogate model: LSTM model . . . 65
6.3.1 The description of LSTM model in the heating network . . . 65
6.3.2 Simulation results based on the metric 1 in the heating network . . . 65
6.3.3 Simulation results based on the metric 2 in the heating network . . . 66
6.4 Performance Summary of Surrogate Models Based on Metric 1 and Metric 2 . . . 67

7 Conclusion & Outlook **69**
7.1 Research Questions . . . 69
7.2 Contributions . . . 71
7.3 Recommendations and Future Works . . . 71

A Transmission Network Modelling **73**

B Test Results of Designed Controllers **79**

C ANOVA Analysis **81**

D LSTM Model in the Electrical Network **83**

Bibliography **84**

List of Figures

1.1 The global land and ocean surface temperature in January between 1880 and 2020 [2]. 1
1.2 The EU's pathway to sustained economic prosperity and climate neutrality, 1990-2050 [4]. 2
1.3 MCES as a multigrid component [11]. . . . 3
1.4 Fraction renewable energy in the Netherlands [20]. . . . 4

2.1 Schematic illustration of the spatial perspective concept [13]. . . . 8
2.2 Classification of energy system models according to modelling approach [38]. . . . 9
2.3 Structure of the co-simulation environment [48]. . . . 10
2.4 Interactions between MCES technologies and networks [11]. . . . 10
2.5 Machine/deep learning in the context of artificial intelligence [70]. . . . 12
2.6 The description of the hold-out method. . . . 13
2.7 Regression example: (a). underfit; (b). good fit; (c). overfit [79]. . . . 13
2.8 An example of the 5-fold cross-validation method [80]. . . . 14
2.9 Number of publications in load, price, wind and solar forecasting returned by respective Scopus search [72]. . . . 15

3.1 The role of Tennet in the electricity network [89]. . . . 17
3.2 Conventional scenario versus emerging scenario in the power system due to the emergence of distributed energy resources [91]. . . . 18
3.3 The distribution of DSOs in Europe [90]. . . . 18
3.4 The energy distribution network operators in the Netherlands [36]. . . . 19
3.5 The NZEB in Rijsdijk, Etten-Leur [94]. . . . 20
3.6 The LV distribution network in the Netherlands. . . . 21
3.7 The Dutch LV distribution network of the rural area considering the PV integration. . . . 26

4.1 The Dutch heating grid of the rural area (EH, electric heater; TES, thermal energy storage). The dash boxes highlight six different groups of the households in the demand system. . . . 30
4.2 The structure of one subsystem of the heater & storage system model in OpenModelica. 30
4.3 The structure of the Dutch heating network based on the rural area in OpenModelica. Four main systems are highlighted in dash boxes. . . . 31
4.4 The structure of one subsystem of the demand system model in OpenModelica. . . . 32
4.5 The structure of one subsystem of the controller model in OpenModelica. The modules related to the control rules are highlighted in red. . . . 34
4.6 The description of FMUWorld. (a) FMUWorld structure (b) Time step determination for simulation setups with variable step sizes for each FMU [23]. . . . 35
4.7 The Dutch hybrid network of the rural area (PV, photovoltaic; EH, electric heater; TES, thermal energy storage). . . . 35
4.8 The simplified scheme of the co-simulation framework based on the multi-carrier energy system in the rural area. . . . 36

5.1 Three main aspects considered in the computationally expensive problems according to Ref. [108]. . . . 40
5.2 Population and sample regression [113]. . . . 41
5.3 Graphical illustration of regressor chain [118]. . . . 42
5.4 Support vector machine [121]. . . . 43
5.5 Basic structure of a three-layer decision tree [125]. . . . 44
5.6 Architecture of a random forest model [129]. . . . 45

5.7 An example of k-nearest neighbor algorithm illustration used to predict the class of a new data point [133]. . . . 46
5.8 A taxonomy of neural network architectures [135]. . . . 47
5.9 The general structure of a perception [138]. . . . 47
5.10 An example of the multi-layer perception [139]. . . . 48
5.11 The structure of the RNN [141]. . . . 48
5.12 The structure of the long short-term memory neural network [143]. . . . 49
5.13 The structure of the LSTM model used in the thesis work. . . . 50
5.14 Two Keras implementations: keras-team (left) and tf.keras (right) [149]. . . . 51
5.15 The simplified scheme of the surrogate models based on the multi-carrier energy system. 52
5.16 The structures of surrogate models used to replace the original networks in the energy system. (a). the surrogate model in the electrical network; (b). the surrogate model in the heating network. . . . 53
5.17 The description of the dataframe of **Ya** and **Yp** in two networks. (a). The dataframe of **Ya** (b). The dataframe of **Yp**. . . . 56

6.1 Performance comparisons of simulation time between the original model and the surrogate models in the electrical network. Solid dots show the simulation time of different surrogate models. Red dash line represents the simulation time of the original model. . 59
6.2 Performance comparisons of simulation time between the original model and the surrogate models in the heating network. Solid dots show the simulation time of different surrogate models. Red dash line represents the simulation time of the original model. . 60
6.3 The simulation results based on the outputs of Bus 8 in the electrical network. . . . 61
6.4 The simulation results based on the outputs of group 1 in the heating network. The left and right column show supply temperatures and storage temperatures of three representative days in each season. The subfigures from top to bottom in each column represent the temperature profiles of January, April, July and October, respectively). . . . 62
6.5 The simulation results of the outputs based on Bus 8 between different surrogate models in the electrical network. The result of the original model is also illustrated as a reference. (a). The simulation result of surrogate models based on one simulated day. (b). Zoomed result of (a). . . . 63
6.6 The simulation results based on the outputs of group 1 in the heating network. The subfigures from left to right represent the temperature profiles of supply temperature and storage temperature, respectively. The results of the original model are also illustrated as a reference. . . . 63
6.7 Performance comparisons of RMSE values for different surrogate models in the electrical network when the simulated days are one year. . . . 64
6.8 Performance comparisons of RMSE values for different surrogate models in the heating network when the simulated days are 12 representative days. . . . 65
6.9 The illustration of LSTM model in the heating network. . . . 65
6.10 Performance comparisons of RMSE values for eight surrogate models in the heating network when the simulated days are 12 representative days. . . . 66
6.11 The simulation results of the outputs based on group 1 in the LSTM model of the heating network. (a). The supply temperature prediction results; (b). The storage temperature prediction results. The result of the original model is shown as a reference. . . . 67

A.1 The transmission grid map in the Netherlands (the study area is highlighted in blue). . 73
A.2 The zoomed transmission grid map of the study area. . . . 74
A.3 The designed transmission network used in the thesis work based on the above zoomed area. . . . 74
A.4 Specifications 380 kV connections as provided by TenneT (source:Gockel.P.N.M, "Steady-state voltage profile and reactive power balance for EHV AC cable systems in the Randstad 380 project", Master Thesis, TU Delft, 2009). . . . 77

B.1 The simulation result based on the tests results of controllers designed in the thesis work with the zoomed period is highlighted in black. . . . 79
B.2 A zoomed result of Figure B.1. . . . 80

D.1 Performance comparisons of RMSE values for bus 8 between eight surrogate models in the electrical network when the simulated days are one year. 83
D.2 The simulation results of the outputs based on Bus 8 in the LSTM model of the electrical network. The result of the original model is illustrated as a reference. 84

List of Tables

2.1 Subset of the total DSOs indicators used to build the large-scale representative distribution networks [31]. . . . 7
2.2 Representative methods creating surrogate models in the application of energy systems. 11

3.1 Comparison of open source element model libraries [21]. . . . 21
3.2 Abbreviation meaning of used parameters. . . . 22
3.3 The simulation parameters of the external grid. . . . 22
3.4 The simulation parameters of different buses. . . . 22
3.5 The simulation parameters of transformers. . . . 23
3.6 The simulation parameters of cable lines. . . . 23
3.7 The eight categories of the NEDU profiles [99]. . . . 24
3.8 The simulation parameters of loads at time=0. . . . 24
3.9 The Dutch DSO indicators and their values in the designed network. . . . 25
3.10 DSOs indicators and reference values [31]. . . . 25
3.11 Typical Transformation capacity indicators and reference values [31]. . . . 26
3.12 Overview of the assumptions of PV panels for the three areas [36]. . . . 26
3.13 The simulation parameters of PV modules at time=0. . . . 27

4.1 The design of parameters using in the heating network. . . . 32
4.2 The simulation parameters of electric heaters in the electrical network at time=0. . . . 37

5.1 Time in seconds on the Madelson data set for various machine learning libraries exposed in Python [24]. . . . 51
5.2 Tuned hyperparameters in the machine/deep learning algorithm. . . . 54
5.3 Key parameters when calculating performance indicators in the electrical and heating network. . . . 56

6.1 The simulation time of the electrical network. . . . 57
6.2 The simulation time of the heating network. . . . 57
6.3 Performance comparisons of the simulation time for different surrogate models in the electrical network whose simulated days are one-year. Bold values mark the best performance in the column of the mean value. . . . 58
6.4 Performance comparisons of the simulation time for different surrogate models in the heating network whose simulated days are 12 representative days. Bold values mark the best performance in the column of the mean value. . . . 58
6.5 A summary based on the first metric reflecting the performances of the original model and different surrogate models in the electrical network. Bold values mark the best performance in the corresponding column. . . . 59
6.6 A summary based on the first metric reflecting the performances of the original model and different surrogate models in the heating network. Bold values mark the best performance in the corresponding column. . . . 60
6.7 Average RMSE values of surrogate models relative to the original model in the electrical network. . . . 64
6.8 Average RMSE values of surrogate models relative to the original model in the heating network. . . . 64
6.9 A summary of simulation time (s), speed-up factor and RMSE of LSTM model in the heating network. . . . 66

6.10 A summary based on two metrics reflecting the performances of the original model and different surrogate models in the electrical network. Bold values mark the best performance in the corresponding column. 67
6.11 A summary based on two metrics reflecting the performances of the original model and different surrogate models in the heating network. Bold values mark the best performance in the corresponding column. 68

A.1 The simulation parameters of the external grid in the transmission network. 75
A.2 The simulation parameters of different buses in the transmission network. 75
A.3 The simulation parameters of cable lines in the transmission network. 75
A.4 The simulation parameters of generators in the transmission network. 76

C.1 The ANOVA analysis based on the original model and surrogate models in the electrical network. 81
C.2 The ANOVA analysis based on the original model and surrogate models in the heating network. 81

D.1 A summary of simulation time (s), speed-up factor and RMSE of LSTM model in the electrical network. 83

List of Algorithms

1 The charging state of the heating network. 33
2 The discharging state of the heating network. 34

List of Acronyms

AI	Artificial Intelligence
ANOVA	Analysis of Variance
ANN	Artificial Neural Networks
API	Application Programming Interface
BENG	Bijna Energieneutrale Gebouwen
CNG	Compressed Natural Gas
DERs	Distributed Energy Sources
DL	Deep Learning
DSOs	Distribution System Operators
DT	Decision Trees
ED	Euclidean Distance
EH	Electric Hater
ENTSO-E	European Network of Transmission System Operators
EU	European Union
EVs	Electrical Vehicles
FMI	Function Mock-up Interface
FMUs	Functional Mock-up Units
GDP	Gross Domestic Product
GHG	Greenhouse Gas
GPs	Gaussian Processes
HLA	High Level Architecture
HV	High Voltage
JRC	Joint Research Centre
k-NN	k-nearest Neighbours
LNG	Liquefied Natural Gas
LR	Linear Regression
LSTM	Long Short-term Memory
LV	Low Voltage
MCESs	Multi-carrier Energy Systems
MES	Multi-energy Systems
ML	Machine Learning
MLP	Multi-layer Perceptron
MSE	Mean Squared Error
MV	Medium Voltage
NCDC	National Climatic Data Centre
NEDU	Nederlandse Energie Data Uitwisseling
NG	Natural Gas
NN	Neural Network
NZEB	Net-zero Emission Building
OMEdit	OpenModelica Connection Editor
P2X	Power-to-X
PV	Photovoltaics
RBF	Radial Basis Function
RC-LR	Linear Regressor Chain
RC-SVR	Linear Support Vector Machine with Regressor Chain
RES	Renewable Energy Sources
RF	Random Forest
RMSE	Root-mean-square-error
RNN	Recurrent Neural Network
RSM	Response Surface Methodology

SITL	System-in-the-Loop
SUF	Speed-up Factor
SVM	Support Vector Machine
TES	Thermal Energy Storage
TF	TensorFlow
TSOs	Transmission System Operators
RMSE	Root-mean-square-error
RNN	Recurrent Neural Network
RSM	Response Surface Methodology
SITL	System-in-the-Loop
SUF	Speed-up Factor
SVM	Support Vector Machine
TES	Thermal Energy Storage
TSOs	Transmission System Operators

1 Introduction

The objective of the chapter is to offer a general introduction of the thesis. First, a brief background is given on the basis of several main topics including climate crisis, renewable energy integration, energy systems and surrogate models. Second, a problem definition is described, and major sub-problems are identified. Third, research objective and research questions are presented in depth. Based on these questions, the corresponding research approaches are discussed. Finally, a clear outline of the thesis work is shown.

1.1 Background

The climate crisis is being paid more and more attention owing to its notable effects on the environment globally. According to the climate report [1], the global atmospheric carbon dioxide concentrations in 2019 increased to the highest point (409.8 ppm) in the past 800,000 years. Moreover, compared with the previous records, the surface temperature of the land and ocean around the world for January 2020 also reached the peak (2.05°F) as shown in Figure 1.1 [2]. The global warming has many negative influences such as extreme weather events, ice melting, sea-level rise, etc. As reported by the National Climatic Data Centre (NCDC) [3], the global mean sea level in 2019 rose up to 3.4 inches (87.61 mm centimetres), which has been the highest value from the satellite records.

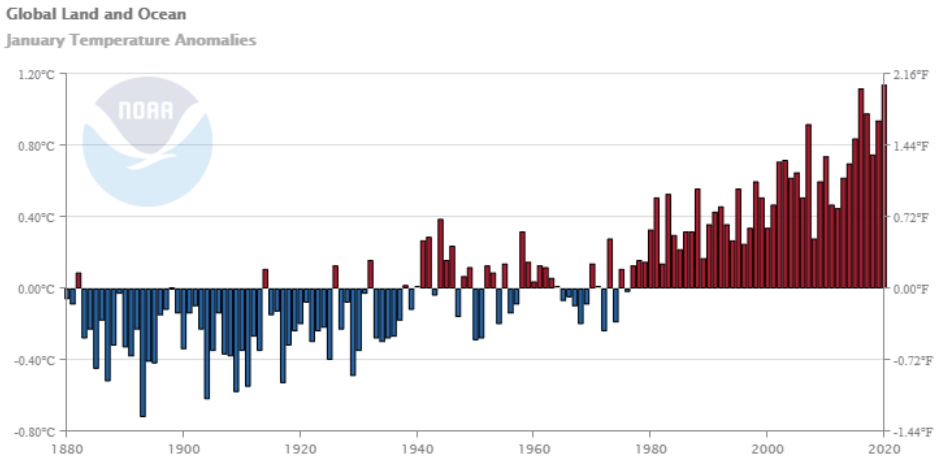


Figure 1.1: The global land and ocean surface temperature in January between 1880 and 2020 [2].

For the sake of tackling the above crisis, the European Union (EU) has developed different strategies. One of which is the “2030 Climate Target Plan” [4] published by the European Commission in Brussel in

September 2020. As illustrated in Figure 1.2, it attempts to achieve a greenhouse gas (GHG) emissions reduction target by at least 55% by 2030. It is worth mentioning that its target has exceeded the previous one, which was set up by the Paris Agreement in 2015, with the aim of reducing the GHG emission by at least 40% by 2030 [5]. Furthermore, the plan contributes to pursuing an ambitious goal with the climate neutrality (net-zero emissions) by 2050. An important reason that the EU takes effort to speed up the process of reducing the GHG emission can be well understood by Figure 1.2 which depicts an obvious trade-off between the GHG emission reduction and GDP growth. The novel decarbonisation goal, according to Ref. [4], is expected to strengthen the whole EU economy which was affected negatively by COVID-19, an unprecedented health crisis worldwide.

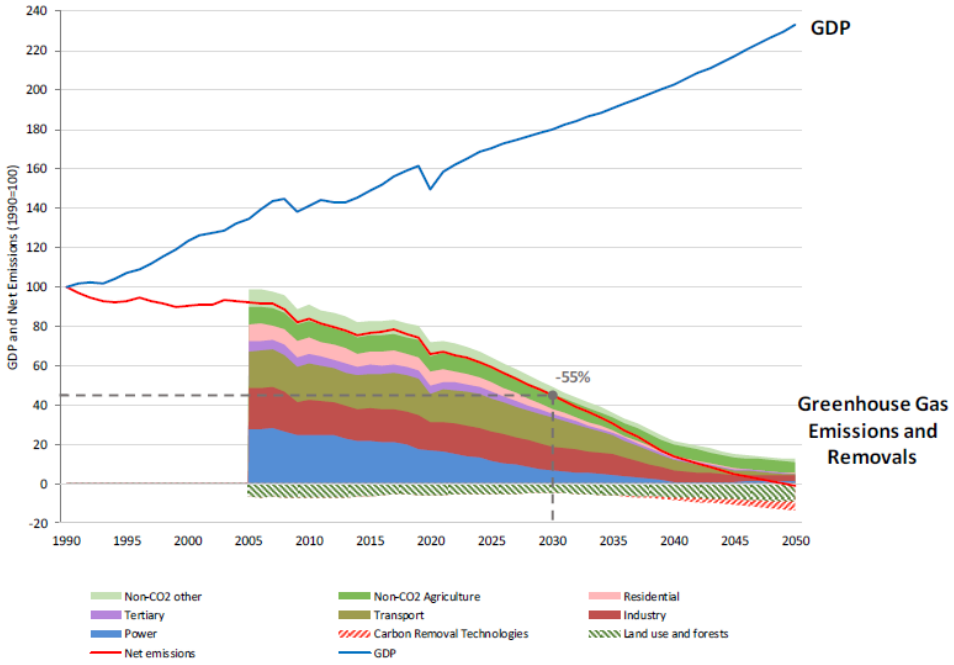


Figure 1.2: The EU's pathway to sustained economic prosperity and climate neutrality, 1990-2050 [4].

Based on the above goal, a list of ideas centred around integrated energy systems has been proposed [6]. Moreover, investments into renewable energy sources (RES), such as wind and solar energy, are supposed to increase so as to reduce the dependency on fossil fuels. To illustrate, the percentage of EU electricity production from RES is desired to go up to nearly 65% [4]. A major challenge of the electricity production derives from the variability and uncertainty of RES that have significant impacts on the grid stability, reliability even economics [7]. To address these problems arising from the increased RES, grid expansion is an effective strategy for many transmission system operators (TSOs) and distribution system operators (DSOs). For instance, a promising solution in the form of sector coupling using Power-to-X (P2X) devices has been proposed [8]. Various benchmarks based on the European network structure, for the sake of assessing the solution, have been studied as described in Ref. [9], [10]. However, the benchmark models provide results that may not be fully applicable to real-world networks. This is because each EU country has their own distribution and transmission networks.

In the context, a generic Dutch grid model is necessary to be established to make up for a deficiency in the availability of the representative energy system in the Netherlands. Traditionally, when conducting the research about energy systems, different energy sectors like electricity are taken into account in an "independent" way, the so-called decoupling methods. However, the methods have apparent drawbacks especially when the integrations of energy sectors become tighter as described in Figure 1.3. For example, the models established by the decoupling methods ignore the interactions between various energy sectors such as electrical vehicles (EVs), biofuels, etc. In consequence, the models cannot represent realistic networks, and the related analysis based on the models is not accurate. Furthermore, apart from the electricity, Figure 1.3 also illustrates the other energy sectors such as heating/cooling,

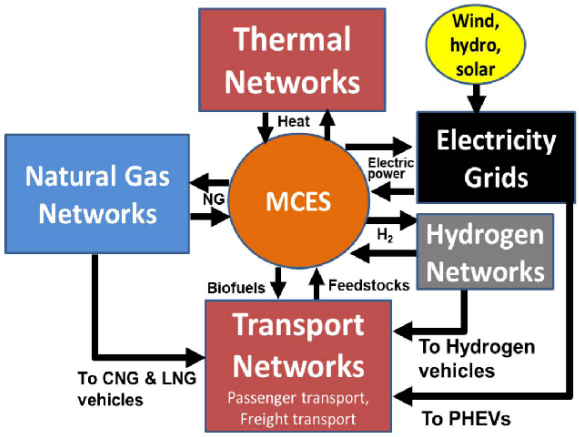


Figure 1.3: MCES as a multigrid component [11].

gas, transport, etc. These sectors account for the vast majority of the GHG emission, and they are more difficult to deal with the decarbonisation issues [12]. Hence, many studies begin to focus on energy systems by taking into account multiple energy sectors, and the relevant definitions have been given such as multi-carrier energy systems (MCESs) [11], multi-energy systems (MES) [13], smart energy systems [14], hybrid networks [15], etc.

1.2 Problem Definition

The thesis work mainly concentrates on **exploring the suitability and applicability of surrogate models for a multi-carrier energy system**. When modelling energy systems, surrogate models have been introduced for the purpose of increasing the computation efficiency [14].There exists an increasing tendency that more detailed models will be considered to narrow the gap between simulation models and realistic models with the progressive research. As a result, the simulation models become more complicated,and the computational burden of simulating such systems increases tremendously. The manifest characteristic of the problems is that it might take minutes to hours, or even days to finish the whole simulation. Therefore, there exists a vital necessity to introduce the surrogate models. In order to tackle the core problem, three sub-problems are presented as follows.

The electrical network is experiencing a notable change towards a decentralised architecture with an increasing fraction of the renewable energy in the Netherlands, as shown in Figure 1.4. As an active member in the EU, the Netherlands is taking effort to increase the utilisation of RES. The high penetration of the RES also poses novel challenges. For example, solar energy sources are vulnerable to weather conditions such as cloudy and rainy days, and thus the sources are intermittent with uncertainties. In order to take on the challenges, different techniques have been proposed and are being researched like EV integration [16], demand response [17], network expansion planning [18], etc. In all of these techniques, one of the indispensable operations is to model electrical networks. Various benchmarks for European networks have been offered in some works [9], [10], however they can only describe the grid structures in general terms without being specific. Within the background, it is imperative to **develop a static electricity network model with the focus on distribution grids representing representative Dutch feeders by means of using open data sources**, which gives us our first sub-problem definition. It is worth mentioning that the distribution network is particularly concentrated in the thesis work because more than 90% of distributed energy resources are installed at DSO level [19]. **This model, as a benchmark, is intended to be openly shared between researchers in the PowerWeb community of TU Delft for further studies**.

However, as mentioned in Section 1.1, it is not enough to consider the single energy sector like electricity to achieve the “2030 Climate Target”. It is necessary to consider more integrations of other energy sectors such as heating/cooling to carry out the holistic analysis of the whole energy system. Therefore, the related definitions like multi-carrier energy systems are introduced, and the modelling approaches

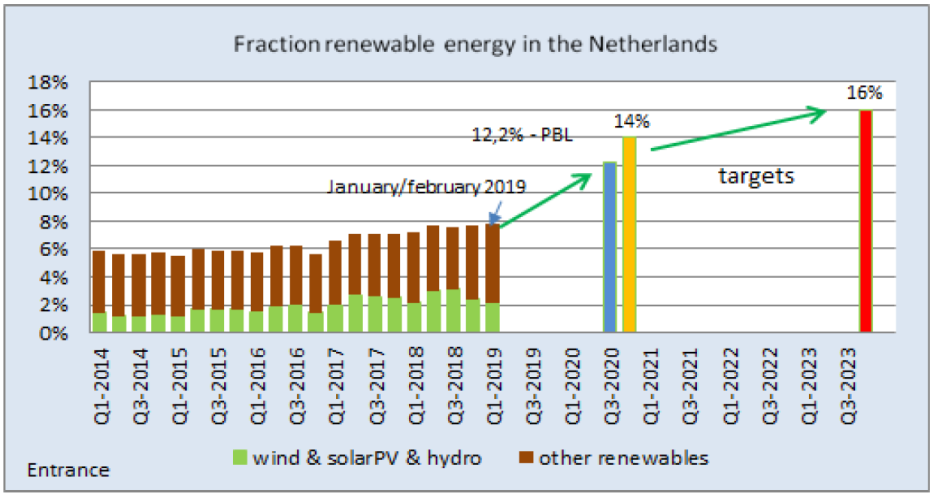


Figure 1.4: Fraction renewable energy in the Netherlands [20].

ought to be improved by using new tools such as multi-domain modelling languages, co-simulation and so forth. Furthermore, control strategies can be added to regulate the behaviour of the whole system. A direct idea of establishing models of the energy systems is to consider the coupling of the district heating network with the electricity distribution network, the so-called multi-carrier distribution networks. This forms our second sub-problem definition, **developing a model of heating distribution network and integrate it with the electrical distribution network to create representative multi-carrier distribution networks**.

Compared with the electrical networks, the multi-carrier energy system model becomes more complex after adding other energy sectors. Owing to the increased coupling between the different energy sectors, computational burden of simulation-based analysis might appear, and simulation time will be longer. For example, if a utility manager needs to forecast PV generation or heat demand of the energy system, it will take a lot of time to conduct a simulation-based analysis. For the sake of resolving the above issues, the applicability of surrogate models is necessary to be investigated. On one hand, the surrogate models are able to avoid the computational complexity introduced by replacing the expensive parts of the original models. As a result, the computation efficiency is enhanced with a faster computation speed. On the other hand, they can quickly assess the impacts of various operating conditions on a tightly integrated system. This offers our third sub-problem definition, **investigating and developing surrogate models best suited for the multi-carrier distribution networks**.

1.3 Objective & Research Questions

The objective of this thesis work is to **develop surrogate models for the multi-carrier distribution networks**. This is achieved by using different machine learning methods to create surrogate models, and their performances are compared with the original models. Moreover, distinct software is applied to carry out the whole simulation, including pandapower [21], OpenModelica [22], EnergySim [23], Python-based libraries (Scikit-learn [24], Tensorflow [25] and Keras [26]).

In order to cope with the above objective, the following research questions and sub-questions are taken into account in the thesis work:

1. **How can a representative multi-carrier distribution network be modelled?**
    - *How can a Dutch electrical distribution network be established as a static power flow model using open data sources?*
    - *How can a representative heating network be designed and modelled to reflect a future tightly integrated distribution energy system?*

2. **What kind of surrogate models are used to create for the multi-carrier distribution networks?**
    - *Which methods are selected to establish surrogate models?*
    - *Which parts of the distribution networks are replaced by surrogate models?*
    - *What are the inputs and outputs of the training/testing data for surrogate models?*
3. **Up to which extent can the simulation performances of the multi-carrier energy distribution networks be enhanced by the surrogate models?**
    - *Which indicators are used to compare the performances between surrogate models and original models?*
    - *Which surrogate models best represent the electrical network?*
    - *Which surrogate models best represent the heating network?*

The contributions of the thesis work are twofold. *First, an electrical distribution network considering three main Dutch distribution system operator areas is established. On the basis of this network, a representative heating network and controllers are taken into account under the environment of the co-simulation. As a result, a representative multi-carrier energy distribution network is modelled. Second, different surrogate models are established, and their performances are compared according to specific indicators. In consequence, the best models are determined in terms of the electrical network and the heating network, respectively*.

1.4 Research Approach

In terms of the first question, the adequate open data sources are determined in the beginning. For example, load profiles can be obtained from Nederlandse Energie Data Uitwisseling (NEDU) [27]. Next, a Dutch electrical distribution network is established. Owing to high shares of VRE resources, PV modules are chosen as one of the classical representatives to add in the above electrical network model. Furthermore, a representative heating network model is built by considering the supply system, the pipe system, the heater & storage system as well as the demand system. In order to resolve the issue introduced by the excess PV power in the electrical network, controllers are designed. Their control strategies are drawn up with considering the electric heaters and storages in the heating network according to rule-based methods [28]. Finally, a co-simulation setup is used to connect the above various networks, and a representative multi-carrier distribution network is developed.

In terms of the second and third question, surrogate models are built with a series of machine/deep learning-based algorithms involving linear regression (LR), linear regression with regressor chains (RC-LR), linear support vector machine with regressor chains (RC-SVR), decision trees (DT), random forest (RF), k-nearest neighbours (k-NN), multi-layer perceptron (MLP) as well as long short-term memory (LSTM). In the thesis work, these methods are adopted to solve regression issues with predicting the multi-output values of the electrical network and the heating network, respectively. Load profiles and PV profiles are used as the inputs in the electrical network, and the outputs are the voltage levels of all the buses in the rural area. Excess PV profiles and demand profiles are employed as the inputs in the heating network. Its outputs are the supply temperature profiles and storage temperature profiles. It should be noted that the methods proposed here are generalised and can be used to predict any other values as required by other users. Furthermore, in order to compare the original models with the surrogate models, prediction accuracy and simulation time are selected as the performance indicators which contain the values of root-mean-square-error (RMSE) and speed-up factor (SUF). Eventually, the best surrogate models are selected for the electrical network and heating network, respectively.

In the thesis work, different simulation platforms are used. The electrical network is modelled by means of pandapower. Both the heating network and controllers are created by OpenModelica. The co-simulation is carried out in the environment of EnergySim. The machine/deep-based surrogate models are simulated via different frameworks containing Scikit-learn and Tensorflow. The programming languages of the above models are all in Python.

1.5 Outline

The remaining part of the thesis work is organised as follows. Chapter 2 offers a literature review based on the previous works. Chapter 3 shows an electrical distribution network with considering PV modules, and three various areas are considered based on main Dutch DSO networks. The modelling of a multi-carrier distribution network is described in Chapter 4 which details the establishment process of a representative heating distribution network and controllers, respectively. Different machine/deep learning algorithms with the modelling process of surrogate models are depicted in Chapter 5. A detailed result analysis is presented in Chapter 6. Finally, Chapter 7 reaches conclusions and outlines the future work.

2

Literature Review

The objective of the chapter is to give a literature review based on previous works, and it is divided into three aspects. The first aspect concentrates on the modelling of electrical distribution networks. The second aspect focuses on multi-carrier energy systems including the related definitions and simulation methods. The third aspect puts the emphasis on surrogate models containing their applications in energy systems and creating approaches. Furthermore, machine/deep learning algorithms are focused as an emerging approach with their characteristics and applications in energy systems.

2.1 Establishment Methods of Electrical Distribution Networks

Traditionally, electrical distribution networks are considered in the perspective of passive electricity dispensation without considering automation or digitisation [19]. However, the “fit and forget” method cannot perform well especially when renewable energy sources are introduced in the networks. Hence, new approaches ought to be proposed. One of the important works is offered by the Joint Research Centre (JRC) which is the Commission’s science and knowledge service [29]. The organisation gives comprehensive overviews of the European electricity distribution networks with solutions by the Distribution System Operators Observatory project launched in 2016 and 2018, respectively [19], [30]. Based on the 2016 observatory, the project comes up with 13 representative distribution networks containing 3 large-scale networks and 10 feeder type networks. Hence, these models can be used to establish the corresponding networks directly. Based on the 2018 observatory, the project offers an alternative with giving 37 DSOs indicators which are key technical input parameters when modelling the networks. These indicators can be categorised in three aspects: network structure & reliability, network design as well as distributed generation. Furthermore, DiNeMo is proposed as a platform for the distribution network simulation based on a subset of the above indicators [31]. Table 2.1 shows the subset of the total DSOs indicators in detail.

Table 2.1: Subset of the total DSOs indicators used to build the large-scale representative distribution networks [31].

ID	DSO Indicators
1	Number of LV consumers per MV consumers
2	LV circuit length per LV consumer
3	LV underground ratio
4	Number of LV consumer per MV/LV substation
5	MV/LV substation capacity per LV consumer
6	MV circuit length per MV supply point
7	MV underground ratio
8	Number of MV supply points per HV/MV substation
9	Typical transformation capacity of MV/LV secondary substations in urban areas
10	Typical transformation capacity of MV/LV secondary substations in rural areas

There also exist other works regarding the modelling of the distribution networks. A European benchmark based on a low voltage microgrid network and its extension containing three types of loads are developed by Papathanassiou et al. in Ref. [10]. The authors concentrate on the network itself without putting more emphasis on micro-sources and control strategies. Based on the benchmark, a series of researches is carried out to explore various technologies such as demand side management [32], peer-to-peer energy trading [33], frequency response analysis [34], etc. Moreover, a work in Ref. [35] establishes Dutch medium voltage networks and explores the impact of future residential loads, which gives meaningful references for the network planning. Ref. [36] develops a benchmark of the low voltage distribution grids in the Netherlands. Three different grid types, including net-zero emission building (NZEB) areas, urban areas and rural areas, are taken into account based on main Dutch DSOs.

2.2 Multi-carrier Energy Systems

Multi-carrier energy systems have received a lot of attention due to high penetrations of renewable energy and electrification. As illustrated in Figure 2.1, the energy systems consider not only the electricity but also other energy sectors like natural gas, heat/cooling, transport, etc. The precise definition has been offered by O'Malley et al. [11] who present an exhaustive review based on the system.

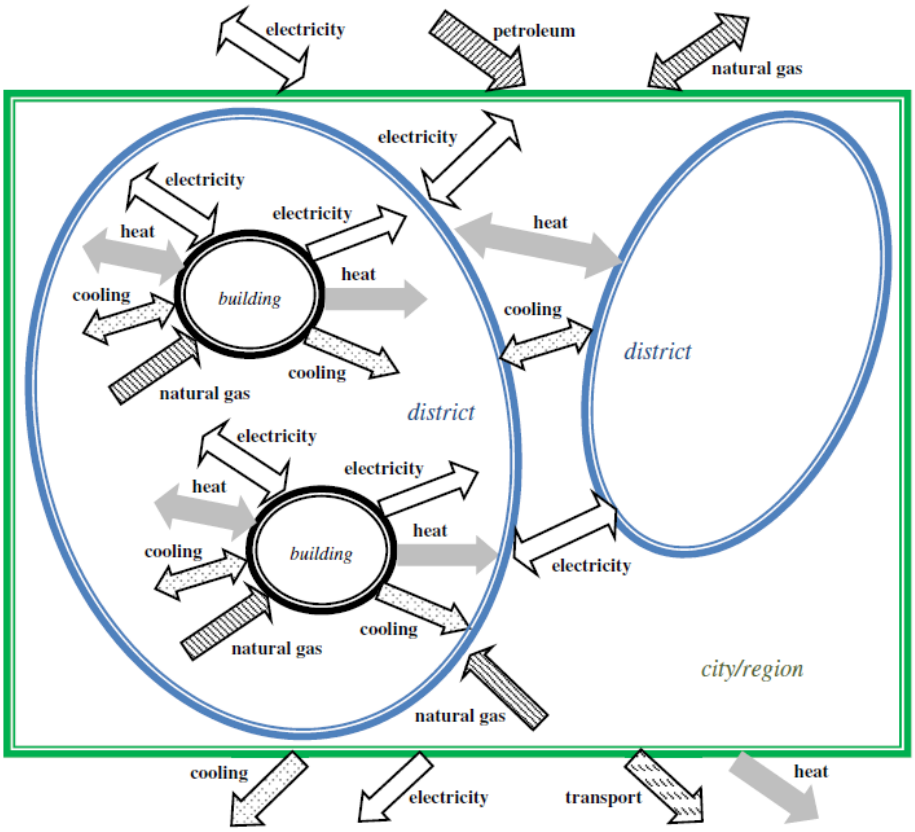


Figure 2.1: Schematic illustration of the spatial perspective concept [13].

Other analogous definitions and terms are also proposed and used in many existing works. For instance, Mancarella [13] gives a comprehensive overview of multi-energy systems involving their models and detailed analysis methods. The same term is used in the work by Chertkov and Andersson [37] who offer an exhaustive overview of the latest developments and opportunities. A work by Lund et al. [14] conducts an excellent review based on smart energy systems containing their definitions and applications, and the integration of the storage is also discussed. A holistic approach is developed by Widl et al. [15] who undertake technical and economic assessments of hybrid thermal-electrical distribution grids. It has further been extended by Ref. [28] which considers a design approach to the optimisation methods and various control strategies.

2.2.1 Modelling and Simulation Methods

Based on the above definitions, a series of modelling and simulation methods is discussed in the literature. A thorough review related to the approaches is presented by Subramanian et al. in Ref. [38]. As seen in Figure 2.2, energy system models are classified into three main branches according to their works.

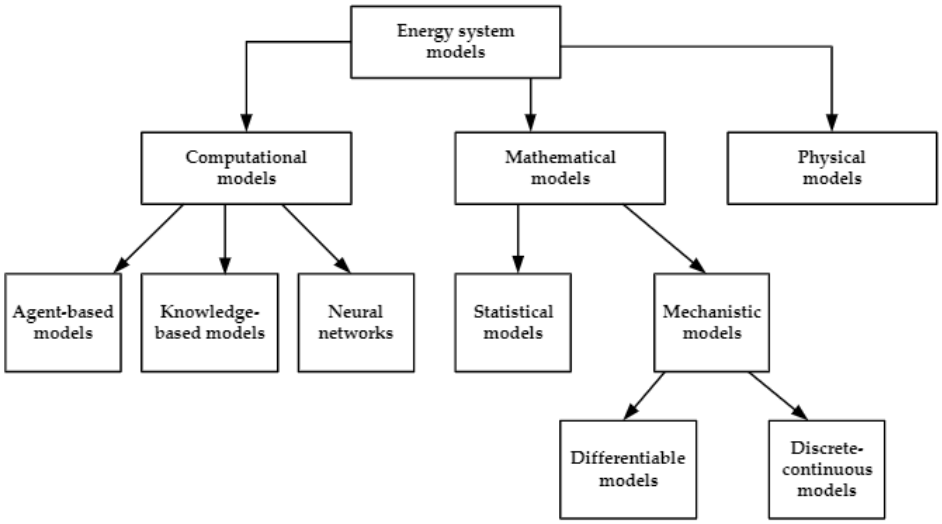


Figure 2.2: Classification of energy system models according to modelling approach [38].

To be specific, the computational models can mimic the intelligent behaviour [39] with different expert systems consisting of agent-based models, knowledge-based models and neural networks. The mathematical models are divided into statistical (black-box) and mechanistic (white-box) models. In terms of the statistical models, a list of techniques has been employed such as self-optimising variables [40], kriging [41] and so forth. In light of the mechanistic models, one typical example is the energy hub model developed by Geidl and Andersson [42]. Its systematic overview, based on its concept potentials and challenges, is provided by Ref. [43], and it is further studied in combination with the urban climate influence on the energy demand by Perera et al. in Ref. [44]. However, as reported by Ref. [45], many studies including Ref. [42] lack the information related to the electricity grid status in the modelling. For the sake of solving the issues, one feasible solution is adopted by Arnaudo et al. [45] who focus on employing an effective tool called co-simulation which has been widely adopted in the literature; see e.g., Ref. [46], [47].

Co-simulation, whose structure is described in 2.3, performs the function in coupling multiple software packages and/or models [48]. It overcomes the barrier that many tools are restricted to simulate a single energy domain system, but also transcends the limitation existing in the specific energy system analysis software like EnergyPLAN without considering the control effects [22]. In addition, as described in Ref. [28], it has more advantages than multi-domain tools such as Modelica, Simulink and MATLAB. Moreover, various frameworks have been established to achieve the function of co-simulation. Mosaik [49], as an open source, is used to carry out the simulation of active components in smart grids. Nonetheless, it is not easy for non-programmers owing to the dependence on simulator APIs [23]. The drawback is overcome by the usage of Functional Mock-up Units (FMUs), and two typical representatives are embodied in VirGIL [50] and Mastersim [51]. Another framework is GridSpice [52], on the basis of cloud systems, which resolves the difficulties existing in the modelling of transmission and distribution systems via the co-simulation by High Level Architecture (HLA) [53]. EnergySim, originally named as FMUWorld, is proposed to model energy systems in different domains as shown in Ref. [23] with the advantage of the multi-time step simulation. Other frameworks contain the Functional Mock-up Interface (FMI) specification [54], System-in-the-Loop (SITL) [55] and so forth.

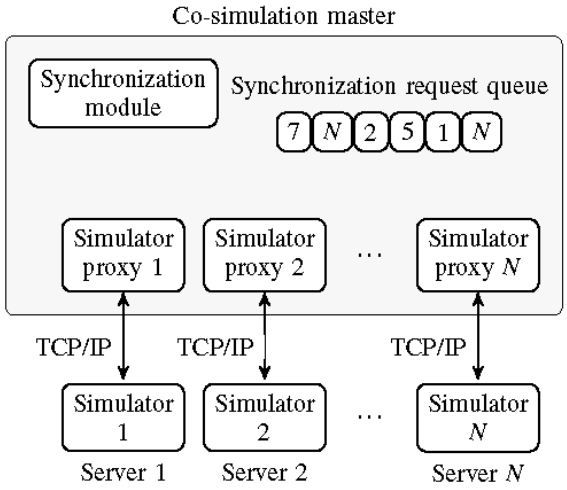


Figure 2.3: Structure of the co-simulation environment [48].

The co-simulation application in energy systems, including hybrid thermal-electrical distribution networks, is well documented in many studies. Ref. [15] builds the network by considering various energy sectors via a FUMOLA tool as their co-simulation platform. Furthermore, a technical-economic assessment is carried out to put forward the recommendation from the perspective of operation and strategy, respectively. In Ref. [45], the authors extend their works by designing performance indicators such as the grid capacity, and distributed heat-pumps are involved in the whole system. Similar design ideas are undertaken by Ref. [47] and Ref. [28]. The former focuses on the technical assessment of power-to-heat use cases, whereas the latter concentrates on the novel control and optimisation strategies. However, these works lack attention to the computation time of the simulation.

2.2.2 Computationally Expensive Problems

The computationally expensive problems are likely to occur in multi-carrier energy systems, and one of the main reasons is the inter-dependencies between various energy sectors, dynamic operation of components within individual energy domain, etc. The tight couplings between technologies and networks in the energy system are shown in Figure 2.4.

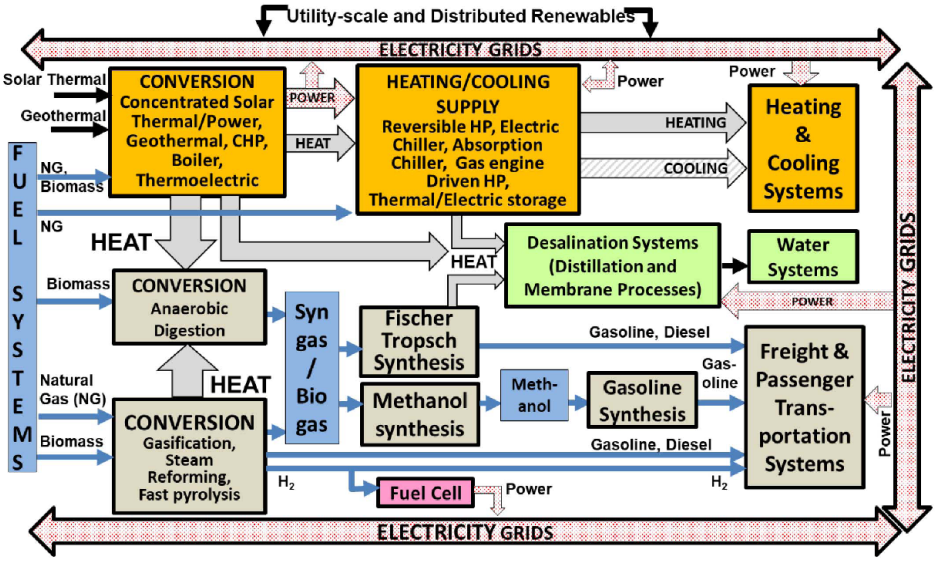


Figure 2.4: Interactions between MCES technologies and networks [11].

It can be seen that various sectors in these systems are taken into account for integrations involving electricity, heat, gas, transportation, among many others. It has existed some works which highlight the expensive problems in the energy field. The authors in Ref. [28] point out the evaluation of objective functions has high computation costs (minutes to hours or even days) in the simulation-based design for the energy system. Ref. [56] indicates that models of the power grid rely on complicated calculations, and hence the simulation costs will be high. Ref. [57], [58] figure out the computationally expensive issues might be introduced when Modelica is used to carry out simulations. Modelica is an equation-based language, and its models may include plenty of non-linear equation systems [59].As a result, the simulation speed is reduced apparently. Similar ideas can be found in Ref. [51], [60]. To illustrate, the authors in Ref. [51] state that carrying out Modelica-only based building simulations is not realistic. Owing to a large quantity of codes in Modelica, it will be time-consuming especially when modelling large buildings.

2.3 Surrogate Models

Surrogate models, also referred to as meta-models or approximate models, have been the focus of a large body of research over the years [39], [61]. The models can be viewed as classical black-box models which concentrate on the inputs and outputs without considering the complicated transformation process. They are mainly introduced to cope with computationally expensive issues with high computation costs.

Table 2.2: Representative methods creating surrogate models in the application of energy systems.

Creating Methods of Surrogate Models	**Applications**
Kriging [41]	Prediction based on the behaviour of the offshore power plants integrating a wind farm to drop the computation difficulty of the optimisation process
Response surface methodology [62]	Covering the uncertainty of forecast loads to reduce the complexity of stochastic electricity grid operations problems
Radial basis functions [63]	Estimation of the system outputs to decrease computation costs when simulating the building systems
Support vector machine (ML algorithm) [64]	Training data based on feasible trajectories to assist the optimisation process of forecasting the multi-period flexibility for LV prosumers
Artificial neutral networks (ML algorithm) [65]	Using transfer learning with hybrid optimisation algorithms to decrease the computation time when optimising the energy systems

2.3.1 General Approaches for Creating Surrogate Models

There are a variety of methods of creating surrogate models, which are summarised and compared by Diaz-Manriquez et al. in Ref. [39]. Table 2.2 provides a brief summary based on the representative methods establishing surrogate models in the application of energy systems.

One of the conventional methods is called response surface methodology (RSM) [62] whose goal is to get the minimum variance value of the responses. Kriging, also known as Gaussian processes (GPs), is another alternative to build surrogate models with various engineering applications [41]. In Ref. [63], the authors apply the radial basis function network in the building energy systems and explore the possibility in high dimensions applications. Machine learning (ML) algorithms also play an indispensable role in creating surrogate models to reduce the computation time. Specifically, the algorithms include a number of approaches, such as support vector machine (SVM) [64] and artificial neural networks (ANN) [65], which are applied in various occasions with their different characteristics.

2.3.2 Applications in Energy Systems

When modelling energy systems, surrogate models can be added in several ways to enhance the computation efficiency. One straightforward idea is to employ the models to replace a part of the simulation models in the whole system directly. In Ref. [66], a dynamic equivalent of the complex distribution grid is created to replace expensive parts in the original network. As a result, the dynamic effects in power systems are simulated and studied in depth. The similar approach is carried out by Balduin et al. in Ref. [67]. The authors carry out the research of the smart grid, and the entire low voltage grid from Ref. [10] is replaced by a deep neural network used as surrogate models. Here, the surrogate models show excellent performances in the aspects of the speed-up factors, simulation time as well as accuracy. However, they fail in taking into account other machine learning methods, and their models have poor performances in the domain-specific experiment. Ref. [68] extends their works and considers more methods of creating surrogate models. Furthermore, the performances of the various methods are compared, and the best models are selected. However, surrogate models in their work are only limited to the application of the electrical network.

Another common technique using surrogate models in energy systems is to assist in the optimisation of complex problems, the so-called surrogate-assisted algorithms [61]. When one optimisation algorithm creates a potential solution, surrogate models can be created to test whether it is feasible. The testing process is accelerated because it is not required to run the whole simulation and then calculating the objective function. Therefore, surrogate models contribute to finding optimal solutions for the algorithms with a faster speed.

There also exist some works which focus on reducing the surrogate modelling process. To illustrate, Ref. [69] presents a tool called MeMoBuilder to assist the surrogate modelling process for three typical simulation models in energy systems consisting of the battery model, photovoltaic plant and fuel cell.

2.4 Machine Learning Algorithms

As mentioned in section 2.2, machine learning (ML) algorithms are one of the popular approaches when creating the surrogate models. First, an introduction related to ML algorithms, including deep learning algorithms, is given. Second, common data splitting approaches in the algorithms are concentrated. Finally, the applications of the algorithms based on the energy field are summarised.

2.4.1 Basic Definitions

Machine learning algorithms play an important role in the artificial intelligence (AI) technology, and

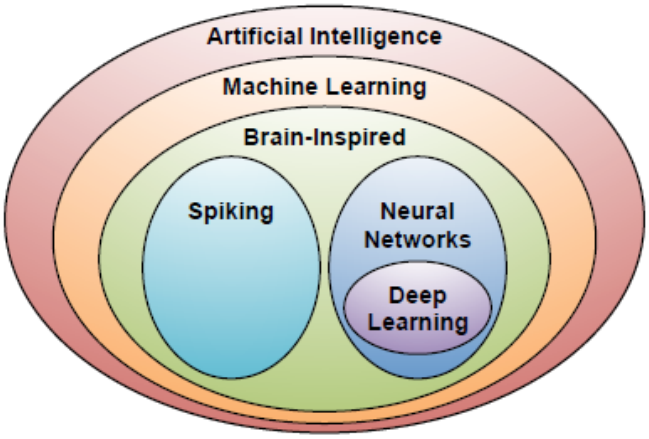


Figure 2.5: Machine/deep learning in the context of artificial intelligence [70].

Figure 2.5 depicts their relationships intuitively. ML algorithms have been experiencing an explosive growth in the research recently. One of their advantages is systems are offered the ability of self-learning without using the complicated programming. Compared with other methods, the algorithms have excellent performances when performing the big data analytics [71] and predictions [72]. One of the applications can be well reflected in the predictive modelling which develops a model employing the historical data to make a prediction on the new data [73]. It includes the classification predictive modelling and regression predictive modelling. They have a similar function with estimating the mapping function from the input variables to output variables [74]. However, their outputs have notable differences that the classification modelling is discrete, whereas the regression modelling is continuous.

Deep learning (DL) algorithm is a subset of the ML algorithm, which is also described in Figure 2.5. As a brain-inspired approach, DL algorithm employs a multi-layered structure called neural networks. Similar as human brains, the neutral networks in the algorithm perform the function of utilising the data to identify patterns and classify various information [75]. Ref. [76] offers some obvious distinctions between the DL algorithm and other traditional approaches in the AL algorithm. For example, compared with other methods in the machine learning algorithms, the model in the DL algorithm can be trained directly by an “end-to-end” way. Moreover, it can handle multi-dimensional tensors. According to Ref. [77], modern DL algorithm puts more emphasis on training deep (many layered) neural network models by the backpropagation algorithm, and main techniques consist of multilayer perception networks, convolutional neural networks and long short-term memory recurrent neural networks.

2.4.2 Data Splitting Approaches

In the machine learning algorithms, a dataset is usually split into the training dataset and the testing dataset initially. Hence, it is necessary to choose the adequate data splitting approach.

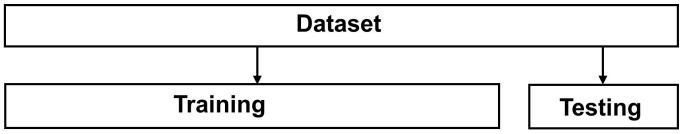


Figure 2.6: The description of the hold-out method.

The hold-out method, also named as the train-test split method, is one of the popular splitting approaches. In this method, an original dataset is directly divided into two subsets including a training dataset and a testing dataset. The training dataset performs the function of extracting features and training to fit a model, and the testing dataset is used to make predictions on the basis of the model obtained from the training dataset. A common splitting percentage is 80% of the dataset for the training data and the remaining 20% of the dataset for the testing data [78]. Although the percentage is fixed, simulation results are still different when carrying out multiple experiments when using the same percentage. It is not difficult to understand because the same dataset is split randomly in each experiment. As a result, various training datasets and testing datasets are created in each experiment. Figure 2.6 shows the hold-out method.

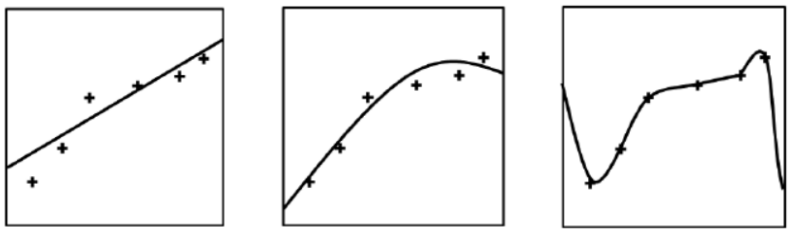


Figure 2.7: Regression example: (a). underfit; (b). good fit; (c). overfit [79].

The above method is easy to implement in the algorithm. However, it suffers from the weakness that it might encounter the overfit issue as shown in Figure 2.7(c). It can be seen that its fitting performance is even better than Figure 2.7(b) which is defined as the “good fit”. Ref. [79] describes one of the characteristics of the overfitting that it fits too closely to the training set, and hence it performs excellently on the training dataset. However, it does not mean it also performs well when predicting the new data. To illustrate, the model in the case is likely to learn the “noise” existing in the dataset without focusing on the variable relationships of the dataset. However, the noise does not belong to the new data. Hence, the case is not an ideal fitting method and ought to be avoided.

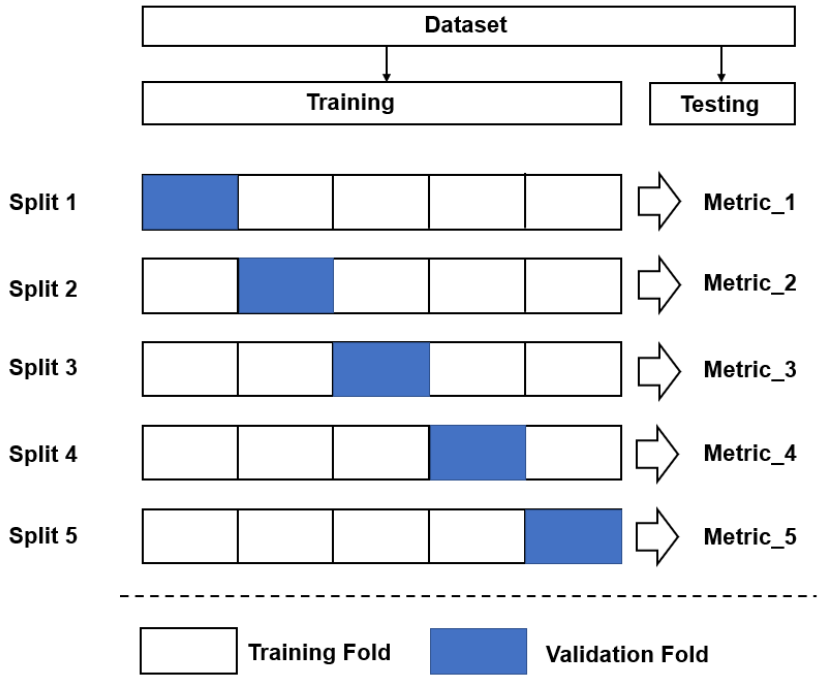


Figure 2.8: An example of the 5-fold cross-validation method [80].

Cross-validation method is introduced to avoid the above overfitting issue, and one popular technique in the method is called as the k-fold cross validation. Compared with the above hold-out method, the dataset is split into k different subsets (or folds) randomly. One of the subsets is employed as the validation fold, and the remaining folds are the training folds. Thus, each subset has one validation fold and k-1 training folds. In each subset, a model is trained on the training subset initially, and a metric value such as RMSE is calculated on the validation subset. As a consequence, K metrics can be obtained for all the subsets. Then, the model can be tested on the test set. Figure 2.8 illustrates an example of the method when k is equal to 5.

2.4.3 Research Directions in Energy Systems

In recent years, ML algorithms and DL algorithms also attract the attention in the energy field research. In Ref. [81], a list of ML algorithms for energy systems is systematically reviewed by Mosavi et al. Here, the authors highlight the distinct ML models and focus on their applications in different energy domains. Moreover, ten ML models are discussed and compared in detail.

One of the major topics in the field of energy systems is energy forecasting which plays a key role in the field of power systems and business from the perspective of planning and operation [72]. Typical forecasting targets involve electricity demand, electricity prices, heating load, renewable energy generation, etc. Figure 2.9 illustrates the increasing researches with publications of the forecasting based on the targets. In Ref. [82], a scientometric overview based on the electricity demand prediction is carried out with a visualisation analysis by Yang et al. Weron undertakes a thorough review of the electricity

price forecasting and offers the

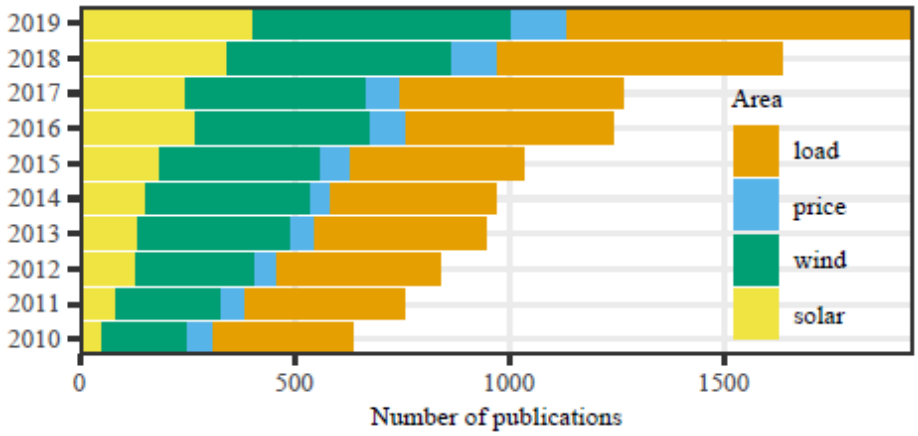


Figure 2.9: Number of publications in load, price, wind and solar forecasting returned by respective Scopus search [72].

future research directions [83]. Some studies that cover the heating load prediction are conducted in Ref. [84], [85]. Sweeney et al. [86] present a good review whose topic is renewable energy forecasting with a focus on wind and solar energy. A comprehensive work by Yagli et al. [87] considers 68 ML and statistical models for hourly solar forecasting.

3

The Electrical Network Model

The objective of this chapter is to model a Dutch electrical network. First, a general introduction of system operators is given, and the background of TSOs and DSOs in Europe is concentrated. Furthermore, a distribution related to these system operators in the Netherlands is concentrated in detail. Finally, a representative Dutch LV distribution model is established with the consideration of RES integration.

3.1 The Background of System Operators

System operators play a key role in guaranteeing the electricity delivery to end users such as consumers, industry, etc. The background of two essential operators including transmission system operators and distribution system operators is illustrated.

3.1.1 TSOs in Europe

The European Network of Transmission System Operators (ENTSO-E) association, including 42 members and 1 observer member from different countries, has been created to liberalise the European gas and electricity markets [88]. Tennet, as a unique Dutch TSO, contributes to European grid connections [89]. Figure 3.1 describes the role of Tennet with three important services [89] in the electrical network. First, it guarantees the extra high- and high-voltage electricity transmission. Second, it offers system services by keeping the balance

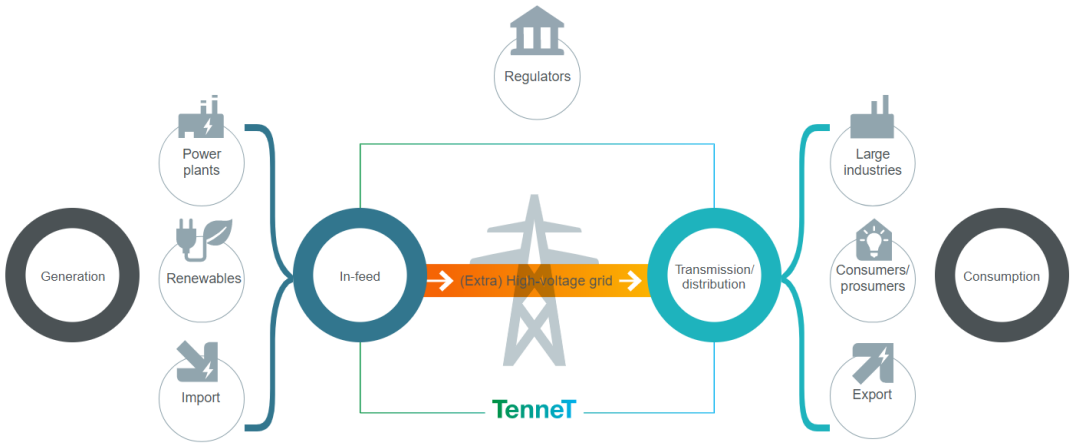


Figure 3.1: The role of Tennet in the electricity network [89].

between generation and consumption. Furthermore, it facilitates the energy market by providing grid

access to the market players such as consumers, prosumers, regulators, etc. As discussed in the problem definition of Chapter 1, transmission network modelling is not the main work in the thesis. Hence, the detailed Dutch transmission grid structures with the relevant design of simulation parameters are described in the Appendix section.

3.1.2 DSOs in Europe

Distribution system operators, as shown in Figure 3.2, play an important role in the connection between TSOs and end customers in the traditional network scenario. In Europe, there exist approximately 98-99% of the customers whose electricity networks are connected with the distribution networks [19]. The responsibilities of the DSOs include planning, maintenance and management of the distribution networks [90]. In terms of the planning, a list of plans is drawn up for the system and its elements to

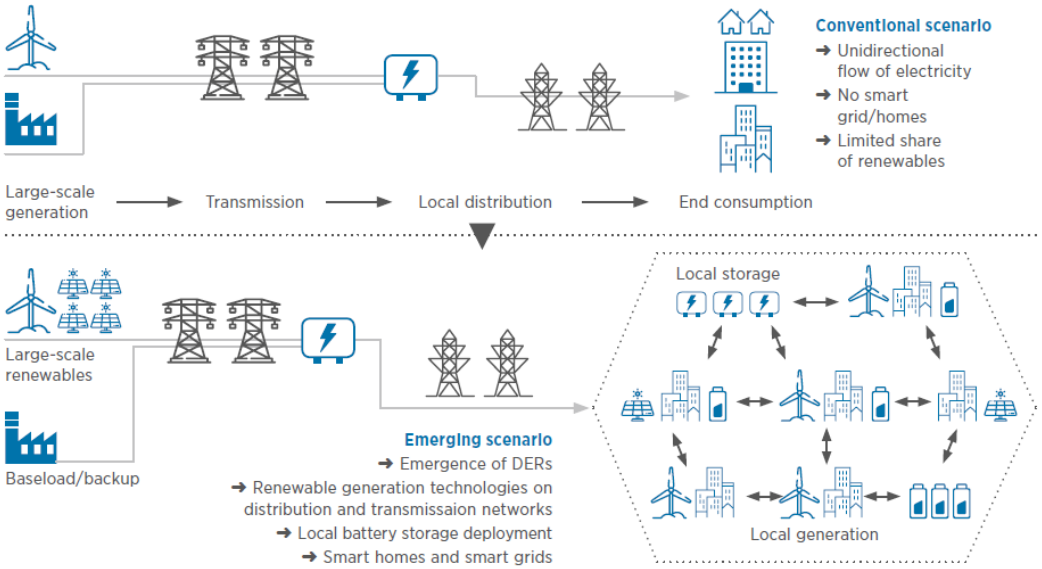


Figure 3.2: Conventional scenario versus emerging scenario in the power system due to the emergence of distributed energy resources [91].

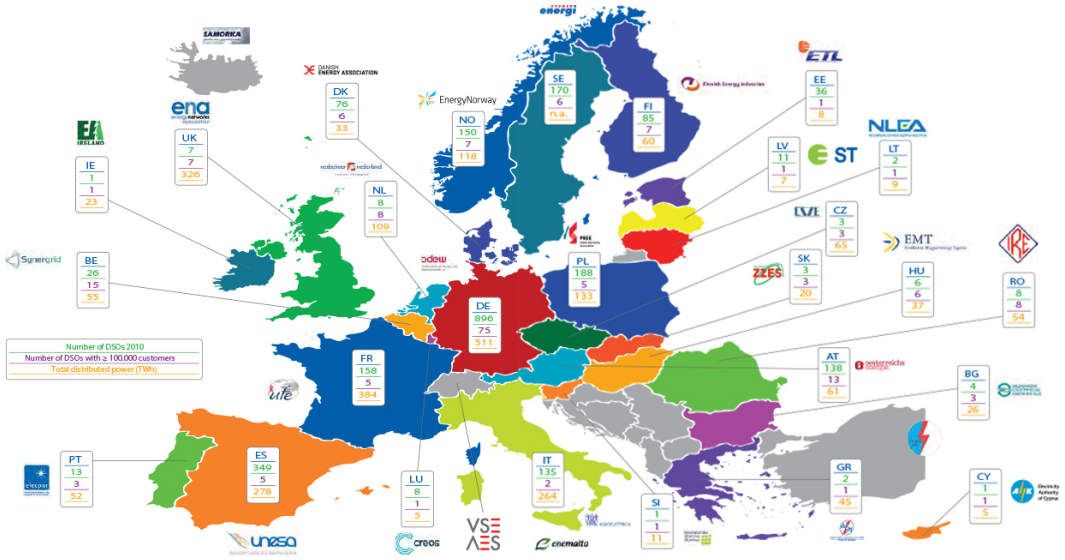


Figure 3.3: The distribution of DSOs in Europe [90].

meet the future development after collecting the grid data [92]. In terms of the maintenance, DSOs perform the function of delivering the electricity to end consumers. For instance, faults will be cleared as fast as possible if they occur in the electrical networks. In terms of the management, DSOs manage not only different voltage levels (HV, MV, LV) of the distribution systems but also the supply outages [91]. Furthermore, the energy of distributed sources is absorbed by DSOs who are responsible for their connection and disconnection.

The roles of DSOs are experiencing a notable variation considering the introduction of distributed energy sources (DERs) and novel market players [93] such as prosumers, market regulators, retail suppliers, microgrid operators, etc. Therefore, they have new responsibilities to meet these challenges in the emerging scenario depicted in Figure 3.2. In light of DERs introduction, DSOs are supposed to guarantee that flexibility services [91], such as peak load management and congestion management, are offered to reduce the redundancy issues from the DERs. With respect to the new market players, take retail suppliers as an example, the cooperation between these suppliers and DSOs ought to be enhanced to bring considerable benefits for customers [93]. To illustrate, when the suppliers offer high-quality electricity supply, DSOs are expected to guarantee their stability and security accordingly.

As illustrated in Figure 3.3, the number of DSOs varies significantly from country to country. For instance, Ireland and Slovenia have only one distribution system operator. Countries like Czechia and Greece exist several DSOs. Countries such as Germany, Spain and Poland possess more than 100 DSOs. Owing to the differences, it is necessary to take into account local situations when modelling the distribution networks for one specific country.

3.2 The Modelling of the LV Distribution Network in the Netherlands

In this section, distribution network models in the Netherlands are established based on main Dutch DSOs with their representative areas. Therefore, some backgrounds on the Dutch DSOs are provided in the beginning. Next, the simulation tool and open data sources used in the thesis work are described. Furthermore, the detailed distribution network with the design of simulation parameters are given. Finally, renewable energy sources are studied by taking into account PV modules on the basis of the established distribution network.

3.2.1 The Distribution of Main Dutch DSOs

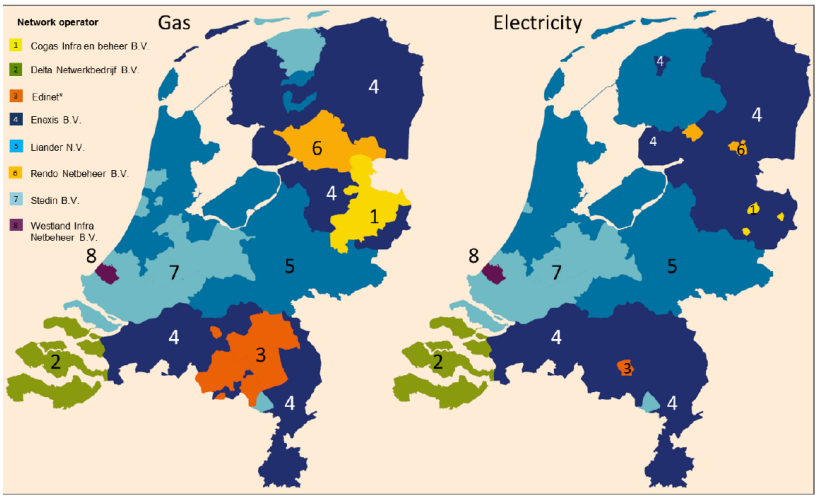


Figure 3.4: The energy distribution network operators in the Netherlands [36].

Figure 3.4 depicts energy (gas and electricity) distribution network operators in the Netherlands. It can be seen that Enexis, Liander and Stedin are three major Dutch DSOs in the electricity network. In Ref. [36], the authors build a distribution grid network focusing on realistic networks in the Netherlands. Furthermore, three types of the distribution grids are studied based on the above three Dutch DSOs. The first type is the residential area with nearly zero energy buildings, which belongs to the realistic grid of Liander. The other two types are the old urban residential area and the rural residential area which are owned by Stedin and Enexis, respectively. In the thesis work, a similar network structure is employed.

Figure 3.5: The NZEB in Rijsdijk, Etten-Leur [94].

It is worth mentioning that NZEB areas, as one of the representatives, are necessary to be concentrated due to EU requirements and local policies. One of the requirements proposed by the Renovate Europe campaign in the EU. They work towards a main goal to decrease the energy demand of the building blocks by 80% by 2050 [95]. In order to achieve the above goal, the energy consumption of the blocks is anticipated to be reduced substantially, and the definition of NZEB is introduced. A representative policy in the Netherlands regulates that all new construction, both residential and non-residential, ought to meet the requirements for Bijna Energieneutrale Gebouwen (BENG) from 1 January 2021. There has been a rapid growth of these buildings in the EU, and Figure 3.5 illustrates an example of NZEB based on Rijsdijk in the Netherlands. The building uses PV panels, a heat pump with the ground source and solar thermal collectors. The heat from the ground storage during the summer is utilized in the winter time [94].

3.2.2 The Simulation Tool & Open Sources

Pandapower [21], as an open source Python tool, is applied to carry out the static analysis, quasi-static analysis and optimisation of the electrical power system. Compared with other calculation tools in the power system, it overcomes their limitations in the specific application as illustrated in Table 3.1. Moreover, its advantage lies in conducting the symmetric distribution system analysis [21]. Thus, pandapower is deployed to model the static power flow of the distribution network in the thesis work.

Open sources can offer reliable parameters of electrical components and various profiles (i.e., generators, loads, etc.) in the distribution network. Electrical components are obtained from design handbooks and the standard simulation settings from pandapower, whereas different profiles such as generators and loads are identified and processed from Nederlandse Energie Data Uitwisseling [27] which is an important data platform for combining various energy sectors in the Netherlands. Additionally, the parameters related to RES involve penetration rates obtained from Ref. [36] and PV profiles accessed from the Open Power System Data Platform [96] that offers free data for power system modelling.

Table 3.1: Comparison of open source element model libraries [21].

	MATPOWER 6.0	PYPOWER 5.1.2	PSAT 2.1.10	OpenDSS 7.6.5	PyPSA 0.10	GridCal	GridLAB-D 3.2	**pandapower 1.4.3**
ZIP-load			√	√		√	√	√
Line	√	√	√	√	√	√	√	√
2-Winding Transformer (π)	√	√	√	√	√	√	√	√
2-Winding Transformer (T)				√	√		√	√
3-Winding Transformer			√	√			√	√
DC Line	√		√	√	√		√	√
Local Switches								√
Voltage Controller Generator	√	√	√	√	√	√	√	√
Static Load/Generation	√	√	√	√	√	√	√	√
Shunt	√	√	√	√	√	√	√	√
Asymmetrical Impedance								√
Ward Equivalents								√
Storage Unit				√	√		√	

3.2.3 Network Structures & Simulation Parameters

Figure 3.6 depicts the structures of a Dutch LV distribution network designed in the thesis work. The designed parameters of each part used in the above network are described in detail. In order to give a clear illustration, Table 3.2 lists the abbreviation meaning of the parameters used in the later table.

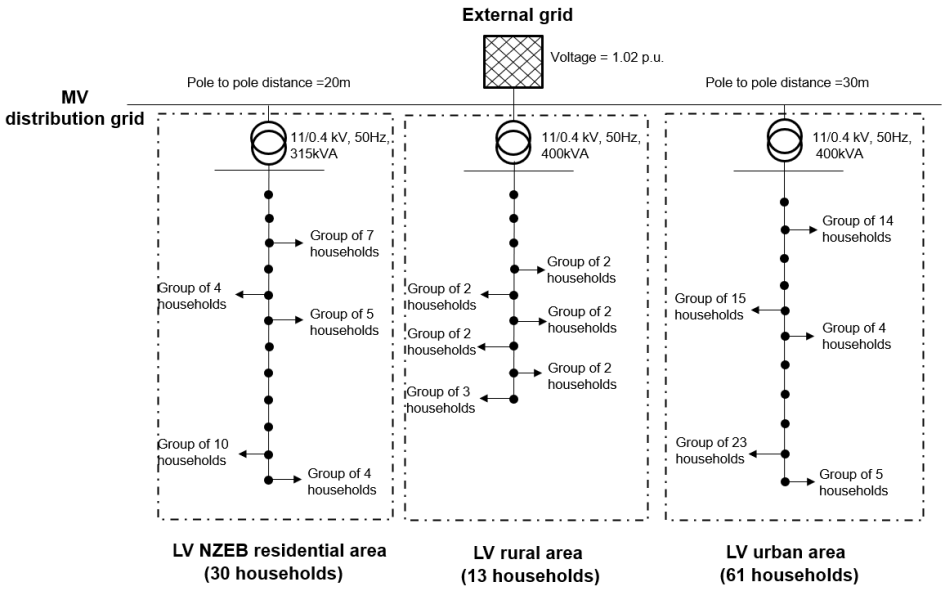


Figure 3.6: The LV distribution network in the Netherlands.

Table 3.2: Abbreviation meaning of used parameters.

Parameter	Meaning of Each Parameter
vm	voltage magnitude at the slack node in per unit
va	voltage angle at the slack node in degree
hv	The bus on the high-voltage side of the transformer
lv	The bus on the low-voltage side of the transformer
sn	rated apparent power
vh	rated voltage on high voltage side
vl	rated voltage on low voltage side
vk	relative short-circuit voltage
vkr	relal part of relative short-circuit voltage
pre	iron losses in kW
i0	open loop losses in percent of rated current
vn	voltage magnitude of each bus
va	voltage degree of each bus
l	line length in km
r	line resistance in ohm per km
x	line reactance in ohm per km
c	line capacitance in nano Farad per km
imax	maximum thermal current in kilo Ampere
p	active power
q	reactive power

The external grid is used to replace the connection of the network to higher voltage levels. Table 3.3 depicts its simulation parameters.

Table 3.3: The simulation parameters of the external grid.

Index	Name	Bus	vm (pu)	va (degree)	In Service
0	grid connection	0	1.02	0	True

Table 3.4: The simulation parameters of different buses.

Index	Name	vn (kv)	va (degree)	In Service
0	Bus0	11	0	True
1	Bus1	0.4	0	True
2	Bus2	0.4	0	True
3	Bus3	0.4	0	True
4	Bus4	0.4	0	True
5	Bus5	0.4	0	True
6	Bus6	0.4	0	True
7	Bus7	0.4	0	True
8	Bus8	0.4	0	True
9	Bus9	0.4	0	True
10	Bus10	0.4	0	True
11	Bus11	0.4	0	True
12	Bus12	0.4	0	True
13	Bus13	0.4	0	True
14	Bus14	0.4	0	True
15	Bus15	0.4	0	True
16	Bus16	0.4	0	True
17	Bus17	0.4	0	True
18	Bus18	0.4	0	True

Table 3.5: The simulation parameters of transformers.

Index	Name	hv	lv	sn (mva)	vh (kv)	vl (kv)	vk (%)	vkr (%)	pre (kw)	i0 (%)
0	Trafo1	Bus0	Bus1	0.315	11	0.4	4.0	1.25	0.8	0.2385
1	Trafo2	Bus0	Bus7	0.4	11	0.4	4.0	1.325	0.95	0.2375
2	Trafo3	Bus0	Bus14	0.4	11	0.4	4.0	0.325	0.95	0.2375

Transformers play an important role in transferring voltage levels from the medium voltage (11 kV) to the low voltage (400 V). In the thesis work, transformers are designed for different residential areas (NZEB areas, rural areas and urban areas). They have the same transformation level (11/0.4 kV) but different rated power values. The transformers in the rural area and the urban area have the same values (400 kVA), whereas the value of the transformer in the NZEB area is 315 kVA. Moreover, as mentioned above, two main voltage levels (MV for 11kV and LV for 400V) of different buses are considered. Table 3.4 and Table 3.5 depict designed parameters of the different buses and transformers, respectively.

Similar as the design of the transformers, different cable lines are taken into account for three residential areas. Their parameters are acquired from the design handbooks [97], [98] and the Ref. [36]. The underground copper cable is selected for the NZEB area. Its type is NYY 4*35, and the total length

Table 3.6: The simulation parameters of cable lines.

Index	Name	Type	From	To	l (km)	r (ohm/km)	x (ohm/km)	c (nf/km)	imax (ka)
0	LV Line0.1	NYY 4*35	Bus1	Bus2	0.055	0.524	0.083	205	0.159
1	LV Line0.2	NYY 4*35	Bus2	Bus3	0.04	0.524	0.083	205	0.159
2	LV Line0.3	NYY 4*35	Bus3	Bus4	0.02	0.524	0.083	205	0.159
3	LV Line0.4	NYY 4*35	Bus4	Bus5	0.08	0.524	0.083	205	0.159
4	LV Line0.5	NYY 4*35	Bus5	Bus6	0.02	0.524	0.083	205	0.159
5	LV Line1.1	NAYY 4*50	Bus7	Bus8	0.03	0.642	0.083	210	0.142
6	LV Line1.2	NAYY 4*50	Bus8	Bus9	0.02	0.642	0.083	210	0.142
7	LV Line1.3	NAYY 4*50	Bus9	Bus10	0.02	0.642	0.083	210	0.142
8	LV Line1.4	NAYY 4*50	Bus10	Bus11	0.02	0.642	0.083	210	0.142
9	LV Line1.5	NAYY 4*50	Bus11	Bus12	0.02	0.642	0.083	210	0.142
10	LV Line1.6	NAYY 4*50	Bus12	Bus13	0.02	0.642	0.083	210	0.142
11	LV Line2.1	NYY 4*70	Bus14	Bus15	0.06	0.268	0.082	220	0.232
12	LV Line2.2	NYY 4*70	Bus15	Bus16	0.09	0.268	0.082	220	0.232
13	LV Line2.3	NYY 4*70	Bus16	Bus17	0.03	0.268	0.082	220	0.232
14	LV Line2.4	NYY 4*70	Bus17	Bus18	0.12	0.268	0.082	220	0.232
15	LV Line2.5	NYY 4*70	Bus18	Bus19	0.03	0.268	0.082	220	0.232

is 215m. The same cable material is used in the urban area. However, its type is NYY 4*70 with a total length of 330m. The length of the cable line in the rural area is chosen as 130m with the aluminium material, and its type is NAYY 4*50. Table 3.6 shows the simulation parameters of the cable lines between two adjacent buses in various areas.

Table 3.7: The eight categories of the NEDU profiles [99].

Profile	>	≤	other	#	#used
E1A	-	3x25A	single tariff	205	101
E1B	-	3x25A	night tariff	222	40
E1C	-	3x25A	single tariff	81	34
E2A	3x25A	3x80A	evening tariff	71	32
E2B	3x25A	3x80A	double tariff	369	99
E3A	3x80A	100 kW	UT2000h	75	32
E3B	3x80A	100 kW	UT3000h	36	18
E3C	3x80A	100 kW	UT3000h	17	5

Table 3.8: The simulation parameters of loads at time=0.

Index	Name	Bus	p (mw)	q (mvar)	Scaling	In Service
0	load0	2	8.72E-04	2.88E-04	1.0	True
1	load1	3	4.99E-04	1.65E-04	1.0	True
2	load2	4	6.23E-04	2.06E-04	1.0	True
3	load3	5	1.25E-03	4.11E-04	1.0	True
4	load4	6	4.99E-04	1.65E-04	1.0	True
5	load5	8	2.49E-04	1.55E-04	1.0	True
6	load6	9	2.49E-04	1.55E-04	1.0	True
7	load7	10	2.49E-04	1.55E-04	1.0	True
8	load8	11	2.49E-04	1.55E-04	1.0	True
9	load9	12	2.49E-04	1.55E-04	1.0	True
10	load10	13	3.74E-04	2.32E-04	1.0	True
11	load11	15	1.74E-03	8.38E-04	1.0	True
12	load12	16	1.87E-03	8.97E-04	1.0	True
13	load13	17	4.99E-04	2.39E-04	1.0	True
14	load14	18	2.87E-03	1.38E-03	1.0	True
15	load15	19	6.23E-04	2.99E-04	1.0	True

In the structure of the network, a total of 104 households with 15 aggregated loads are taken into account. They belong to three typical residential areas which are owned by main Dutch DSOs. Load profiles are attained by NEDU. There is a lot of data in the platform, and hence it is a key procedure of processing the data used in pandapower. Table 3.7 illustrates eight categories of the NEDU profiles. Similar as the study in Ref. [99], the profile values in E1A are chosen in the thesis work. The active power values in the load profiles of the single customer i are extracted from the NEDU data based on the equation 3.1 provided in Ref. [100]:

$$\widetilde{v_i}(t) = AEC_I \cdot \widetilde{w_i}(t) \tag{3.1}$$

Here AEC is the annual electricity consumption in kWh, and w(t) shows the normalised measurement series in h^{-1}. The value of AEC is nearly 3500 kWh per capita in the Netherlands according to Ref. [101], and the values of w(t) are equal to the profile values in E1A based on the annual electricity data in 2019.

The reactive power values in the load profiles are calculated on the basis of the active power values. The formula is depicted in the following equation:

$$Q = P \cdot \tan \phi \tag{3.2}$$

Here ϕ is the phase angle. To be specific, $\cos \phi$ is defined as the power factor. In the thesis work, the values of power factors are 0.95, 0.85 and 0.9 for the NZEB area, the rural area and the urban area, respectively.

It should be noted that the time resolution of the profile values in E1A is in 15 mins. Hence, the time resolution of calculated profiles of each load is also in 15 mins. Table 3.8 illustrates calculation results, using the above method, based on the electricity profiles of 15 aggregated loads. The time period in the table is chosen as 0:00-0:15 a.m. on Jan 1, 2019 with its timestep is noted as 0. The electricity profiles based on other timesteps can be calculated and set up via a similar procedure in pandapower.

3.2.4 Model Validation by Reference Values

As discussed in the literature review, JRC offers holistic overviews of the European electricity distribution networks with solutions by the Distribution System Operators Observatory project launched in 2016 [30] and 2018 [19]. In the thesis, the model from Ref. [36] is used, and it is validated against the reference values offered in Ref. [31].

Table 3.9: The Dutch DSO indicators and their values in the designed network.

ID	DSO Indicators	Actual Values in the Designed Network
1	Number of LV consumers per MV consumers	104
2	LV circuit length per LV consumer (km/LV consumer)	Minimum (0.02), Maximum (0.12)
3	LV underground ratio (%)	Minimum (19.3), Maximum (48.9)
4	Number of LV consumer per MV/LV substation	Minimum (13), Maximum (61)
5	MV/LV substation capacity per LV consumer (kVA/LV consumer)	–
6	MV circuit length per MV supply point	–
7	MV underground ratio (%)	–
8	Number of MV supply points per HV/MV substation	–
9	Typical transformation capacity of MV/LV secondary substations in urban areas	400
10	Typical transformation capacity of MV/LV secondary substations in rural areas	400

Table 3.10: DSOs indicators and reference values [31].

ID	Indicators	Average Value	Median Value	Min Value	Max Value
1	Number of LV consumers per MV consumers	671	401	22	1946
2	LV circuit length per LV consumer (km/LV consumer)	0.03	0.025	0.012	0.16
3	LV underground ratio (%)	66	75	11	99
4	Number of LV consumer per MV/LV substation	86	76	17	230
5	MV/LV substation capacity per LV consumer (kVA/LV consumer)	4.76	3.88	2.1	13
6	MV circuit length per MV supply point (km/MV supply point)	1.06	1.04	0.54	1.77
7	MV underground ratio (%)	59	61	10	100
8	Number of MV supply points per HV/MV substation	155	127	33	460

Table 3.11: Typical Transformation capacity indicators and reference values [31].

ID	Indicators	Common Values
9	Typical transformation capacity of MV/LV secondary substation in urban areas (kVA)	400,630,1000
10	Typical transformation capacity of MV/LV secondary substation in urban areas (kVA)	50,100,250,400,630

Table 3.9 illustrates 10 key Dutch DSO indicators, offered by DSO Observatory Project 2018 [19], when designing MV and LV distribution networks. Their values, calculated based on Figure 3.6, are also listed in the table. Moreover, the reference values are determined according to the EU level and listed in the corresponding tables as illustrated in Table 3.10 and Table 3.11. It can be observed that the actual values used in the designed network are in the same range as the reference values, and hence the model used in the thesis is in accordance with the JRC reference.

3.2.5 RES Integration in the Network

The above distribution network does not take into account the penetration of RES. In the thesis work, in order to study the effect of the RES integration, PV modules are added on the basis of the distribution network. Especially, the rural area is selected as a use case to assess the simulation performance. The number of households with PV modules in the rural area can be calculated according to the PV penetration rate [36] as shown in Table 3.12. Figure 3.7 shows the structure of the distribution network after considering the penetration of RES, which extends the original rural area network in Figure 3.6.

Table 3.12: Overview of the assumptions of PV panels for the three areas [36].

Area	Households	PV Penetration Rate	Households with PV
NZEB area	30	100%	30
Rural area	13	35%	5
Urban area	61	19%	12

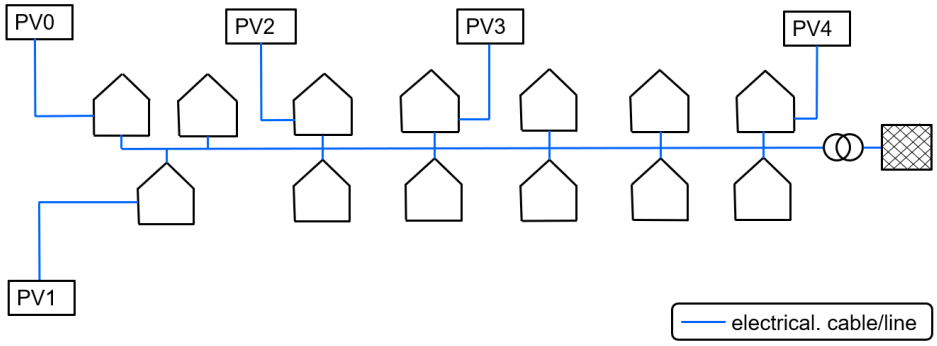


Figure 3.7: The Dutch LV distribution network of the rural area considering the PV integration.

A PV module can be modelled as one ‘static generator’ in pandapower. PV profiles are offered based on the typical Dutch rural data from the Open Power System Data Platform [96]. Similar as the load profiles, the obtained PV profiles are in 15-min time resolution. Table 3.13 depicts the simulation parameters of PV modules using in the simulation when its timestep is equal to 0. The PV profiles based on other timesteps can be calculated by a similar procedure.

Table 3.13: The simulation parameters of PV modules at time=0.

Index	Name	Bus	p (mw)
0	sgen0	8	0
1	sgen1	10	0
2	sgen2	11	0
3	sgen3	13	0
4	sgen4	13	0

Having established the structure and the above data (inputs) in the electrical network, the simulation can be carried out in pandapower. As mentioned earlier, the simulation time is based on one year. Considering the data (csv format) of load profiles and PV profiles is in 15-min time resolution, the number of timesteps is 35040 in one year. In consequence, the outputs (voltage profiles) in 15-min time resolutions will be obtained after undertaking the simulation.

4

The Multi-carrier Energy System Model

The objective of the chapter is to describe a representative multi-carrier energy system in detail. First, simulation tools are given to offer the adequate simulation platforms used in the thesis work. Second, a heating network is designed and parameterised. Then, it is coupled with the electrical network developed in the last chapter using electric boilers. Third, controllers with a rule-based operation strategy are developed, and they play an essential role in controlling the energy system. Eventually, the applied co-simulation methodology is depicted in depth.

4.1 Simulation Tools

OpenModelica, based on the object-oriented Modelica language, is an open-source tool for system modelling and simulation. Compared with other simulation languages, it has notable advantages in acausal modelling, multi-domain modelling and so forth [22]. Moreover, models in the tool are easy to be created, connected and simulated by means of a user-friendly graphical interface called OpenModelica Connection Editor (OMEdit). Furthermore, Openmodelica offers various libraries such as fluid components, control systems, etc. Hence, OpenModelica is selected to simulate models of the heating network and controllers. The representative libraries for simulating the heating network contain IBPSA, DisHeatLib, AixLib, etc. The Modelica IBPSA library [102] is a free open-source library employed to simulate buildings and district energy systems. On the basis of the IBPSA library, the Modelica DisHeatLib library [103] is developed and focuses on modelling the district heating network. The AixLib library [104] is also a free open-source Modelica library which concentrates on building performance simulations. In the thesis work, the AixLib library is used since it is the only one that works well with the established model in OpenModelica, whereas others are found to not be fully compatible with OpenModelica.

4.2 The Modelling of the Heating Network

Figure 4.1 illustrates the structure of a representative heating network which is designed according to the work in Ref. [28]. Following the idea of treating the rural area as a use case when studying the RES integration, the rural area with 13 households is also focused and considered as a use case in the heating network. There are four main systems in the heating network: supply system (a thermal plant), pipe system (thermal flow pipes and return pipes), heater & storage system (3 subsystems of electric heaters and thermal energy storages) as well as demand system (13 households). It is worth mentioning that the demand system is classified into six different groups,which are highlighted in Figure 4.1, owing to similar heating profiles in each group. Based on the above design methods and examples provided by libraries of OpenModelica, the heating network is modelled, and its OpenModelica structure is shown in Figure 4.3 where the heater & storage system and the demand system are packaged. Furthermore, it can be seen that different temperature profiles are obtained by focusing on supply temperatures and storage temperatures in the thesis work. They are measured by considering the

temperatures of flow pipes and the layer of storage tanks, respectively. Due to three heater & storage subsystems are designed, three groups of supply temperatures and storage temperatures are acquired.

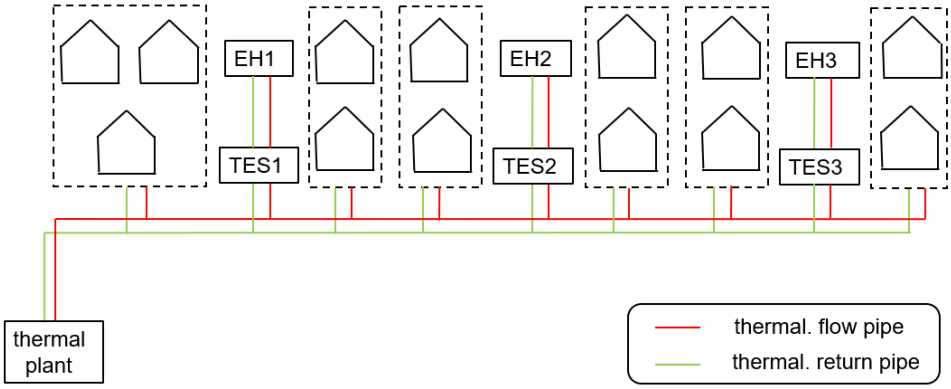


Figure 4.1: The Dutch heating grid of the rural area (EH, electric heater; TES, thermal energy storage). The dash boxes highlight six different groups of the households in the demand system.

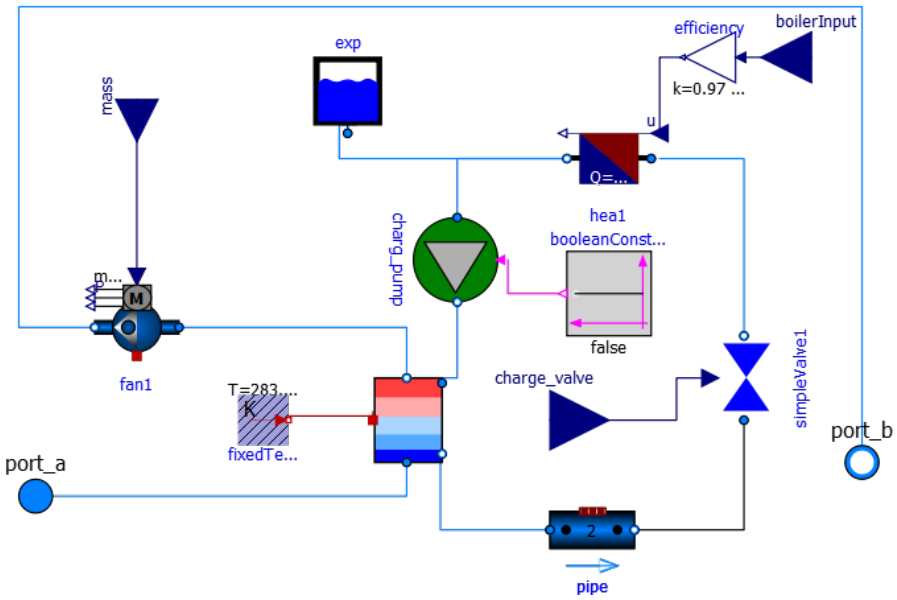


Figure 4.2: The structure of one subsystem of the heater & storage system model in OpenModelica.

In the heater & storage system, three heater-storage subsystems are designed, and each of them connects the corresponding pipes. Figure 4.2 illustrates the detailed structure of the heater and storage based on one subsystem. The inputs of the subsystem are three external signals (boiler input, charge valve on/off, mass flow rate). The electric boiler connects to the controller input of the heater directly after converting from the electrical energy to the heat energy with a fixed efficiency (0.97 in the thesis work). Then, the water flows into the storage tank via a charging pump. The charge valve on/off signal is linked to a simple valve. If the signal is in the charging state, water from the storage tank will flow back to the heater via the pipe. The mass flow rate signal offers a rate value to a fan whose efficiency is computed based on the efficiency and pressure curves [104].

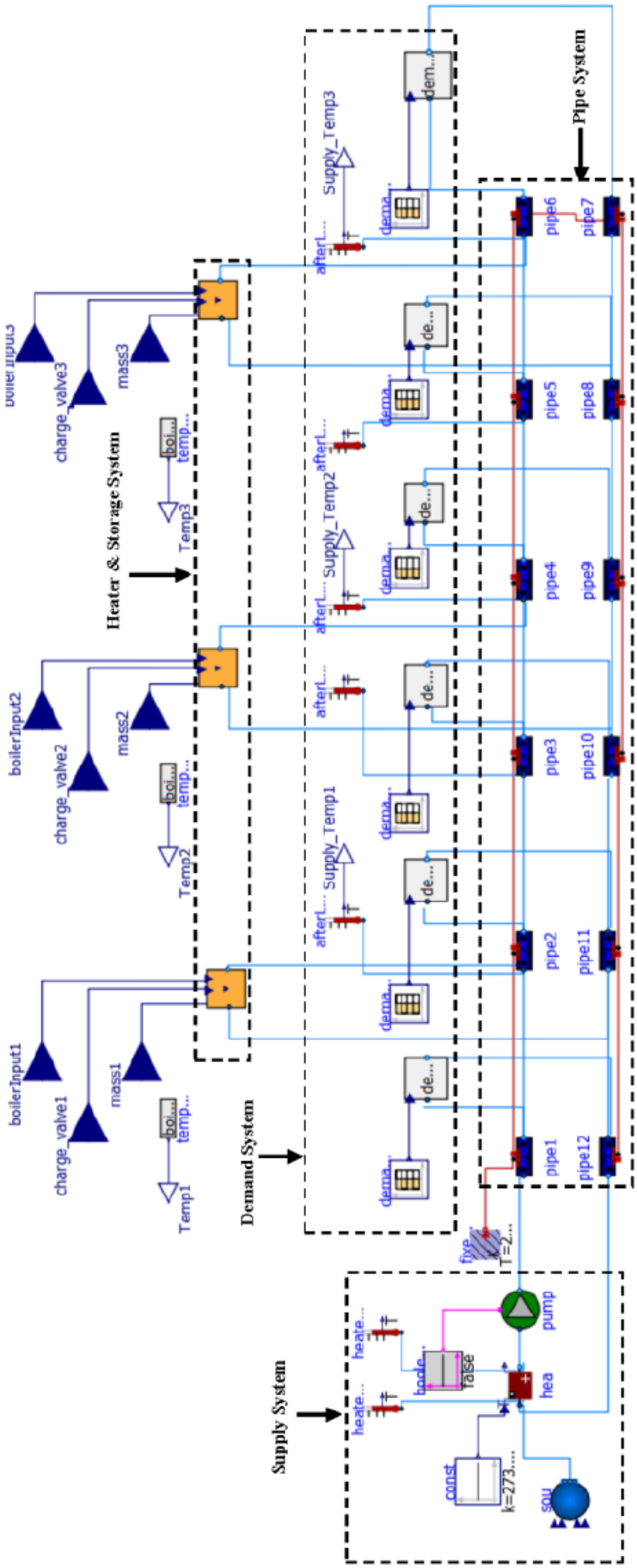


Figure 4.3: The structure of the Dutch heating network based on the rural area in OpenModelica. Four main systems are highlighted in dash boxes.

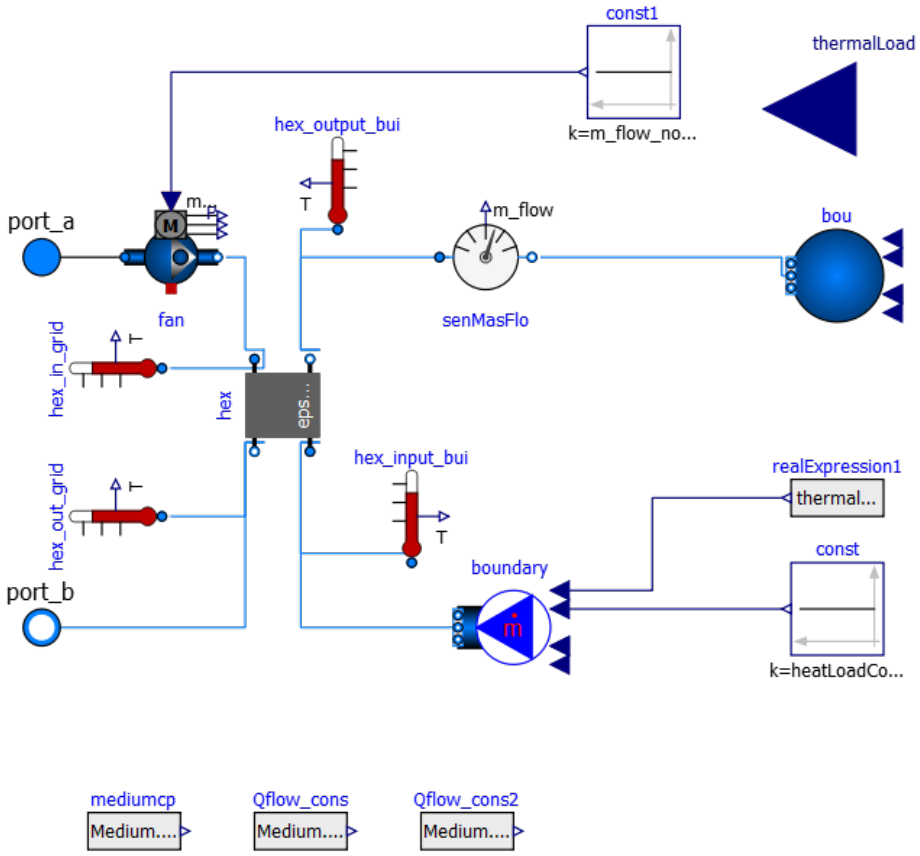


Figure 4.4: The structure of one subsystem of the demand system model in OpenModelica.

Table 4.1: The design of parameters using in the heating network.

Systems	Parameters	Values	Units
Supply System	Fixed value of pressure	306000	Pa
	Pressure difference	400	Pa
	Maximum heat flow rate for heating (positive)	200000	W
Pipe System	The length of pipe1	200	m
	The length of other pipes	20	m
	Return temperature setpoint	55	degC
	Hot water temperature setpoint	70	degC
Demand System	Nominal mass flow rate	0.05	kg/s
	Fixed value of pressure	103125	Pa
	Heat exchanger effectiveness	85%	-
	Peak load demand	2000000	W
	Electric boiler efficiency	97%	-
	Heat flow rate	200000	W
Heater & Storage system	Nominal mass flow rate	0.1	Kg/s
	Pressure difference	200	Pa
	Diameter of circular pipe	0.05	m

In the demand system, six demand subsystems are considered as mentioned earlier, and Figure 4.4 depicts its structure based on one subsystem. The relevant demand (thermal load) profiles can be collected from Open Power System Data Platform [96], NEDU [27], Modelica DisHeatLib Library [103], etc. In the thesis work, the DisHeatLib Library is finally selected because its thermal data (space

heating and the domestic hot water) has the same time resolution (15-min resolution) as the data in the electrical network. One of the important components in the subsystem is the heat exchanger with a constant effectiveness. The formula related to the heat transfer can be depicted as follows:

$$Q = Q_{max} \cdot \varepsilon \tag{4.1}$$

where ε is a constant effectiveness, and Q_{max} is the maximum heat that can be transferred.

Table 4.1 depicts the major parameters applying in the heating network. As mentioned earlier, they are obtained from the related design handbooks, reference values from OpenModelica examples and typical rural area systems, etc.

4.3 The Modelling of Controllers

In the electricity network, PV modules are added to study the renewable energy integration. However, one of the main problems which cannot be ignored is the excess power from the PV modules, the so-called PV surplus issue. For example, when the PV modules produce too much power in specific weather conditions, the generated power values will be far more than the consumed power values. In consequence, power imbalances are likely to occur with frequency variations, which are harmful to system components even lead to the blackout for the whole system. [105].

In order to deal with the above issue, system-level controllers are proposed to utilise the extra power from the PV modules. They consist of two typical types: rule-based and model-based control schemes. Ref. [28] offers an explicit methodology of the controllers. In the work, the rule-based controllers are selected because a list of control rules is necessary to be established for tackling the PV surplus issue. Furthermore, the controllers are case-specific with the expert knowledge, and they follow a series of rules which is described as follows:

$$(\phi_1 \rightarrow \psi_1) \wedge \ldots \wedge (\phi_i \rightarrow \psi_i) \tag{4.2}$$

Here, ψ_1 and ϕ_i represent a certain system state and a control action of the system inputs, respectively.

Operational strategies are used to describe control rules of the system controllers. The strategies in the thesis work are formulated by taking into account the working states of the electric heaters and thermal storage tanks in the heating network. They can be depicted by the following two algorithms. Algorithm 1 and Algorithm 2 describe the charging state and the discharging state of the heating network, respectively.

Algorithm 1: The charging state of the heating network.

if *(a1). excess PV power is produced* ***and*** *(b1). the temperatures in the storage tanks of their top layers are lower than 95 °C* **then**
  The heater&storage system is in the charging state;
else
  The storage tanks do not charge;
end

Algorithm 2: The discharging state of the heating network.

if *(a2). the district heating demand is sufficient* ***and*** *(b2). the temperatures of the top layers of storage tanks are higher than the supply temperature of the district heating* ***and*** *(c2). the supply temperature of the district heating is more than 55 °C* **then**
  The heater&storage system is in the discharging state;
else
  The storage tanks do not discharge;
end

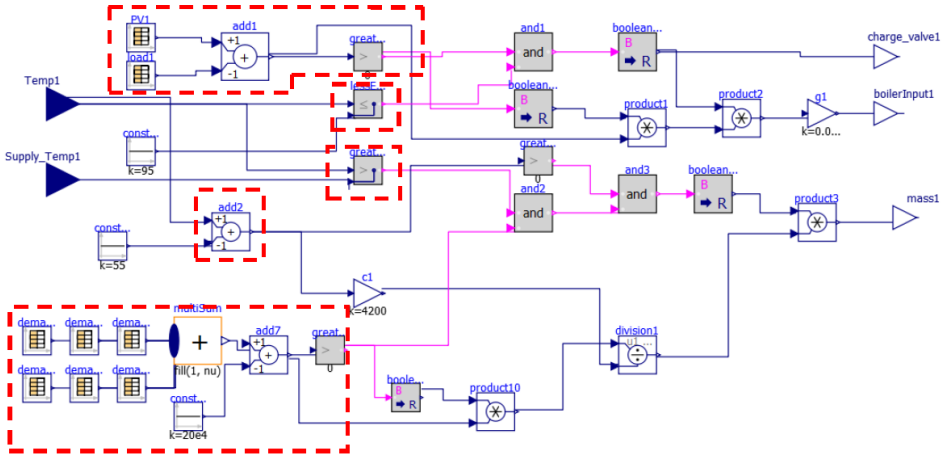


Figure 4.5: The structure of one subsystem of the controller model in OpenModelica. The modules related to the control rules are highlighted in red.

The controllers are modelled in OpenModelica with offering control signals to three heater & storage subsystems. Considering each subsystem has the same control signals, one heater & storage subsystem is zoomed with its structure to give a clear illustration as shown in Figure 4.5. The rules based on the charging state and the discharging state are highlighted. In terms of the charging state, on/off signals of the charge valve and the boiler input signal are considered as outputs to indicate the charging state. In terms of the discharging state, the mass flow rate value is used as the only output to decide whether the model is in the discharging state, and it is calculated by the following formula:

$$m = \frac{Q}{c \cdot \Delta t} \tag{4.3}$$

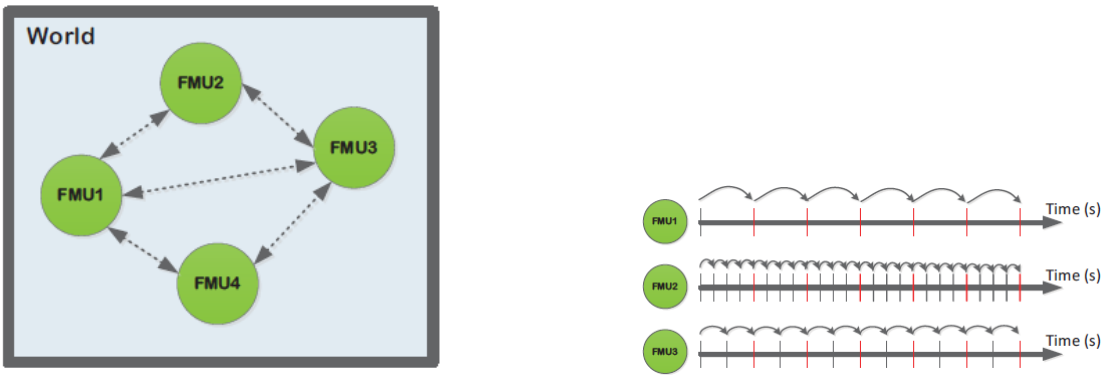
In equation 4.3, Q is the district heating demand energy which is calculated by the difference between the value of the total demand and the peak load demand. c represents the specific heat capacity that is equal to 4200J/kg °C (water). Δt is the temperature change. Considering m is more than zero in practice, one extra rule is added to determine whether the value of the temperature difference is positive. In the thesis work, the temperature difference is established based on the comparison between the actual water temperature and the return temperature setpoint (55°C).

4.4 The Modelling of the Multi-carrier Energy System

The electrical network in Chapter 3 is a python-based power flow model, the heating network and the controllers, modelled in OpenModelica, are time-continuous models. Moreover, the time step of these models are also different. Hence, a tool is necessary to combine different networks to form the multi-carrier energy system and undertake the multi-domain simulation.

4.4.1 The Methodology of Co-simulation

Co-simulation, as a powerful tool, plays an important role in analysing the interaction between various networks in energy systems. It overcomes the limitation occurring in traditional tools which concentrate on simulating one detailed structure with simplified models and ignore the interactions between models [48]. EnergySim, previously called as FMUWorld, is one of the platforms of the co-simulation. Its structure is shown in Figure 4.6 (a), and different FMU modules (i.e., four FMUs) are connected in a canvas/World [23]. The simulation time of each energy domain model (electricity networks, thermal networks, etc.) in energy systems have obvious distinctions. In EnergySim, these models are packaged as FMUs. Hence, the time step of each FMU is different. One of the prime functions of EnergySim is to find an optimal time size (highlighted in red) for the different FMUs described in Figure 4.6 (b), and the FMUs can exchange the related information in the World. Therefore, EnergySim is used to undertake the the multi-domain simulation with a combined analysis in the thesis work.



(a) (b)

Figure 4.6: The description of FMUWorld. (a) FMUWorld structure (b) Time step determination for simulation setups with variable step sizes for each FMU [23].

4.4.2 The Establishment of the Multi-carrier Energy System

A Dutch multi-carrier energy system, also called as a hybrid network, is designed based on the rural area with 13 households as illustrated in Figure 4.7. In the system, it combines the structure of the electricity distribution network integrating PV modules (Figure 3.7) with the heating network (Figure 4.1). If the excess PV power occurs in the electricity network, it will be utilised by electric heaters and storage tanks in the heating network according to the above operational strategies of the controllers.

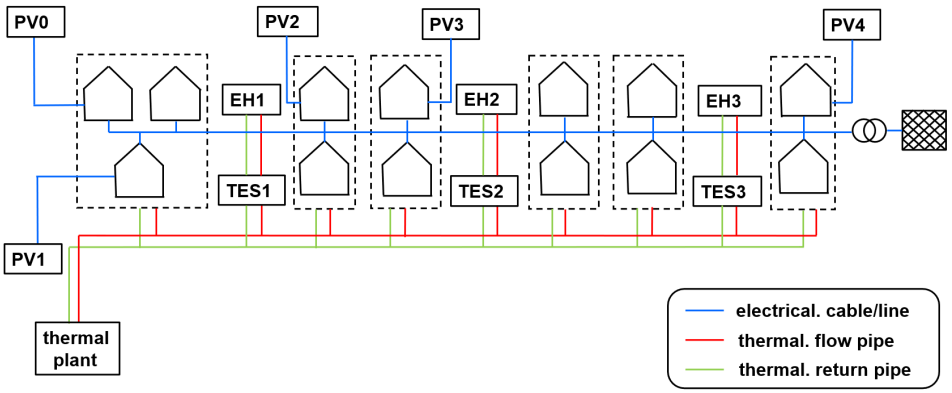


Figure 4.7: The Dutch hybrid network of the rural area (PV, photovoltaic; EH, electric heater; TES, thermal energy storage).

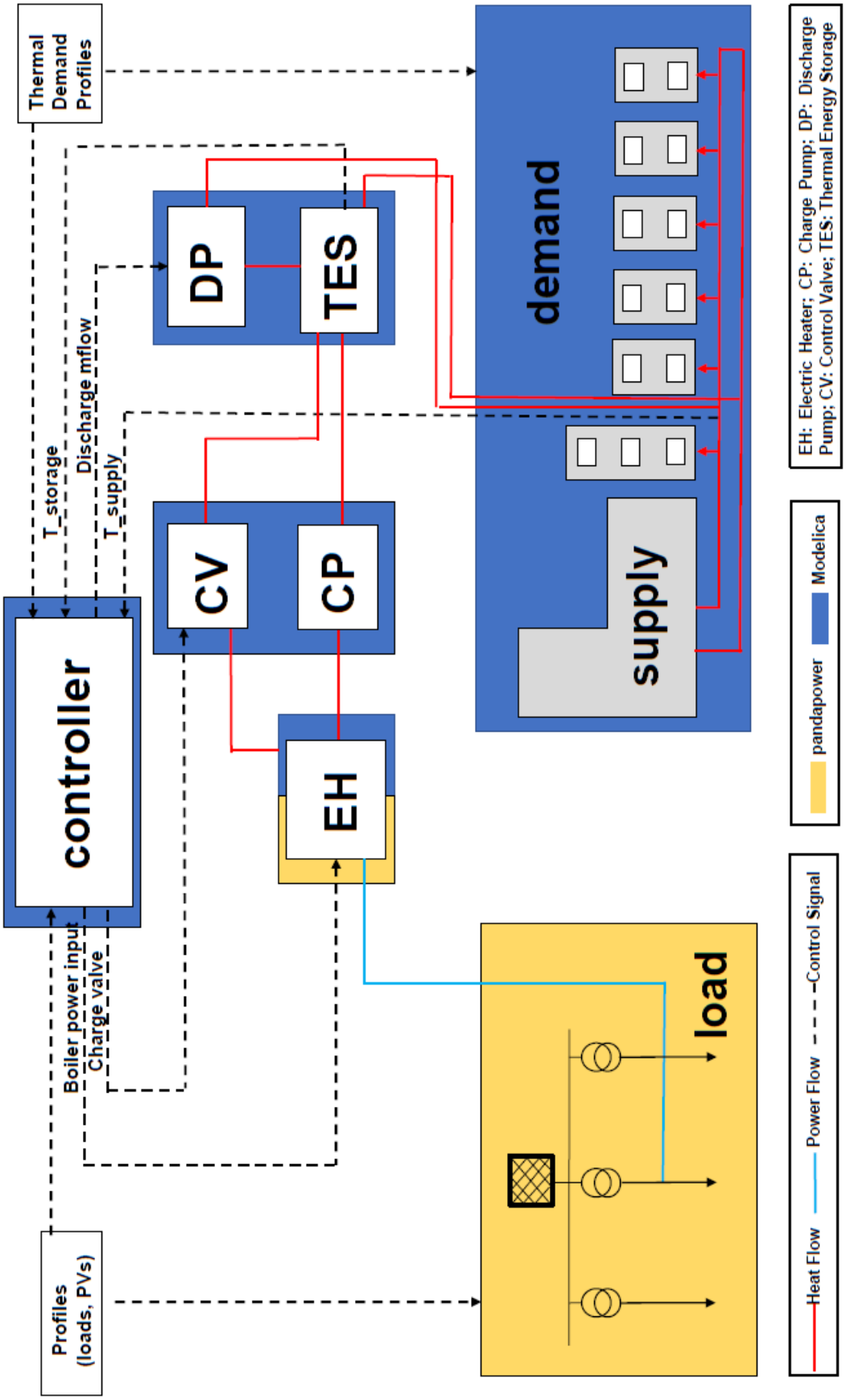


Figure 4.8: The simplified scheme of the co-simulation framework based on the multi-carrier energy system in the rural area.

Figure 4.8 depicts the simplified scheme of the co-simulation framework applied in the thesis work. The Dutch multi-carrier energy system, mainly focusing on the rural areas, is modelled by means of the co-simulation (EnergySim) which connects various networks including the electrical distribution network (pandapower), the heating network (OpenModelica) as well as the controllers (OpenModelica). Moreover, the interaction between the heating network and the controller can be depicted intuitively in this figure. It can be seen that the controllers send the corresponding signals to the heating network according to the operation strategies listed earlier. The heating network returns the real-time temperature measurement signals (the top-layer temperature of the specific storage tank and the supply temperature) to the controller.

It should be noted that electric heaters are taken into account in the heater & storage system of the heating network. Their characteristic is to transfer electrical energy to heat energy with a specific efficiency (0.97 in the thesis work). Hence, the electrical heaters are necessary to be added as extra loads in the established electrical network in the rural area. Its load profile, when the timestep is equal to 0, can be illustrated in Table 4.2. The main parameter of the electric heaters in the heating network are the heat flow rate whose value is shown in Table 4.2. Furthermore, heat pumps are also key components in the heater & storage system.

Table 4.2: The simulation parameters of electric heaters in the electrical network at time=0.

Name	Bus	p (mw)	q (mvar)	Scaling	In Service
EH1	8	0	0	1	True
EH2	10	0	0	1	True
EH3	13	0	0	1	True

Having determined the structure and the above data, co-simulation can be undertaken by EnergySim. It will take several hours even days if the simulation period of the energy system is settled as one-year. Hence, 12 representative days (3 days per season) are finally chosen as its simulation period. Considering the data (csv format) is in 15-min time resolutions, the number of timesteps is 1152 timesteps based on 12 days. As a result, the outputs (3 groups of storage temperature profiles and supply temperature profiles) are obtained after carrying out the co-simulation.

It is also worth mentioning that different losses are taken into account in the above multi-energy energy system. In the electrical network, when using pandapower, the parameters of cable lines are designed, and hence power losses, based on the power flow calculation, are included in the simulation. In the heating network, four largest parameters affecting the losses involve the amount of insulation around the pipes, the pipe dimension, supply and return temperature and the geographical distribution of the heat demand [106]. These parameters are considered and designed before carrying out the simulation in OpenModelica. Thus, the losses in the heating network are also included in the simulation.

5

Surrogate Models for the Multi-carrier Energy System

The objective of the chapter is to create surrogate models based on the multi-carrier energy system established in the previous chapter. First, computationally expensive issues are described, and this is why surrogate models are introduced. Second, surrogate models with their modelling processes are given by a list of mathematical expressions. Furthermore, different creating methods of surrogate models are described with the focus on several main approaches used in the thesis work. Finally, the application of surrogate models in the multi-carrier energy system is depicted in detail.

5.1 The Background of Surrogate Models

Computationally expensive problems have received notable research attention in the engineering field. The work in Ref. [107] offers its background in depth. During the conventional engineering design process, several prototypes used to be established for the sake of providing benefits to the real system design and optimisation. However, the process is often slow because different prototypes, rather than a single one, need to be considered. A dramatic progress is achieved by computer simulations with the corresponding models which decrease the number of prototypes, and hence the traditional design process is accelerated efficiently. Then, a series of approaches, in order to make these simulation models closer to the real case, is proposed. For instance, these models get more complicated [108], and they might have more interactions [11]. As a consequence, simulation time increases significantly, and computational costs are very expensive. These direct effects can be categorised into computationally expensive problems [109], [110]

In the context, many researches have focused on dealing with computationally expensive problems. In Ref. [108], the authors offer a detailed description of the problems by three different aspects: approximation models, optimisation algorithms and multi design objectives. Figure 5.1 depicts the aspects intuitively. Furthermore, they also figure out effective solutions. In light of approximation models, their goals are to employ a list of techniques to replace the expensive structures in the original models. The techniques include model-driven approaches, data-driven approaches and their hybridisation [107]. Model-driven approaches, also named as model order reduction approaches, are limited to the application in specific domains. Moreover, they make approximations of original simulators by means of mathematical methods. Data-driven approaches, compared with the model-driven approaches, are not problem-specific. Moreover, a term called *surrogate models* [110] is introduced because the approaches ignore internal states of simulators treated as black-box models. Hybrid approaches focus on applying the problem-specific knowledge to data-driven models. Hence, the approaches combine model-driven approaches with data-driven approaches. In light of optimisation algorithms, there exist two main branches: local search optimisation methods and multimodal global optimisation methods. On one hand, a series of classical algorithms, such as gradient-based algorithms, shows excellent performances in the local search methods. On the other hand, a number of advanced methods, such as

heuristic algorithms, have more advantages in the multimodal global methods. In light of multi design objectives, it ought to be focused in particular because computationally expensive problems are likely to appear in many engineering applications involving multiple objectives and complicated constraints.

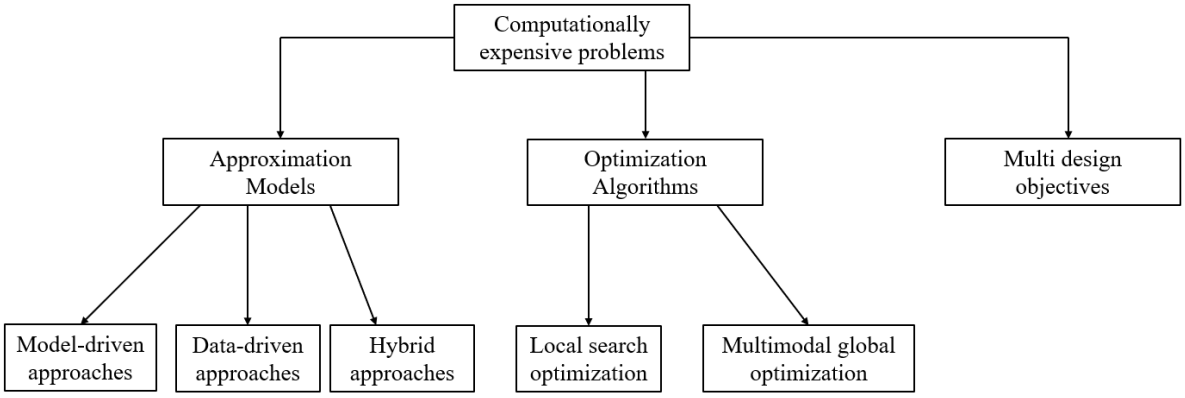


Figure 5.1: Three main aspects considered in the computationally expensive problems according to Ref. [108].

The main focus of interest at the thesis work is employing data-driven approaches to tackle the computationally expensive problems due to a great amount of data is used in the energy system simulation. Surrogate models are established to replace expensive models in the whole system to reduce high computation costs.

5.2 The Mathematical Description of Surrogate Models

Surrogate models, also called meta-models or approximate models [39], are commonly used to replace expensive models in the data-driven approaches. Moreover, the models depict the relationship between independent variables (x) and dependent variables (y). Ref. [111] describes the approximation between original models and the corresponding surrogate models. If the original model is represented as

$$y = f(x) \tag{5.1}$$

Then a surrogate model can be expressed as

$$\hat{y} = \hat{f}(x) \tag{5.2}$$

with $y = \hat{y} + \epsilon$, where ϵ refers to the disturbance terms.

Moreover, in Ref. [107], the authors focus on the internal process of surrogate modelling, and its mathematical expression is given as follows:

$$argmax_{t\in T} argmin_{\theta\in\alpha} - \Lambda(\kappa, \widetilde{f}_{t,\theta}, D) \tag{5.3}$$

subject to

$$\Lambda(\kappa, \widetilde{f}_{t,\theta}, D) \leq \tau \tag{5.4}$$

where $\widetilde{f}$ represents an expensive function, D is the data sample collection with the corresponding evaluations, κ represents an error function such as mean squared error (MSE), root mean square error, etc., Λ is the model quality estimator, τ is the targeted quality value, t and T are model types and available model types. θ and α represent hypermetres and available hypermetre values, respectively.

5.3 Detailed Method Illustrations of Creating Surrogate Models

Different creating methods of surrogate models have been summarised in Chapter 2. As one of the

approaches, machine learning algorithms have recently seen rapid growth in the research of the energy domain. Hence, ML algorithms are employed to establish surrogate models, and the multi-output regression problem will be concentrated. Different approaches have been proposed to deal with the problem. In the thesis work, eight typical methods (LR, RC-LR, RC-SVR, DT, RF, k-NN, MLP, LSTM) are chosen and used in the algorithms, and their descriptions are given as follows.

5.3.1 Linear Regression

Regression analysis, as one of the statistical methods [112], is used to estimate the relationships between a dependent variable [1] (y) and one or more independent variables (x). One of the common

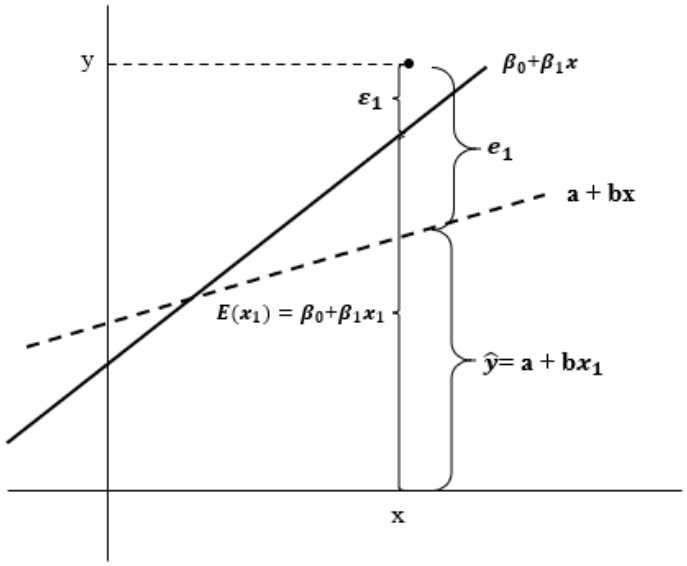


Figure 5.2: Population and sample regression [113].

types in the regression analysis is linear regression models. Its linear relationship between x and y, the so-called regression equation [114], is expressed as:

$$y = X\beta + \epsilon \tag{5.5}$$

Equation 5.5 can be expanded to the form of matrix, which can be written as the equation 5.6:

$$y = \begin{bmatrix} 1 & x_1 & x_2 & \dots & x_K \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{bmatrix} + \epsilon \tag{5.6}$$

The equation 5.6 can be organised and expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_K x_K + \epsilon \tag{5.7}$$

where β_0, β_1, β_2, β_K represent the regression coefficients.

Following the above definition, one of the simplest equations is a single linear regression model [115] when K is equal to 1. Figure 5.2 shows an example of the model. It can be depicted as:

[1] In this thesis work, the terms dependent variable, target variable and response variable will be used interchangeably.

$$y_i = \beta_0 + \beta_i x_i + \epsilon_i \tag{5.8}$$

with:

$$\epsilon_i \sim N(0, \epsilon^2) \; and \; cov(\epsilon_i, \epsilon_j) = 0, when \; i \neq j \tag{5.9}$$

herein i refers to a particular observation, and ϵ_i is a random error or disturbance terms.

The single linear regression model is a good example to illustrate the function of regression analysis. As discussed above, its prior goal is to estimate the relationship between y and x. From the equation 5.8, it is easy to get the estimated parameters are β_0 and β_i. A typical method of the estimation is the least squares regression which forms the following optimisation problem with no constraints [115]:

Minimize

$$\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_i x_i)]^2 \tag{5.10}$$

It is worth mentioning that the above linear regression is simple to implement. However, according to Ref. [116], it might introduce the underfitting or overfitting issue. Ridge regression, as a variant of linear regression, is mainly taken into account to overcome the issue in the thesis work.

5.3.2 Regressor Chain

Regressor chain method is mainly applied in multi-output regression problems with the idea of chaining single-target models [117]. When predicting all target variables together is difficult, the method is very

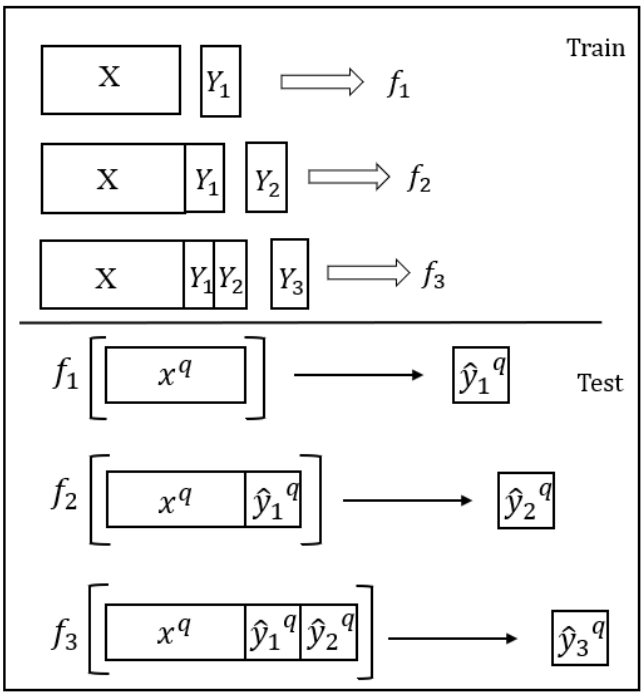


Figure 5.3: Graphical illustration of regressor chain [118].

useful by considering prediction values sequentially. For instance, an easy application of the regressor chain is to create a linear sequence of models on the basis of single-output regression models. Figure

5.3 clearly shows its working principle by taking into account three output variables (Y_1, Y_2, Y_3) and one input variable (X) [118].

In the training process, a model (f_1) is fitted for the first dependent variable (Y_1) when its input is an independent variable (X). Then, a new model (f_2) is fitted for the second dependent variable (Y_2) by considering two inputs: the dependent variable (X) and the target variable (Y_1). Similarly, a model (f_3) is fitted for the third response variable (Y_3) when its input variables are the dependent variable (X) and the independent variables (Y_1, Y_2).

In the testing process, the first stage begins with acquiring the predicted output variable (Y_1) by using the first model (f_1) and the test input data (x^q). Then, a prediction for the second output variable ($\hat{y}_2^q$) is made by the second model (f_2) , the same test data (x^q) as well as the first predictions ($\hat{y}_1^q$). In the last stage, the third output variable ($\hat{y}_3^q$) is predicted by using the concatenated data (x^q, $\hat{y}_1^q$, $\hat{y}_2^q$) and the third model (f_3).

5.3.3 Linear Support Vector Machine

Support vector machine algorithm, first proposed by Vladimir Vapnik and his colleagues, is a prominent supervised machine learning method for classification and regression [119]. The advantage of the algorithm is that it offers more choices of penalties and loss functions. Furthermore, it is expected to scale better when the numbers of samples are large [120].

If a training set $(x_1, y_1), ..., (x_i, y_i)$ is given, then the goal of the SVM can be described to find a fitting function and make sure that it is as flat as possible [121]. In other words, the deviation between the function and the target (y_i) is less than ε. The fitting function can be depicted by

$$f(x) = \omega x + b \tag{5.11}$$

Here ω is the weighted vector, and b represents the bias. Due to the fitting function is linear, the definition of linear SVM is introduced.

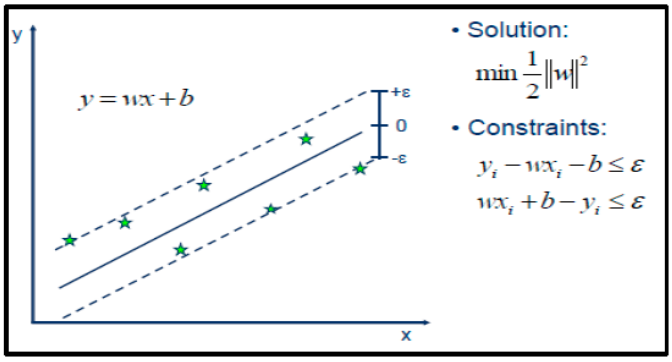


Figure 5.4: Support vector machine [121].

Figure 5.4 can explain the algorithm intuitively. The solid line represents the targeted fitting function. The dash lines are called as the negative hyperplane and positive hyperplane, respectively. The vector points closest to the hyperplanes are known as the support vector points. The distance between the vectors and the hyperplane is named as the margin. Hence, the goal of SVM can be depicted as a typical optimisation problem with searching for the maximum margin value. The solutions and the corresponding constraints in the problem are also illustrated in the figure.

However, according to Ref.[122], the algorithm cannot support the multi-output regression. An effective solution is to divide the regression into multiple single-output regressions. The solution has the similar

idea as the above regressor chain algorithm. Hence, the linear SVM is combined with the regressor chain to achieve the multi-output regression in the thesis work.

5.3.4 Decision Tree

Decision tree, as a non-parametric supervised learning method, is applied in two important aspects: classification and regression [123]. Hence, the definitions of classification trees and regression trees are introduced. One of the main differences between these two trees is dependent variables [124]. The classification trees take unordered values as their dependent variables, whereas the regression trees take continuous values or ordered discrete values. Figure 5.5 shows the basic structure of a three-layer decision tree.

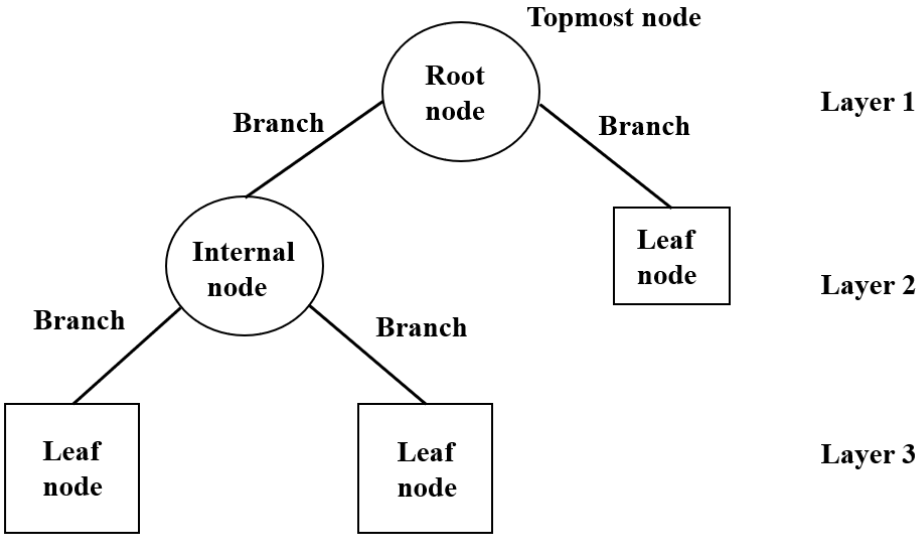


Figure 5.5: Basic structure of a three-layer decision tree [125].

A decision tree has nodes, branches and layers which originate from the definitions in the graph theory. In a graph G = (V, E), it includes a finite, non-empty set of nodes V and edges E [126]. With regard to the nodes, there are three various nodes: root (topmost) node, leaf node and internal node. Take the three-layer decision tree as an example, the whole tree begins from the node which is called a root node or a topmost node. The nodes without any extension are named as leaf nodes or terminals. The other nodes in the tree are called internal nodes except for root nodes and leaf nodes. With regard to the edges, they are known as branches in a tree. The branches, as a subsection of the decision tree, play an essential role in connecting various nodes. In regression trees, particularly, each internal node corresponds to a test on a feature. Each branch and each leaf node represent the decision rule and the prediction result, respectively. There also exist other definitions. To illustrate, the layer of the tree is described by the maximum depth of a node defined as the path length from the root node to the leaf node [127].

5.3.5 Random Forest

The above decision tree method is easy to understand and implement. However, the method is vulnerable to overfitting issues particularly when the layers of a tree are extremely deep [128]. One available remedy is to use the random forest algorithm introduced by Breiman [129]. Random forest is defined as an ensemble which combines various tree predictors $h(x;\theta_k), k = 1, 2, ...K$. Here x refers to the input data with associated random vector X, and θ_K represents a series of identical and independent distributed random vectors. Figure 5.6 illustrates an architecture of a random forest model. It can be observed that each individual tree offers a prediction value (i.e., $k1$). Then, the final prediction value (k) of the whole model is selected by votes in the classification or averages in the regression with the

highest weighted values.

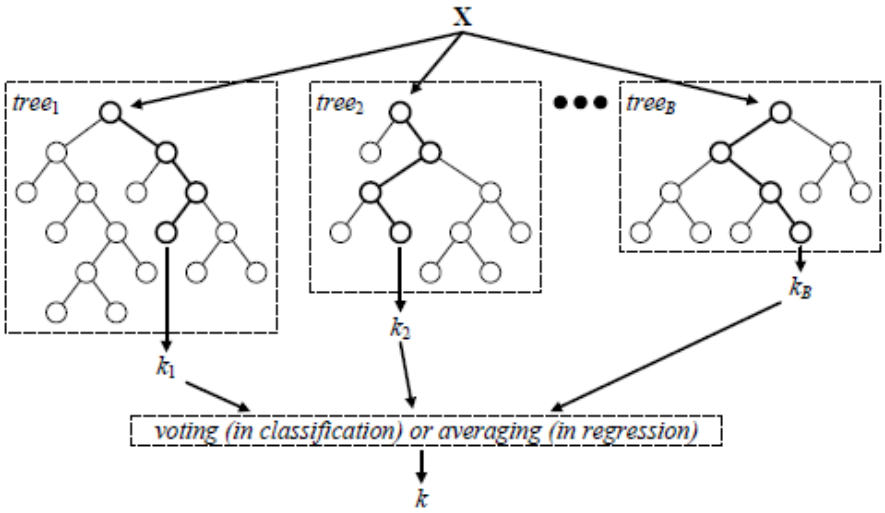


Figure 5.6: Architecture of a random forest model [129].

The reason why the random forest algorithm can avoid the overfitting issues can be illustrated by the following mathematical expressions [130]. In terms of the regression, the prediction of the random forest can be depicted as:

$$\overline{h}(x) = \frac{1}{K}\sum_{k=1}^{K} h(x;\theta_k) \tag{5.12}$$

When $k \to \infty$, the Law of Large Numbers gives:

$$E_{X,Y}(Y - \overline{h}(x))^2 \to E_{X,Y}(Y - E_\theta h(X;\theta))^2 \tag{5.13}$$

where Y is the outcome of the whole model. Hence, the random forest algorithm is able to tackle the overfitting issues due to the convergence shown in the expression.

Having established the models of the random forest, one of the important aspects in the regression is to enhance its accuracy. In order to give a detailed description, several definitions need to be introduced in advance [130]. First, the average prediction error of an individual can be expressed as:

$$PE_t{}^* = E_\theta E_{X,Y}(Y - h(X;\theta))^2 \tag{5.14}$$

Second, an assumption is added that the tree is unbiased for all θ:

$$EY = E_X h(X;\theta) \tag{5.15}$$

If the convergence result in Formula 5.13 is noted as $PE_f{}^*$, then

$$PE_f{}^* \leq \overline{\rho} PE_t{}^* \tag{5.16}$$

where $\overline{\rho}$ represents the weighted correlation between $Y - h(X;\theta)$ and $Y - h(X;\theta')$. Therefore, it can be seen that the low weighted correlation value and average prediction error are two key aspects for obtaining the high accuracy in the random forest regression.

5.3.6 k-nearest Neighbours

The k-nearest neighbours algorithm is a supervised machine learning method [131]. Ref. [132] depicts the definition of the kth nearest neighbour. Define $x \in R^d$ and the data $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ based on the increasing order of distance values between (X_i) and x, noted as $\| X_i - x \|$. Then, the data sequence can be obtained:

$$(X_{(1,n)}(x), Y_{1,n}(x)), ..., (X_{(n,n)}(x) Y_{(n,n)}(x)) \tag{5.17}$$

In the equation 5.17, $X_{(k,n)}(x)$ is defined as the kth nearest neighbour of x.

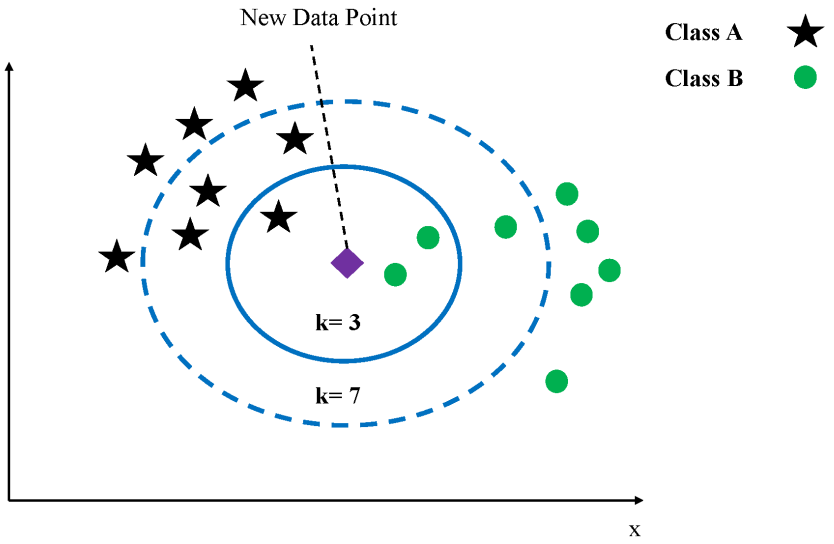


Figure 5.7: An example of k-nearest neighbor algorithm illustration used to predict the class of a new data point [133].

Figure 5.7 shows an example of the k-nearest neighbour algorithm illustration by predicting the class of a new data point. There are two initial class labels (Class A and Class B), and the purple diamond is the new data point. When k is equal to 3, the algorithm will just consider the three closest neighbours (two points belonging to Class B and one point belonging to Class A) inside the blue circle with the dotted line. Hence, the new data point is in the category of Class B. When k is equal to 5, seven nearest neighbours (four points in Class A and three points in Class B) will be taken into account inside the blue circle with the solid line. In this case, the new data point belongs to Class A. Therefore, in the algorithm, the value of k plays an essential role in determining the prediction of the new data point. Furthermore, it is worth mentioning that the circles in the figure are drawn according to the distance metric [133]. The commonly used metric is called Euclidean distance (ED) [134] whose expression is shown as follows:

$$ED(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{5.18}$$

Herein x and y are depicted by feature vectors $x = (x_1, x_2, ..., x_n)$ and $x = (y_1, y_2, ..., y_n)$, respectively. n refers to the dimensionality of the feature vectors.

5.3.7 Multilayer Perceptron

A neural network (NN), also called an artificial neural network, is inspired by the organisation of the human brain [135]. It consists of a collection of mathematical or computational models that are made of natural or artificial neurons. Its categories can be illustrated in Figure 5.8.

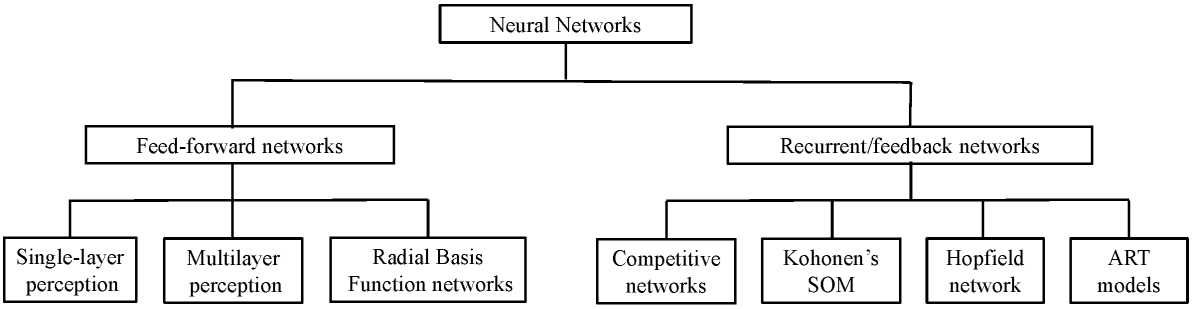


Figure 5.8: A taxonomy of neural network architectures [135].

A single-layer perceptron, as one of the branches in feed-forward networks, is proposed by Frank Rosenblatt [136]. It is a simple learning machine whose inputs are a vector of real numbers and output is a real number. Figure 5.9 shows the structure of a perception. The mathematical expression of the perception can be described as follows [117], [137]:

$$y = f(\omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_n x_n + b) = f(\omega^T x + b) \tag{5.19}$$

where $w_1, w_2, \ldots, w_n$ represent different weights, b is a bias, f is an activation function, $x_1, x_2, \ldots, x_n$ represent inputs, and y is an output. It is worth mentioning that the activation function, in the process of training, will be optimised to reduce the error between the output value of the perception for the corresponding input and the target output value.

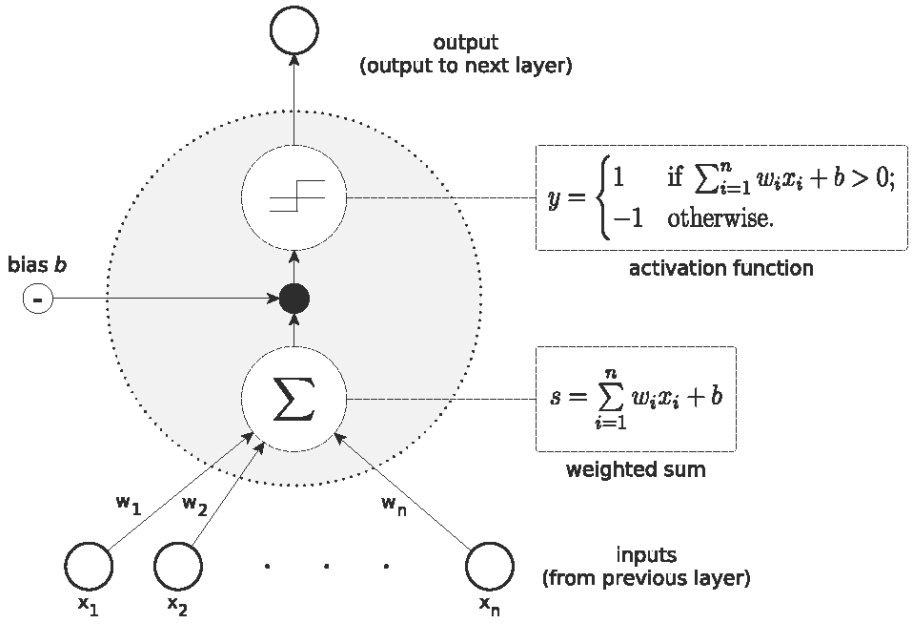
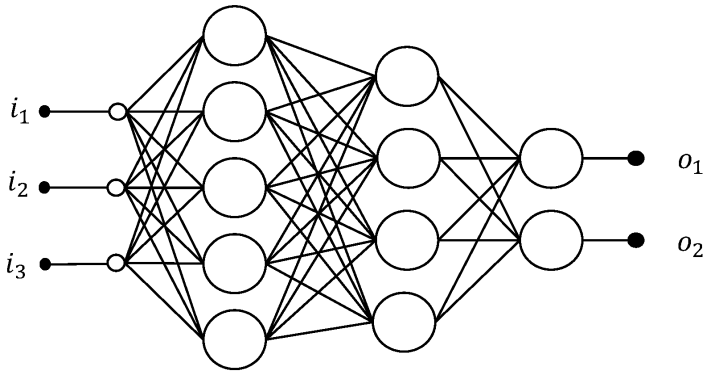


Figure 5.9: The general structure of a perception [138].

The above perception algorithm is simple to implement, however it does not handle non-linearly separable data [137]. Hence, a multi-layer perception algorithm is introduced to overcome the drawback [139]. MLP is a neural network which incorporates additional multilayer perceptions. Moreover, the structure of a MLP is fully connected that means each node of the MLP is connected with every node in the adjacent layers. Figure 5.10 illustrates an example of a MLP with two hidden layers. In particular, if the structure of a MLP has many hidden layers, it can be called a deep neural network [140].



$$\underline{i} = [i_1, i_2, i_3] = \text{input vector}$$
$$\underline{o} = [o_1, o_2] = \text{output vector}$$

Figure 5.10: An example of the multi-layer perception [139].

The mathematical expression of a MLP can be depicted [139]:

$$Y_{MLP} = f(X_{MLP}) \tag{5.20}$$

where $X_{MLP} = [n \times k]$, $Y_{MLP} = [n \times j]$, X_{MLP} is the input vector and Y_{MLP} is the output vector. f is the activation function. n represents the number of training patterns, k and j are the number of input nodes and output nodes, respectively.

5.3.8 Long Short-term Memory

A recurrent neural network (RNN), as shown in Figure 5.11, is another important branch in the neural

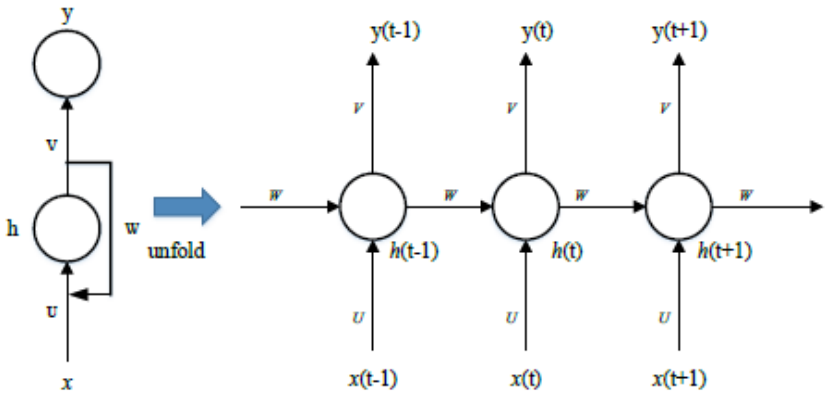


Figure 5.11: The structure of the RNN [141].

networks. The network is mainly used to handle time-series data in comparison with other algorithms [142]. A simple structure of RNN is shown in Figure 5.11. After being unfolded, the right part in the figure can be described as follows [141]:

$$h_{(t)} = \sigma(Ux(t) + Wh_{(t-1)} + b) \tag{5.21}$$

$$y_{(t)} = \sigma(Vh(t) + c) \tag{5.22}$$

where $x_{(t)}$, $y_{(t)}$ and $h_{(t)}$ refer to the input vectors, output vectors and hidden vectors, respectively at t; σ is the activation function; b and c are the bias terms; U, W and V represent connection weights.

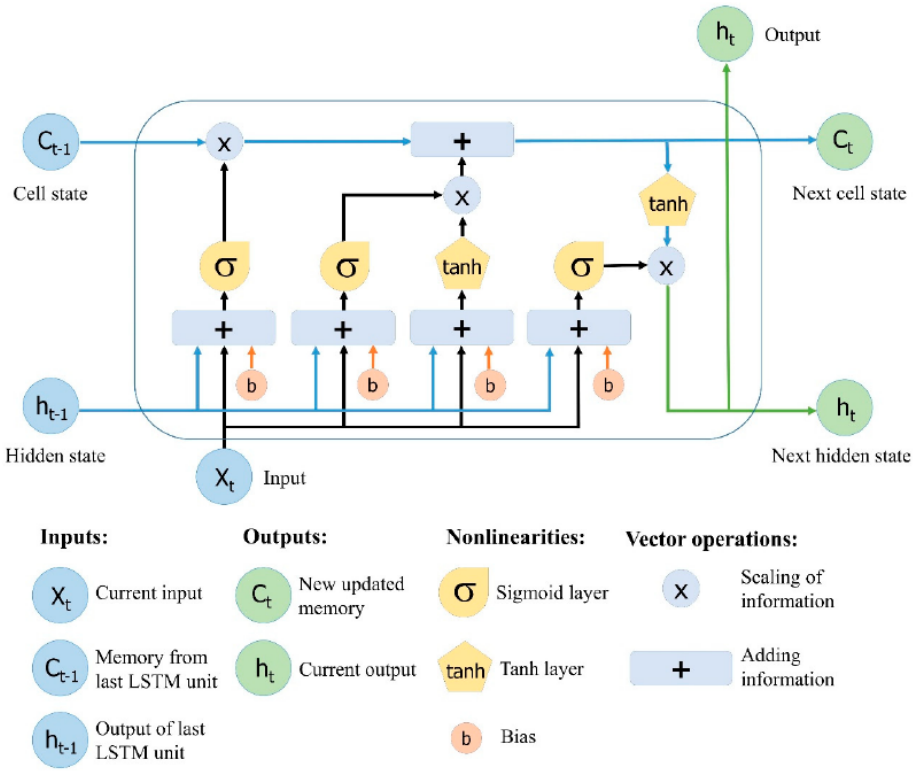


Figure 5.12: The structure of the long short-term memory neural network [143].

RNN networks, as discussed above, mainly manage the temporal data. However, the typical RNN models are limited to short-term memories like 5-10 time steps according to Ref. [144]. Hence, a neural network called long short-term memory is introduced, and Figure 5.12 illustrates its structure. Ref. [143] offers the mathematical expressions of the LSTM structure in detail. Three gates (forget gate, input gate and output gate) can be described as:

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \tag{5.23}$$

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \tag{5.24}$$

$$O_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \tag{5.25}$$

where f_t,i_t,O_t are the forget gate, input gate and output gate, respectively. W and b represent the weight matrices and bias of the corresponding gates; σ is the nonlinear function, and the sigmoid function is selected as usual because it is used to update or omit the new information (1 or 0).

An intermediate state N_t in LSTM can be expressed as:

$$N_t = tanh(W_n[h_{t-1}, X_t] + b_n) \tag{5.26}$$

here W_n and b_n represent the corresponding weight matrices and bias; $tanh$ refers to the tanh function which determines the weight values.

In consequence, the updating values of the memory call and hidden state can be shown as:

$$C_t = C_{t-1} f_t + N_t i_t \tag{5.27}$$

$$h_t = O_t tanh(C_t) \tag{5.28}$$

where C_{t-1} and C_t refer to the cell states at $t-1$ and t, respectively; h_t is the hidden state at t.

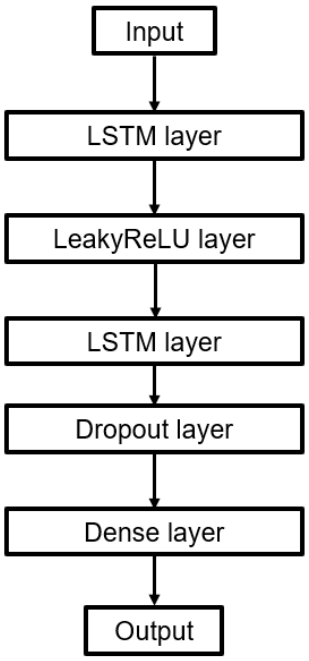


Figure 5.13: The structure of the LSTM model used in the thesis work.

It should be mentioned that LSTM algorithm has different models, and stacked LSTM is selected due to its simple structure in the thesis work. Its structure is depicted in Figure 5.13. It can be seen that, beside LSTM layers, different layers are covered in the structure. LeakyRelu is one of the common activation functions. The dropout technique is used for regularisation [145]. The dense layer is another basic layer in neural networks. Furthermore, in order to reduce the computation time, a callback called EarlyStopping in tf.keras is taken into account in its simulation.

5.4 The Modelling of Surrogate Models

The above eight methods are used to create surrogate models. Scikit-learn and Tensorflow, as simulation platforms, are depicted in detail. Furthermore, the modelling process of surrogate models including their structures and datasets are illustrated. Finally, several indicators are described to compare the performances between the surrogate models and the original model.

5.4.1 Simulation Platforms

Scikit-learn, mainly written in Python, is an important library by offering algorithms to deal with machine learning problems [24]. It is established on the basis of several libraries such as Numpy [146], Scipy [147] and Cython [148]. Numpy has advantages when handling large arrays and multi-dimensional matrices. Scipy is a powerful scientific computing tool. Cython performs the function of integrating C language in python.

Scikit-learn plays an essential role in enhancing the computation efficiency. Table 5.1 shows computing time comparisons between six algorithms when using different machine learning libraries on the Madelson data set. It can be seen that scikit-learn, compared with the other five libraries, shows the best performances in Supper Vector Classification, LARS, Elastic Net and PCA. In the algorithm of K-Nearest Neighbours, the computing time of scikit-learn is similar to pymvpa which has the most excellent performance. In the algorithm of k-Means, the performance of Scikit-learn is limited owing to NumPy's array operations [24]. In the work, Scikit-learn is used to create different algorithms including linear regression, regressor chain, support vector machine, decision tree, random forest, k-nearest neighbours as well as multilayer perceptron.

Table 5.1: Time in seconds on the Madelson data set for various machine learning libraries exposed in Python [24].

	scikit-learn	mlpy	pybrain	pymvpa	mdp	shogun
Support Vector Classification	**5.2**	9.47	17.5	11.52	40.48	5.63
Lasso (LARS)	**1.17**	105.3	-	37.35	-	-
Elastic Net	**0.52**	73.7	-	1.44	-	-
k-Nearest Neighbors	0.57	1.41	-	**0.56**	0.58	1.36
PCA (9 components)	**0.18**	-	-	8.93	0.47	0.33
k-Means (9 clusters)	1.34	0.79	⋆	-	35.75	**0.68**
License	BSD	GPL	BSD	BSD	BSD	GPL

-: Not implemented. ⋆: Does not converge within 1 hour.

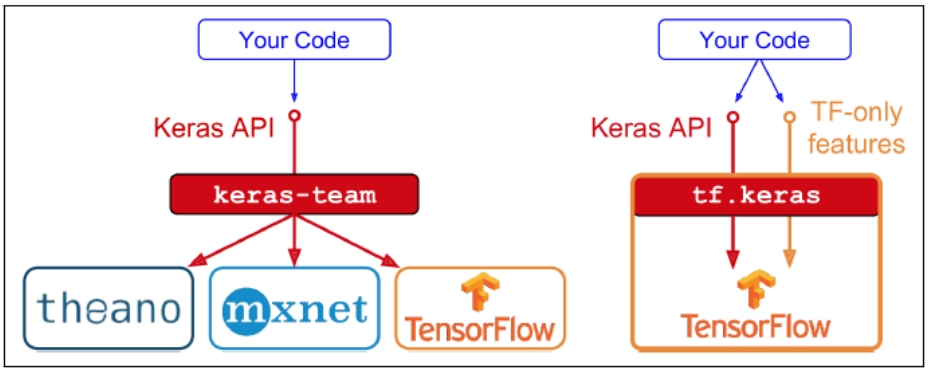


Figure 5.14: Two Keras implementations: keras-team (left) and tf.keras (right) [149].

Keras, an Application Programming Interface (API) for deep learning, is mainly employed to build neural network models [26]. Here the term called keras-team describing the implementation of Keras is followed as stated in Ref. [150]. Keras-team is established on the basis of Theano [149] and TensorFlow [151]. Theano, a python-based library,arrays. Tensorflow is an open-source library applied for building computational graphs. Moreover, keras-team depends on a computation backend which can be selected from one of the libraries including TensorFlow, Theano, Mxnet, etc. Particularly, when TensorFlow is chosen as the backend of Keras, it has the own implementation, named as tf.keras, with TensorFlow (TF)-only features as depicted in the right part of Figure 5.14. In the thesis work, long short-term memory algorithm is established by Keras, and tf.keras is selected as the implementation due to no additional TF features are considered.

5.4.2 The Establishment of the Surrogate Models in the Energy System

The multi-carrier energy system in the rural area has been established as shown in Chapter 4. However, its computation costs are expensive when carrying out its simulation in the system. Hence, surrogate models based on the system are necessary to be created to enhance the computation efficiency. Figure 5.15 shows the simplified scheme of the surrogate models in the multi-carrier energy system. In the thesis work, the electrical network and the heating network are replaced by surrogate models, respectively. In order to give a detailed description, the surrogate models are zoomed and depicted in Figure 5.16. The surrogate models in two networks have similar functions such as enhancing the computation efficiency and predicting multiple outputs. However, they have notable differences, and hence they are split into two models.

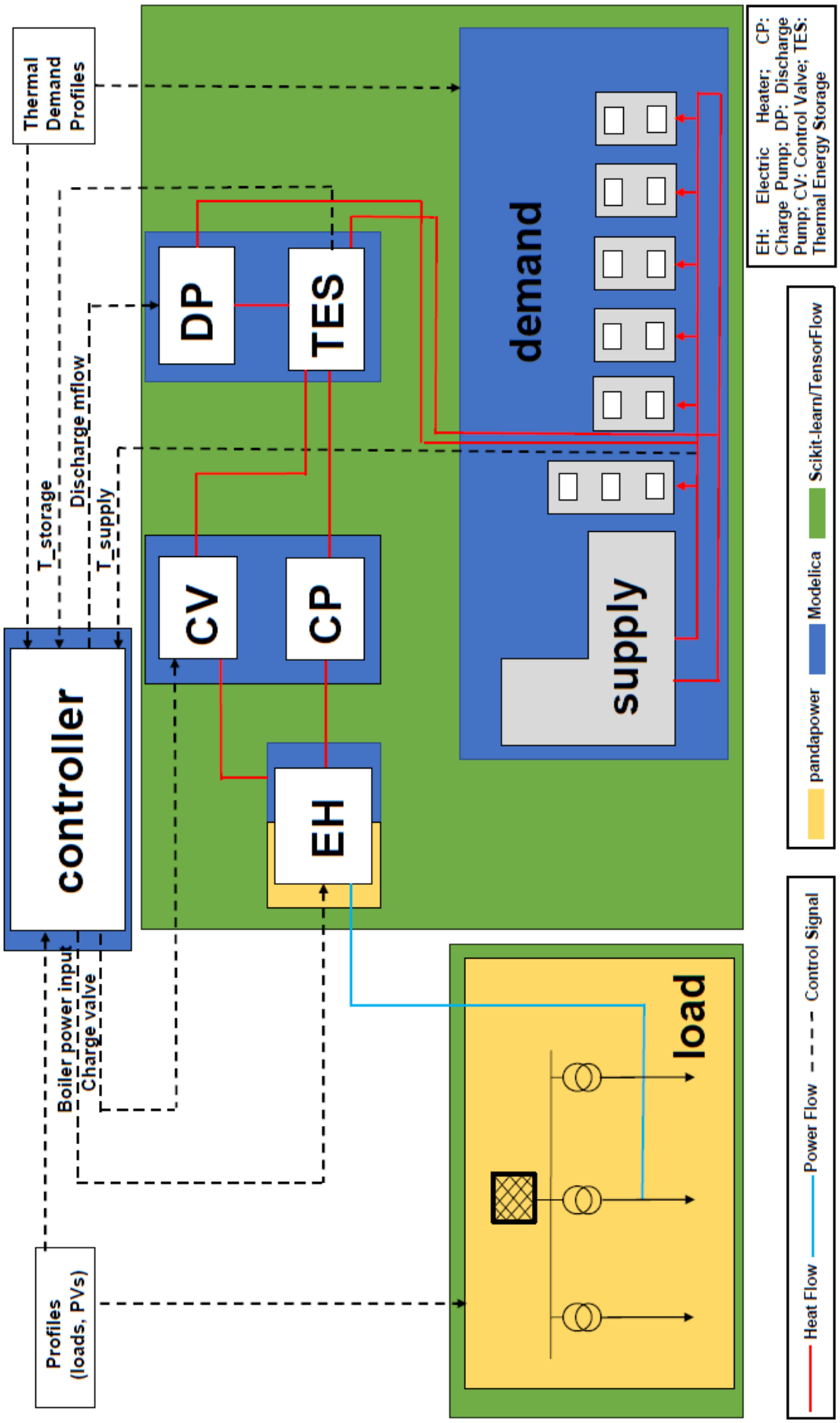


Figure 5.15: The simplified scheme of the surrogate models based on the multi-carrier energy system.

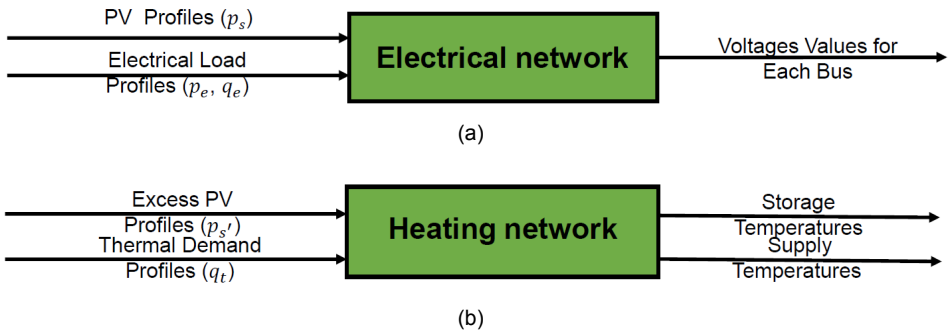


Figure 5.16: The structures of surrogate models used to replace the original networks in the energy system. (a). the surrogate model in the electrical network; (b). the surrogate model in the heating network.

First, the difference is employed in various characteristics of the simulation in these two networks. In the electrical network, its output variables (voltages) are a result of static-power flow solution implemented by pandapower. Thus, its surrogate model replaces the power flow functionality for the electrical network. For a given set of inputs (PV and load setpoints) at a particular time, the power flow model and the surrogate model calculate the voltages for each bus at that time. In the heating network, its output variables (temperatures) are a result of dynamic simulation implemented by Open Modelica. Hence, the surrogate model here therefore replaces the dynamic simulation. For a given set of inputs (excess PV and thermal demand setpoints), the dynamic simulation model and the surrogate model calculate the storage and supply temperatures at specified points in the heating network.

It must be noted here that both the surrogate models developed here take inputs as an array of values at a particular time instant, and convert the array into the required output variables corresponding to those input values at that time. The surrogate models here do not capture sequential information in the input data, and therefore the outputs of each surrogate model at any instant of time are independent of the model state at previous time instants. In essence, surrogate models for both networks are snapshot models. Furthermore, the surrogate models without considering the sequential information in the input data are evaluated initially. Later, surrogate models which have the ability of dealing with the sequence data, such as LSTM, will be investigated. Moreover, the two approaches to surrogate modelling will be compared in detail.

Second, the input dataset and the output dataset of the surrogate models have obvious distinctions. The electrical network simulation is for a period of one year. In each 15 min time steps, a power flow is calculated based on the values of loads and pv modules, and then bus voltages are measured. Therefore, generated dataset has dimensions of (35040, 37) for inputs and (35040, 7) for outputs, which are corresponding to 37 inputs for pv modules, loads, and 7 outputs for 7 bus voltages. The dynamic simulation for the heating network is carried out along with electrical network and controller model. The thermal demand is given as setpoint within the heating network, while the controller determines the valve positions of storages and electric heater setpoints within the heating network. The dynamic simulation calculates the temperatures at various points in the network with a timestep of 0.1 sec. However, since the co-simulation for an entire year for the combined controller, electrical and heating network can take several hours and even simulated days, only 12 representative simulated days (three representative days for each season) with simulation time step of 0.1s. In order to train the surrogate models, however, the dataset is generated by sampling the results of high-resolution temperature profile from the dynamic simulation at every 15 minutes. In each 15 min time steps, a dynamic simulation is carried out based on the values of excess PV profiles and demand profiles, and then storage temperatures and supply temperatures are measured. Hence, generated dataset has dimensions of (1152, 10) for inputs and (1152, 6) for outputs, which are corresponding to 10 inputs for excess PV profiles, demand profiles, and 6 outputs for 3 storage temperature profiles and 3 supply temperature profiles.

5.4.3 The Description of the Modelling Approach

Having determined the above input dataset and output dataset, they can be trained and tested in the machine learning algorithms to create the corresponding models in the electrical network and the heating network, respectively. In each network, the modelling approach of the algorithms are similar.

As mentioned in the literature review, train/test data splitting is one of the key procedures in the machine learning algorithms. In the thesis work, except the LSTM algorithm, the above input/output dataset of the electrical network and the heating network is split by 80-20 ratio in the aforementioned machine/deep learning algorithm. 80% data is used for training the dataset, and 20 % data is used for testing. Then, data scaling is carried out as a pre-processing procedure. After that, except LSTM model, n-fold cross validation method with n = 10 based on the splitting percentage is employed to avoid the overfitting issue. Furthermore, the appropriate hyperparameters in each algorithm are selected in the process. The detailed description of the cross validation method is offered as depicted in Figure 2.8.

Table 5.2: Tuned hyperparameters in the machine/deep learning algorithm.

Machine/Deep Learning Algorithms	Hyperparameters and Tuning Ranges
LR	• alpha: [1, 0.1, 0.01, 0.001] • solver: [svd, cholesky, lsqr, sparse_cg] • max_iter: [100, 1000, 10000]
RC-LR	• based_alpha: [1.0, 0.1, 0.01] • based_solver: [svd, cholesky, lsqr, sparse_cg] • based_max_iter: [10, 100, 1000]
RC-SVR	• base_estimator__C: [0.1, 1, 10, 100] • base_estimator__tol: [0.00001, 0.0001, 0.001]
DT	• max_depth: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] • min_samples_leaf: [1, 2, 3, 4, 5] • min_samples_split: [2, 3, 4, 5, 6, 7, 8, 9, 10]
RF	• n_estimators: [10, 30, 100] • max_depth: [10, 20, 30, 40, 50] • min_samples_split: [2, 3, 4, 5] • min_samples_leaf: [1, 2, 3, 4, 5]
k-NN	• n_neighbors: range (1, 50) • p: [1, 2]
MLP	• hidden_layer_size: [(50,50), (100,50)] • activation: [tanh, relu] • solver: [sgd, adam] • alpha: [0.0001, 0.05] • learning rate: [constant, adaptive]
LSTM	• hidden_nodes: [128, 64, 32] • dropout: [0.2, 0.3] • batch_size: [32, 64] • epoch: [10, 50]

Hyperparameters,as a key branch of model parameters, play an important role in the machine learning algorithms. Compared with the ordinary parameters, hyperparameters need to be established manually before the machine learning model is trained [152]. Table 5.2 depicts selected hyperparameters and their tuning ranges used in the thesis work. It can be seen that each algorithm has a set of various hyperparameters. Hence, it is essential to obtain the appropriate hyperparameters by specific optimisation approaches. Two common approaches, including grid search and random search, are mainly considered in the thesis work. The major difference between them is their search space [153]. The random search focuses on a bounded domain where sample points are chosen randomly. In the thesis work, the approach is used to find the appropriate hyperparameters in the algorithms of RC-SVR, DT, RF and MLP. Whereas, the grid search concentrates on a grid where each position is assessed. The

approach is considered in LR, RC-LR and k-NN algorithms to assist in searching for their appropriate hyperparameters.

It should be noted that the above two optimisation approaches are still not employed in the LSTM algorithm. The hyperparameters of the LSTM model are easily tuned by means of the method mentioned in Ref. [154].

5.4.4 Performance Indicators

As discussed above, surrogate models are used to replace the original simulation models. Hence, it is necessary to introduce relevant performance indicators (metrics) to determine whether the surrogate models can perform the above function. It is worth mentioning that the term “original model” and “surrogate model” are used in the following description to distinguish two models. In the thesis work, two indicators are implemented to compare the performances between the surrogate models and the original models.

The first indicator is the speed-up factor which aims to observe the enhancement based on the average calculating time of each surrogate model compared with the original model [68]. The SUF is defined on the basis of the simulation time, and it can be expressed as:

$$SUF = \frac{\frac{1}{N}\sum_{n=1}^{N} t_{ori,n}}{\frac{1}{N}\sum_{n=1}^{N} t_{sur,n}} \tag{5.29}$$

where $t_{ori,n}$ is the simulation time of each repetition (noted as n) based on the original model; $t_{sur,n}$ is the simulation time of each repetition based on the surrogate model; N is the number of repetitions, and its value is equal to 10.

The second indicator is the RMSE whose main objective is to estimate the prediction accuracy of the surrogate models. RMSE can be defined in the following equation.

$$RMSE = \sqrt{\frac{1}{N \cdot M \cdot L}\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{l=1}^{L}\left(y_{nml} - \hat{y}_{nml}\right)^2} \tag{5.30}$$

where n is the number of repetitions of the experiment. y_{nml} is an element in **Ya** (matrix of actual values of outputs), while $\hat{y}_{nml}$ is an element in **Yp** (matrix of predicted values of outputs). The dimension of **Ya** and **Yp** is $L \times M$. M is the number of output variables which represent predicted bus and temperature profiles in the electrical network and heating network, respectively. L is the number of samples. As mentioned earlier, all the surrogate models considered in this section do not consider the sequential information in the input data. It means each $l \in L$ is a snapshot. As for the electrical network, the snapshot can be depicted that voltage values of buses are obtained using power flow calculation. As for the heating network, the snapshot can be described that temperatures are sampled at every 15 minutes from the result of the dynamic simulation. In short, the surrogate models in each network predict values of desired output variables for a given snapshot of input variables. To give a better illustration, Figure 5.17 (a) and 5.17 (b) describe the dataframe of **Ya** and **Yp** when n is equal to 1. Table 5.3 shows the key parameters when calculating the above performance indicator.

Samples	Output Variable 1	Output Variable 2	…	Output Variable m
1	$y_{1,1}$	$y_{1,2}$	…	$y_{1,m}$
2	$y_{2,1}$	$y_{2,2}$	…	$y_{2,m}$
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
l	$y_{l,1}$	$y_{l,2}$	…	$y_{l,m}$

(a)

Samples	Output Variable 1	Output Variable 2	…	Output Variable m
1	$\hat{y}_{1,1}$	$\hat{y}_{1,2}$	…	$\hat{y}_{1,m}$
2	$\hat{y}_{2,1}$	$\hat{y}_{2,2}$	…	$\hat{y}_{2,m}$
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
l	$\hat{y}_{l,1}$	$\hat{y}_{l,2}$	…	$\hat{y}_{l,m}$

(b)

Figure 5.17: The description of the dataframe of **Ya** and **Yp** in two networks. (a). The dataframe of **Ya** (b). The dataframe of **Yp**.

Table 5.3: Key parameters when calculating performance indicators in the electrical and heating network.

	Number of Total Samples	**L**	**M**	**N**
Electrical Network	35040	7008	7	10
Heating Network	1152	231	6	10

6

Results & Discussions

The objective of the chapter is to provide simulation results based on the original model and surrogate models in the electrical network and heating network, respectively. First, seven surrogate models are taken into account initially, and their performances based on two metrics (speed-up factor and root mean square error) in each network are given, respectively. The framework of these models are based on Scikit-learn. The models consist of linear regression, linear regression with regressor chains, linear support vector machine with regressor chains, decision tree, random forest, k-nearest neighbours as well as multilayer perceptron. Second, one extra surrogate model called long short-term memory model is considered, and its framework is based on TensorFlow. Its performance is offered and compared with the seven surrogate models in the heating network. The simulation in the thesis work is mainly coded in Python 3.7 and performed on a HP personal computer with Intel (R) Core (TM) i7-8750 CPU @ 2.20 Ghz and 16 GB RAM.

6.1 Metric 1: Speed-up Factor

Surrogate models are introduced to improve the computation efficiency by replacing expensive parts in the original model. Hence, the metric related to the simulation time is taken into account initially. First, the results based on the simulation time of the original model is given. Second, the results based on the simulation time of the surrogate models are provided. Finally, the performances between the original model and the surrogate models are compared. Speed-up factor, defined on the basis of the simulation time, is used to carry out the related assessment.

6.1.1 Simulation Time of the Original Model

As mentioned in Chapter 5, in order to avoid the randomness brought by the single simulation and make the results persuasive, the simulation of each network in the original model is repeated 10 times. Table 6.1 shows the the simulation time of the electrical network when the simulated days are

Table 6.1: The simulation time of the electrical network.

	Number of Simulations	Max. Time (s)	Min. Time (s)	Mean Time(s)
Electrical Network	10	1742.649	1439.393	1565.541

Table 6.2: The simulation time of the heating network.

	Number of Simulations	Max. Time (s)	Min. Time (s)	Mean Time (s)
Heating Network	10	43304	32544	38851

one-year, and Table 6.2 illustrates the simulation time of the heating network when the simulated days are 12 representative days. It can be seen that computation costs are expensive in two networks especially the time in the heating network reaches 38851s when the simulated days are only 12 days. Hence, it is necessary to introduce surrogate models to enhance the computation efficiency.

6.1.2 Simulation Time of Surrogate Models

Similarly, the simulation of each surrogate model is also repeated 10 times, and the simulation time is recorded. Table 6.3 and Table 6.4 show results based on the simulation time of the electrical network and heating network, respectively.

Table 6.3: Performance comparisons of the simulation time for different surrogate models in the electrical network whose simulated days are one-year. Bold values mark the best performance in the column of the mean value.

	Number of Simulations	Max. Time (s)	Min. Time (s)	Mean Time (s)
LR	10	29.856	20.407	**23.430**
RC-LR	10	83.076	68.334	74.481
RC-SVR	10	4403.709	3989.231	4244.119
DT	10	71.411	55.996	63.267
RF	10	5758.628	1844.002	3101.228
k-NN	10	284.651	247.183	258.927
MLP	10	1324.711	712.607	1028.135

Table 6.4: Performance comparisons of the simulation time for different surrogate models in the heating network whose simulated days are 12 representative days. Bold values mark the best performance in the column of the mean value.

	Number of Simulations	Max. Time (s)	Min. Time (s)	Mean Time (s)
LR	10	1.512	1.081	1.193
RC-LR	10	2.827	2.157	2.324
RC-SVR	10	10.876	8.878	9.914
DT	10	0.428	0.331	**0.383**
RF	10	22.995	13.943	18.590
k-NN	10	1.866	1.960	1.904
MLP	10	245.397	210.411	223.798

In the electrical network, it is clear that LR model has the fastest computation speed (23.430 s), followed by the DT model (63.267 s) and RC-LR model (74.481 s). However, the performances of RC-SVR model (4244.119 s) and RF model (3101.228 s) are not satisfactory. They have longer simulation time than the other models, and hence their computation speeds are lower. In the heating network, it can be found that DT model has the best performances with the highest computation speed (0.383 s), followed by the LR model (1.193 s) and k-NN model (1.904 s). Whereas, MLP model has the worst performance with the longest simulation time (223.798 s).

6.1.3 Performance Comparisons Based on Metric 1

Having obtained the simulation results of both the original model and the surrogate modes, it is important to compare their differences in light of the simulation time. First, one-way analysis of variance (ANOVA) is used to know about whether there are any statistically significant differences of the simulation time between these models. The detailed results of the analysis, offered in Appendix, prove that the differences between these models are significant. Second, in order to reflect the differences between them explicitly, the values of the speed-up factor are calculated.

Table 6.5 combines the simulation time of the original model with the surrogate models in the electrical network. Here, the time of the surrogate models uses the mean time of 10 repeated simulation results based on Table 6.1 and Table 6.3. Furthermore, the simulation time of different models is plotted together as shown in Figure 6.1. Except RC-SVR model and RF model, the other surrogate models show faster computation speed compared with the original model. It can be better depicted by one of the performance indicators, speed-up factor. As mentioned in Chapter 5, it can reflect how much faster the surrogate models' computation speed in comparison with the original model's computation speed. It can be observed that LR model has the most remarkable speed-up factor (66.818) with the fastest computation speed. Furthermore, the performances of DT model (24.745) and RC-LR model (21.019) are also notable. However, the values of the speed up factors of RC-SVR model and RF model are even less than 1, which means their simulation time are longer than the original model. Hence, their performances in terms of the speed-up factor are not satisfactory.

Table 6.5: A summary based on the first metric reflecting the performances of the original model and different surrogate models in the electrical network. Bold values mark the best performance in the corresponding column.

	Number of Simulations	Simulation Time (s)	Speed-up Factor
Original Model	10	1565.541	1
LR	10	**23.430**	**66.818**
RC-LR	10	74.481	21.019
RC-SVR	10	4244.119	0.369
DT	10	63.267	24.745
RF	10	3101.228	0.505
k-NN	10	258.927	6.046
MLP	10	1028.135	1.523

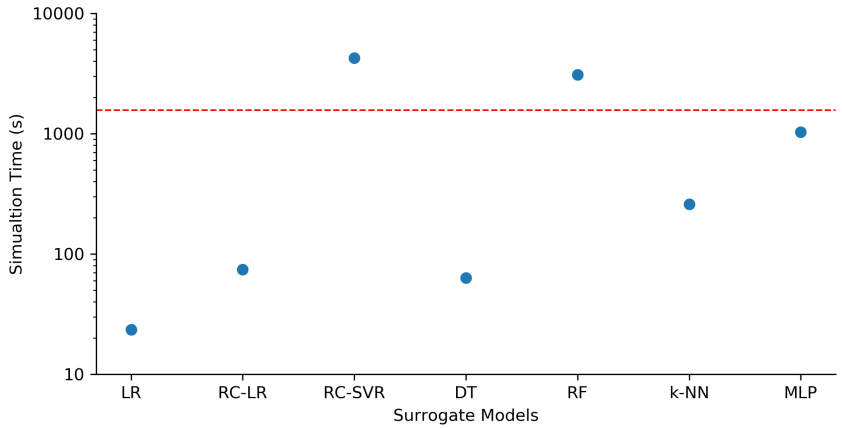


Figure 6.1: Performance comparisons of simulation time between the original model and the surrogate models in the electrical network. Solid dots show the simulation time of different surrogate models. Red dash line represents the simulation time of the original model.

The simulation time of the original model and the surrogate models in the heating network is illustrated in Table 6.6. Here, the time of the surrogate models also uses the mean time based on Table 6.2 and Table 6.4. Figure 6.2 describes their performances of the simulation time. Similar as the previous work in the electrical network, the ANOVA analysis is carried out, and it shows the differences between these models are significant. The results are also illustrated in Appendix. It can be seen that all the surrogate models show faster simulation speed in comparison with the original model. Especially, DT model has the most excellent performances (1.01E+05), followed by LR model (3.26E+04) and k-NN model (2.04E+04). Whereas, MLP model has the lowest speed-up factor (1.74E+02).

Table 6.6: A summary based on the first metric reflecting the performances of the original model and different surrogate models in the heating network. Bold values mark the best performance in the corresponding column.

	Number of Simulations	Simulation Time (s)	Speed-up Factor
Original Model	10	38851	1
LR	10	1.193	3.26E+04
RC-LR	10	2.324	1.67E+04
RC-SVR	10	9.914	3.92E+03
DT	10	**0.383**	**1.01E+05**
RF	10	18.590	2.09E+03
k-NN	10	1.904	2.04E+04
MLP	10	223.798	1.74E+02

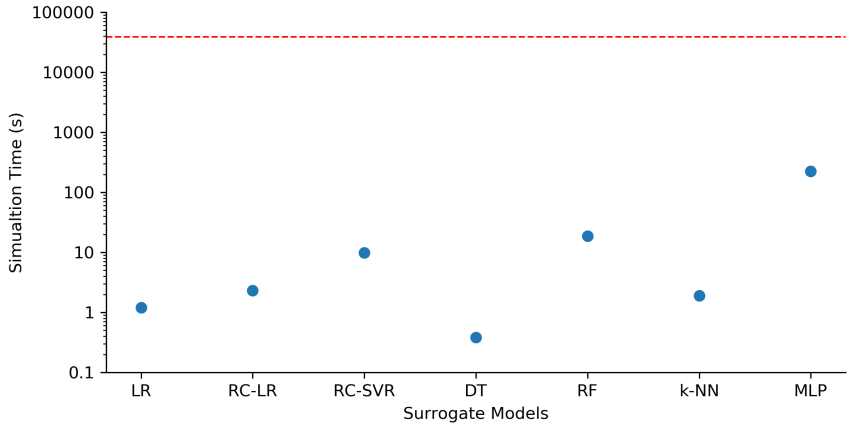


Figure 6.2: Performance comparisons of simulation time between the original model and the surrogate models in the heating network. Solid dots show the simulation time of different surrogate models. Red dash line represents the simulation time of the original model.

6.2 Metric 2: Root Mean Square Error

As stated in Chapter 5, the essential characteristic of the surrogate model is to mimic the behaviour of the computation codes in the original model. Thus, it is necessary to evaluate accuracy of the models, and RMSE is considered as the second metric. Similar as the first section, simulation results of the original model and surrogate models in each network are given.

6.2.1 Simulation Results of the Original Model

In the electrical network, the output data focused in the thesis work is the voltage magnitude of each bus based on the rural area. Figure 6.3 illustrates yearly voltage profiles in per unit based on one typical bus (Bus 8) in the area. It can be seen that the voltage values change from 1.019 per unit to 1.026 per per unit. The results can be validated according to standard EN 50160 which reflects voltage characteristics of public distribution systems [155]. In the European standard, it offers references for important voltage parameters with the related requirements. For instance, it exposes the limitation that the voltage magnitude variations should have a tolerance of +/- 10% of the nominal voltage (0.9 per unit to 1.1 per unit) for MV and LV networks. It can be found that the voltage magnitude values in the figure (1.019 per unit to 1.026 per unit) vary in the range (0.9 per unit to 1.1 per unit), which proves that the results are reasonable according to the standard.

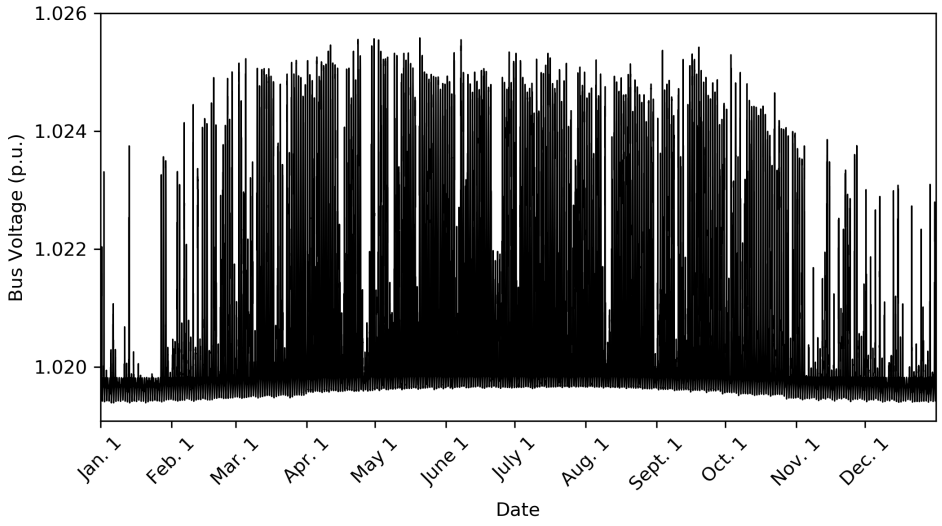


Figure 6.3: The simulation results based on the outputs of Bus 8 in the electrical network.

In the heating network, its outputs are supply temperatures and storage temperatures based on the rural area. As mentioned earlier, it will take several days to undertake the simulation in the network when its simulated days are one year. Thus, in order to simplify the workload, 12 representative days (3 simulated days per season) in one year are selected. It is worth mentioning that the simulation in each group (season) is carried out independently. As a result, the output data of each group can be collected. After combining the data of these groups, the temperature profiles based on 12 typical days are finally obtained. Figure 6.4 depicts temperature profiles in Kelvin (K) based on 12 simulated days. The output data of group 1 ("Supply_Temp1" and "Temp1") in Figure 4.3 is taken as the supply temperature profile and storage temperature profile, respectively.

It can be found that supply temperatures keep constant after experiencing a short increase in the initial period. The storage temperatures have obvious increasing and decreasing periods based on the operation strategies (charging state and discharging state) of the designed controllers.

6.2.2 Simulation Results of Surrogate Models

In each network, surrogate models are established based on different machine learning algorithms. As mentioned in Chapter 5, the algorithms are used to cope with the regression problem in the thesis work. Considering regression is one of the main branches in the predictive modelling, and hence the prediction accuracy of the algorithms should be concentrated. An intuitive method is to observe the outputs of the surrogate models and compare them with the outputs of the original model.

Figure 6.5 (a) illustrates actual results and prediction results of one simulated day in the electrical network, and Figure 6.5 (b) illustrates the zoomed result of the Figure 6.5 (a) for the sake of showing differences between surrogate models clearly. The bus selected here (Bus 8) is the same as in Figure 6.3 based on the rural area. Also, in order to make a better observation, the actual results provided by the original model are also plotted in the figures to offer a reference. It can be seen that the performances of surrogate models are various in the electrical network. Compared with other surrogate models, DT model and k-NN model show bigger prediction errors.

Figure 6.6 describes prediction results of one simulated day in the heating network. It is clear that prediction performances of surrogate models have obvious differences between supply temperatures

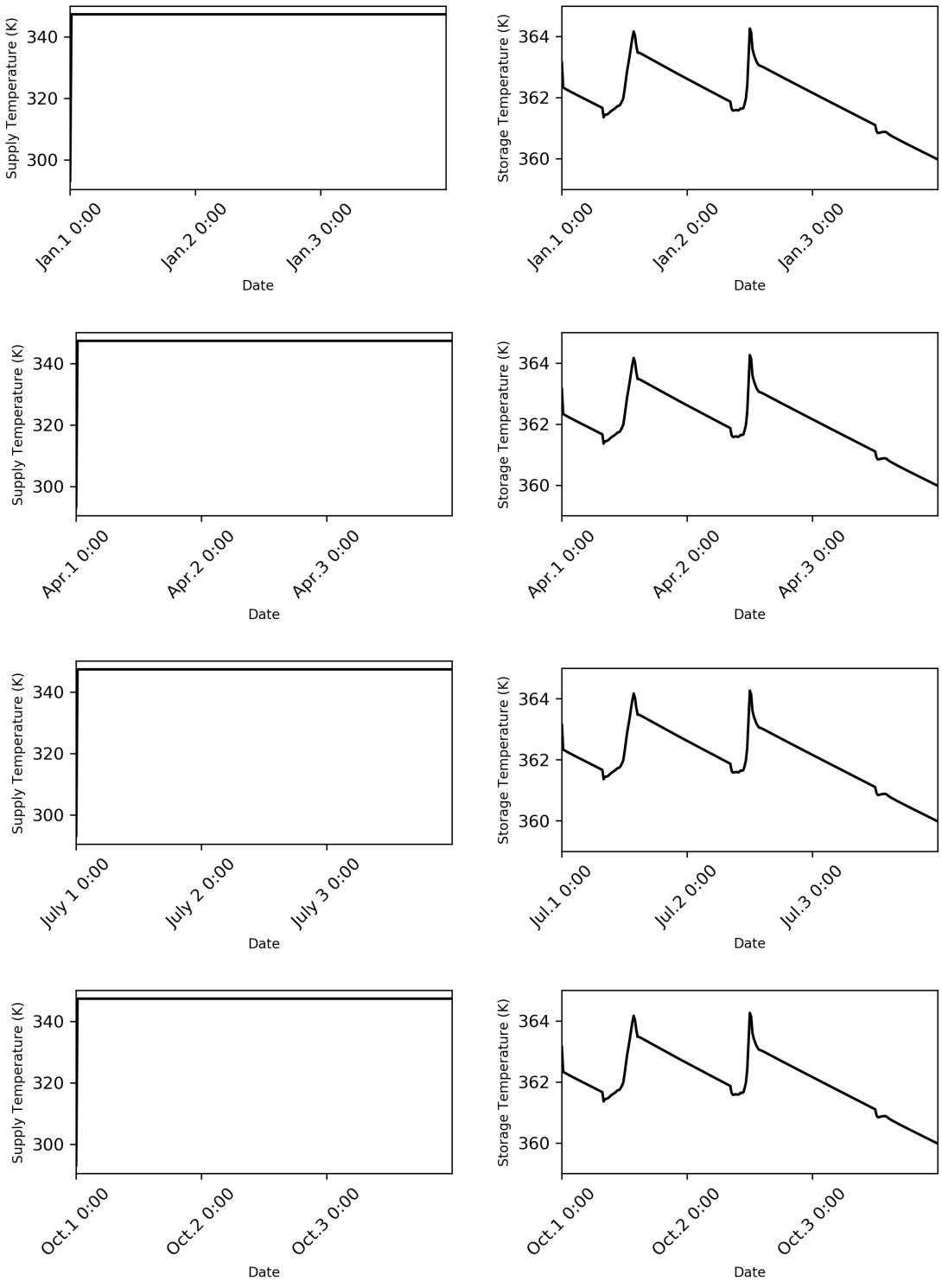


Figure 6.4: The simulation results based on the outputs of group 1 in the heating network. The left and right column show supply temperatures and storage temperatures of three representative days in each season. The subfigures from top to bottom in each column represent the temperature profiles of January, April, July and October, respectively).

and storage temperatures. In terms of the supply temperatures, except MLP model, other surrogate models have excellent performances with small errors. The performance of MLP model is not satisfactory. In terms of the storage temperatures, surrogate models show various performances clearly. It can be found that some surrogate models can give correct prediction results at specific points. However, none of the models have satisfactory performances in the whole period.

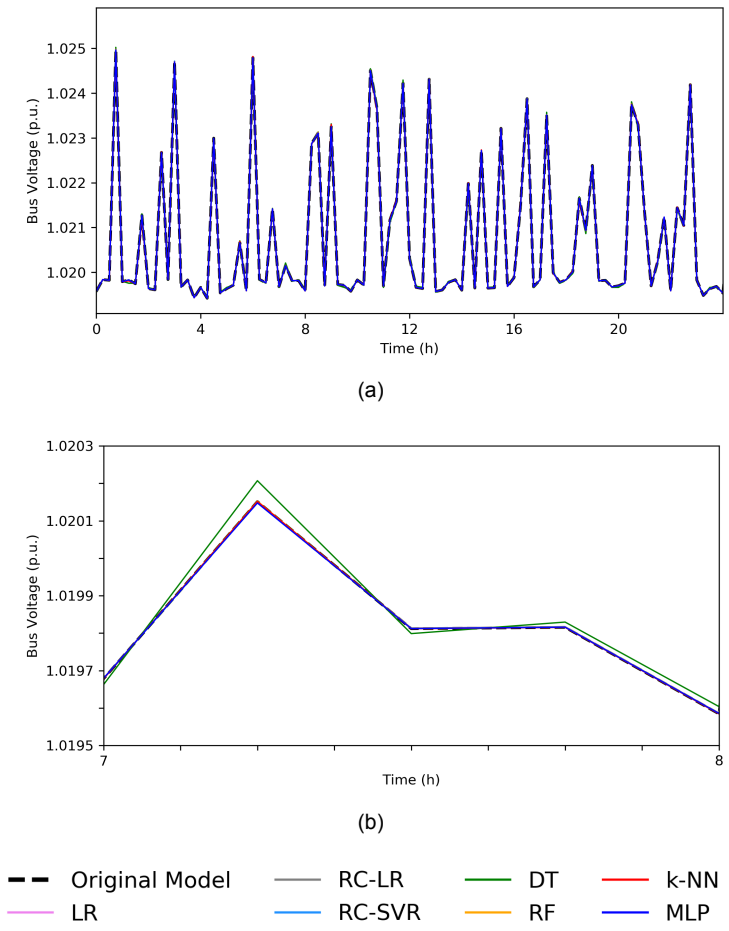


Figure 6.5: The simulation results of the outputs based on Bus 8 between different surrogate models in the electrical network. The result of the original model is also illustrated as a reference. (a). The simulation result of surrogate models based on one simulated day. (b). Zoomed result of (a).

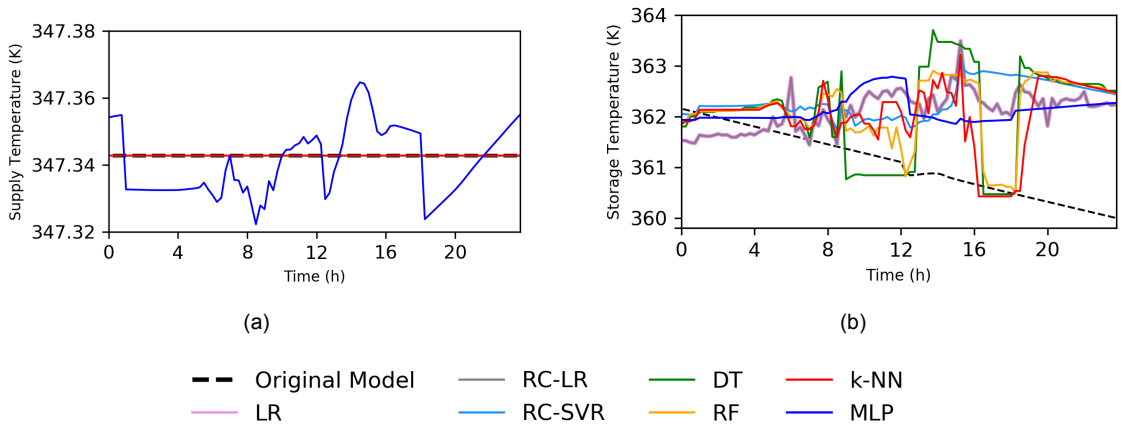


Figure 6.6: The simulation results based on the outputs of group 1 in the heating network. The subfigures from left to right represent the temperature profiles of supply temperature and storage temperature, respectively. The results of the original model are also illustrated as a reference.

6.2.3 Performance Comparisons Based on Metric 2

When comparing various surrogate models in the last subsection, the differences between the original model and the surrogate models in each network can also be observed in Figure 6.5 and Figure 6.6, respectively. In order to quantify the comparison results, RMSE is introduced to calculate the differences between the actual values and the predicted values.

Table 6.7 shows average RMSE values of different surrogate models in the electrical network, and Figure 6.7 depicts the boxplot of yearly RMSE values of the surrogate models. It can be found that the performances of seven surrogate models are different. RF model (6.516E-06) has the best performance, followed by LR model (7.494E-06) and RC-LR model (7.498E-06). Compared with other models, the performances of DT model (30.716E-06) and k-NN model (25.878E-06) are not satisfactory. The results corresponds to the previous observations in Figure 6.5.

Table 6.7: Average RMSE values of surrogate models relative to the original model in the electrical network.

	Original Model	LR	RC-LR	RC-SVR	DT	RF	k-NN	MLP
RMSE ($\times 10^{-6}$)	-	7.494	7.498	8.109	30.716	**6.516**	25.878	12.802

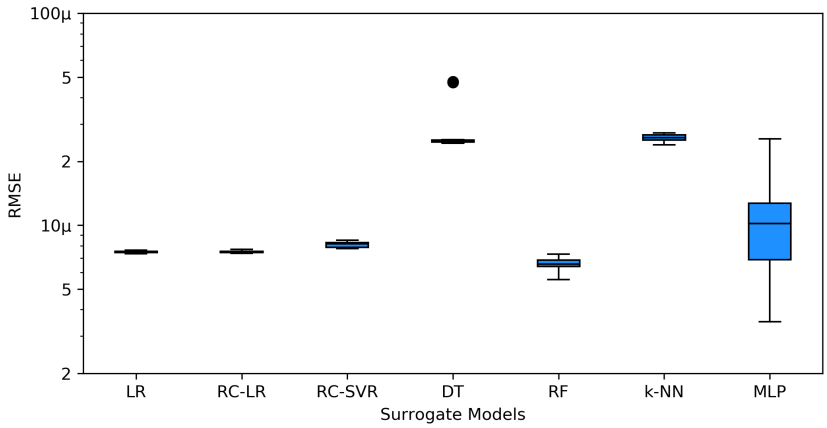


Figure 6.7: Performance comparisons of RMSE values for different surrogate models in the electrical network when the simulated days are one year.

Surrogate models in the heating network are also studied by calculating their RMSE values.Table 6.8 shows average RMSE values of seven surrogate models in the heating network, and Figure 6.8 depicts the boxplot of RMSE values in the surrogate models. It can be seen that RF model has the lowest RMSE values (0.685). However, its advantage is not notable compared with other surrogate models. All the surrogate models have high RMSE values which are above 0.5. The results can also be reflected by the observation in Figure 6.6.

Table 6.8: Average RMSE values of surrogate models relative to the original model in the heating network.

	Original Model	LR	RC-LR	RC-SVR	DT	RF	k-NN	MLP
RMSE	-	0.747	0.757	0.783	0.757	**0.685**	0.738	0.727

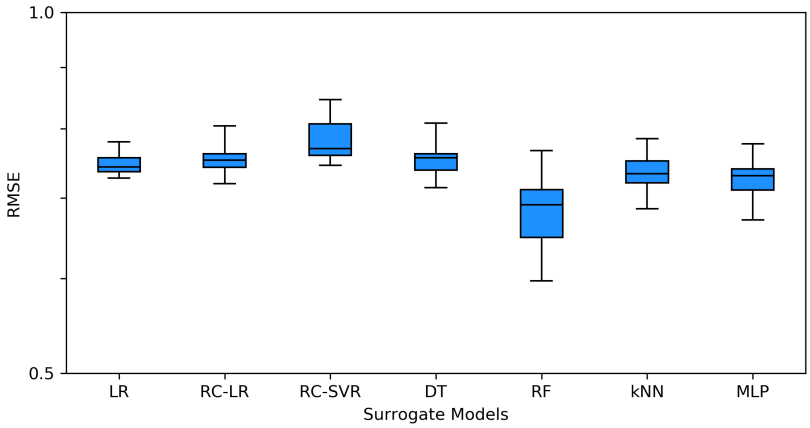


Figure 6.8: Performance comparisons of RMSE values for different surrogate models in the heating network when the simulated days are 12 representative days.

6.3 Another surrogate model: LSTM model

Having obtained the above simulation results, it can be found that the performances of all the surrogate models used in the thesis work are not satisfactory in the heating network. Hence, it is necessary to consider other surrogate models. Long short-term memory model is studied as another surrogate model in the thesis work. As mentioned in Chapter 5, as a deep learning model, LSTM network has obvious advantages when handling time-series data. Stacked LSTM model is concentrated, and its structure is shown in Figure 5.13. Similar as the other models, the same metrics are employed to assess the performances of the model in the heating network. It should be noted that, although LSTM model is introduced for the heating network in the thesis work, its performance is also studied in the electrical network to conduct a comprehensive analysis, and its simulation results are listed in the appendix chapter.

6.3.1 The description of LSTM model in the heating network

Compared with other surrogate models in the heating network as shown in Figure 5.16 (b), it can be found that the main difference is the input dataset of the LSTM model takes into account previous values of the output dataset (temperature profiles in the heating network). Then, the snapshot data is transferred to timeseries data via the timeseries generator. In consequence, the output values are predicted, and hence the model is capable of capturing the sequential information in the input data. Figure 6.9 shows the LSTM model in the heating network.

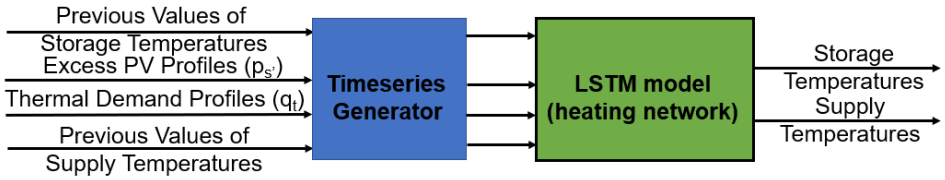


Figure 6.9: The illustration of LSTM model in the heating network.

6.3.2 Simulation results based on the metric 1 in the heating network

A summary based on simulation results of the LSTM model in the heating network is given by Table 6.9. Similar as other surrogate models, the simulation time here also uses the average time based

on 10 repeated simulation. It can be seen that its simulation time (51.880) is not notable (only faster than the MLP model) in comparison with the aforementioned surrogate models as listed in Table 6.6. However, it still has a better performance in terms of the computation speed when comparing with the original model in the network.

Table 6.9: A summary of simulation time (s), speed-up factor and RMSE of LSTM model in the heating network.

	Simulation Time (s)	Speed-up Factor	RMSE
heating network	51.880	748.863	0.376

6.3.3 Simulation results based on the metric 2 in the heating network

In terms of the metric 2, the RMSE value of the LSTM model is also shown in Table 6.9. In order to make a clear comparison, RMSE values of various surrogate models after considering the LSTM model in the heating network is depicted in Figure 6.10. Compared with the other surrogate models, the LSTM model shows its obvious advantage with the lowest RMSE value (0.376) which is lower than 0.5.

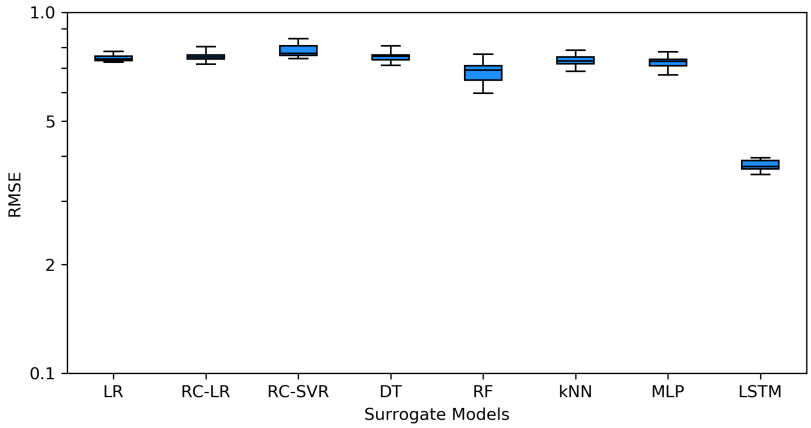


Figure 6.10: Performance comparisons of RMSE values for eight surrogate models in the heating network when the simulated days are 12 representative days.

Figure 6.11 shows the prediction results of the supply temperature and storage temperature of the LSTM model based on one simulated day in the heating network. It is worth mentioning that, although the prediction results of the LSTM model is better than the above surrogate models, it still exist errors. One of the main reasons is the model has many hyperparameters such as size of the hidden state, number of layers, batch size, epoch number, activation function type, optimiser type, etc. These hyperparameters not only affects the performance of the model in terms of the metric 2, but also plays an important role in influencing the performance in terms of the simulation time (metric 1). An advanced tuning approach, such as HyperOpt [156], can be taken into account to find the appropriate hyperparameters for the LSTM model in the future work. Moreover, the LSTM model has other types such as bidirectional LSTM, encoder-decoder LSTM, CNN-LSTM, ConvLSTM, etc. Hence, the model needs more studies in the further research.

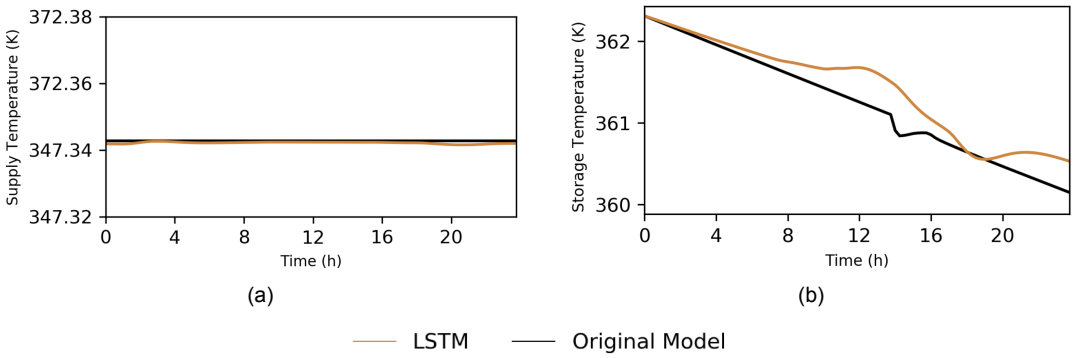


Figure 6.11: The simulation results of the outputs based on group 1 in the LSTM model of the heating network. (a). The supply temperature prediction results; (b). The storage temperature prediction results. The result of the original model is shown as a reference.

6.4 Performance Summary of Surrogate Models Based on Metric 1 and Metric 2

Table 6.10 summarises the simulation results based on Table 6.5, Table 6.7 and the LSTM performance. In the electrical network, in terms of the metric 1, LR model has the best performances, followed by DT model and RC-LR model. However, RC-SVR model and RF model have the worst performance, and their simulation time is even more than the original model. In terms of the metric 2, the performance of RF model is the best, followed by LR model and RC-LR model. Whereas, LSTM model (analysed in Appendix) has the worst performance. DT model and k-NN model are not satisfactory with higher RMSE values. Hence, according to the general performances based on two metrics, LR model shows the most excellent performance compared with other surrogate models used in the thesis work.

Table 6.10: A summary based on two metrics reflecting the performances of the original model and different surrogate models in the electrical network. Bold values mark the best performance in the corresponding column.

	Number of Simulations	Simulation Time (s)	Speed-up Factor	RMSE
Original Model	10	1565.541	1	-
LR	10	**23.430**	**66.818**	7.494E-06
RC-LR	10	74.481	21.019	7.498E-06
RC-SVR	10	4244.119	0.369	8.109E-06
DT	10	63.267	24.745	3.072E-05
RF	10	3101.228	0.505	**6.516E-06**
k-NN	10	258.927	6.046	2.588E-05
MLP	10	1028.135	1.523	1.280E-05
LSTM	10	2381.586	0.657	8.335E-05

Table 6.11 summarises the simulation results based on Table 6.6, Table 6.8 and the LSTM performance. In the heating network, in light of the metric 1, DT model has the fastest simulation time, followed by LR model and k-NN model. MLP model has the worst performance compared with other surrogate models. In light of the metric 2, the performances of seven models (LR, RC-LR, RC-SVR, DT, RF, k-NN, MLP) are all not satisfactory with high prediction errors, and their prediction output curves cannot follow the actual output curve. Hence, LSTM model is studied as an extra surrogate model. Although its performance of simulation time is not notable, it has a unique advantage in terms of RMSE value with a smaller prediction error. Therefore, based on the general performances based on two metrics, LSTM model shows the best performance compared with other surrogate models used in the thesis work.

Table 6.11: A summary based on two metrics reflecting the performances of the original model and different surrogate models in the heating network. Bold values mark the best performance in the corresponding column.

	Number of Simulations	Simulation Time (s)	Speed-up Factor	RMSE
Original Model	10	38851	1	-
LR	10	1.193	3.26E+04	0.747
RC-LR	10	2.324	1.67E+04	0.757
RC-SVR	10	9.914	3.92E+03	0.783
DT	10	**0.383**	**1.01E+05**	0.757
RF	10	18.590	2.09E+03	0.685
k-NN	10	1.904	2.04E+04	0.738
MLP	10	223.798	1.74E+02	0.727
LSTM	10	51.880	7.49E+02	**0.376**

7

Conclusion & Outlook

The objective of the chapter is to provide final conclusions and future outlooks. First, research questions raised in the introduction are answered. Second, main contributions in the thesis work are presented. Finally, several recommendations with potential research directions are illustrated for the future work.

7.1 Research Questions

The core research questions with their sub-questions posed in Chapter 1 are taken up again with detailed answers:

1. **How can a representative multi-carrier distribution network be modelled?**

   As described in Chapter 2, electricity is not the only energy sector in the multi-carrier energy system. In the thesis work, a detailed hybrid network related to the energy system is established with containing the electrical network with the focus on the distribution grid and the heating network. In order to combine two networks, co-simulation is employed by the simulation platform named as EnergySim. Furthermore, controllers are considered in the co-simulation and built by Open Modelica. The detailed modelling processes of the electrical network and the heating network are illustrated by answering the sub-questions below.

   - *How can a Dutch electrical distribution network be established as a static power flow model using open data sources?*

     Dutch electrical distribution network is built with the design of different electric parameters by means of pandapower. Furthermore, various areas (NZEB area, rural area and urban area) owned by three chief DSOs in the Netherlands are taken into account, and the low-voltage network with 104 households is especially focused. Moreover, owing to the rising distributed energy resources, PV modules are considered in the rural area which is treated as a use case.

   - *How can a representative heating network be designed and modelled to reflect a future tightly integrated energy system?*

     The heating network, also concentrating on the rural area, is built via OpenModelica with designing various heating parameters. The network involves supply system, pipe system, electric heater & storage system as well as demand system. To be specific, the electric heater & storage system plays an important role in the design of controllers whose control strategies consist of the charging state and the discharging state to offer the solution when excess PV power is produced in the electrical network. Furthermore, electric heaters are both considered in the network and the electrical network because the electrical power is converted to the heat power with a specific efficiency in their working states.

2. **What kind of surrogate models are used to create for the multi-carrier distribution networks?**

   When modelling the multi-carrier energy system (the original model) according to the solutions of the above question, one of the main limitations is the low computation efficiency with the long simulation time. Hence, surrogate models are introduced to replace the expensive parts in the original model. The following sub-questions are centred around the modelling process of the surrogate models.

   - *Which methods are selected to establish surrogate models?*

     As discussed in Chapter 2, various techniques of creating surrogate models have been proposed. As one of the popular approaches, machine learning algorithms are chosen as the main approach to establish the surrogate models. Furthermore, these exists a list of methods when dealing with the multi-output regression problems in the machine learning algorithms. In the thesis work, eight typical algorithms are taken into account to establish the corresponding surrogate models including linear regression model, linear regression with chain model, linear support vector with chain model, decision tree model, random forest model, k-nearest neighbour model, multilayer perceptron model as well as long short-term memory model.

   - *Which parts of the distribution networks are replaced by surrogate models?*

     In terms of the multi-carrier energy system in the thesis work, surrogate models are built for the electrical network and the heating network, respectively. In each network, eight machine learning algorithms are taken into account, and their performances are compared.

   - *What are the inputs and outputs of the training/testing data for surrogate models?*

     When machine learning algorithms are selected as creating methods of surrogate models, dataset processing should be taken as an important goal initially. The same datasets containing input datasets and output datasets are applied for different algorithms in the respective network. In the electrical network, its input datasets include PV and load power profiles, and its output datasets consist of voltage profiles in each bus. In the heating network, its input datasets contain excess PV profiles and demand profiles, and its output datasets involve storage temperatures and supply temperatures. These input datasets and output datasets are trained and tested in the machine learning algorithms to create the corresponding surrogate models.

3. **Up to which extent can the simulation performances of the multi-carrier distribution networks be enhanced by surrogate models?**

   Having established the original models and the surrogate models, it is essential to compare their performances. Hence, the comparison indicators need to be determined initially. Then, the best surrogate models, which represent the electrical network and the heating network in MCES, can be selected and summarised. The comparison and selection in each network are described by answering the subquestions below.

   - *Which indicators are used to compare the performances between surrogate models and original models?*

     One of the significant characteristics of the surrogate model is to mimic the behaviour of the computation codes in the original models. Thus, it is necessary to introduce performance indicators to compare the differences between the original models and the surrogate models. Two indicators are taken into account in the thesis work. The first indicator is speed-up factor defined on the basis of the simulation time. The second indicator is RMSE which

performs the function of comparing the performances with regard to the prediction ability.

- *Which surrogate models best represent the electrical network?*

  In the electrical network, the rural area of the MCES is concentrated, and the simulation period is chosen as one year. In terms of the speed-up factor, linear regression model has the best performances. Furthermore, the performances of decision tree model and linear regression with regressor chains are also notable. Whereas, the performances of linear support vector machine with regressor chains model and random forest model are not satisfactory with their simulation time is even more than the original model. In terms of the RMSE value, the performance of RF model is the best, followed by linear regression model and linear regression with regressor chains model. Whereas, long short-term memory model has the worst performance. Hence, linear regression model can be employed as a surrogate model to represent the electrical network based on the rural area of the MCES.

- *Which surrogate models best represent the heating network?*

  In the heating network, its simulation is still based on the rural area of the MCES, and the simulation period is selected as 12 representative days in one year. In light of the speed-up factor, decision tree model has the fastest simulation time, followed by linear regression model and k nearest neighbour model. Whereas, multilayer perceptron model has the worst performance compared with other surrogate models. In light of the RMSE value, long short-term memory model has the best performance with the lowest prediction errors. Thus, according to the general performances of two metrics, long short-term memory model can be taken as a surrogate model to represent the heating network based on the rural area of the MCES.

7.2 Contributions

The thesis work focuses on the research field of the machine learning algorithms employed in the energy system. The main contributions are outlined.

- A LV Dutch electrical distribution network has been established with the design of different electric parameters via pandapower. Furthermore, various areas (NZEB, rural, urban) owned by main Dutch DSOs are taken into account. Especially, the rural area is expanded with adding PV modules as a representative to evaluate the impact of RES integration. Therefore, it can be viewed as a benchmark when carrying out the further study of the Dutch electrical network.

- On the basis of the electrical distribution network, a representative multi-carrier energy system based on the rural area has been modelled by EnergySim. The co-simulation platform also connects other networks including the heating network and controllers built by Open Modelica. The heating network considers different subsystems, and controllers with their operational strategies resolve the excess PV power issue existing in the above electrical network. Hence, the energy system modelling helps to explore the potential of achieving the carbon emission reduction target.

- Surrogate models have been created for the electrical network and the heating network of the above energy system, respectively. In each network, eight representative machine learning algorithms are employed, and their performances are compared with considering various indicators. As a result, the best surrogate models are selected, which can be taken as a useful replacement when the computation costs of the original simulation models are expensive.

7.3 Recommendations and Future Works

The work in the thesis yields some promising results related to the renewable energy penetration, the

establishment of the multi-carrier energy system as well as surrogate models. However, several gaps still need to be filled and further research is required. Some recommendations and potential future research topics are described below:

- When modelling the electrical distribution network, the low voltage level is mainly concentrated in the thesis work without considering other levels such as the medium voltage level and the high voltage level. In order to carry out a comprehensive analysis, it is expected to build their models on the basis of the established LV network. Moreover, due to the rising penetration of the renewable energy, PV modules are especially taken into account in the rural area, and a similar treatment in the heating network is carried out. However, three areas are covered in the original LV network. Hence, other areas such as the NZEB area and the urban area can be further expanded. Then, the operation strategies of controllers are anticipated to have differences because detailed situations will vary from areas to areas. Furthermore, solar energy is not the only renewable source, other sources can be included in the further study.

- When modelling the multi-carrier energy system, the hybrid network consists of the electrical distribution network and the heating network. There exist other different energy sectors such as hydrogen energy, gas energy and so forth. Thus, it will be a future direction when adding other energy sectors or advanced techniques such as Power-to-X on the basis of the hybrid network with considering more interactions. In addition, the simulation of the energy system mainly puts an emphasis on two aspects: simulation time and simulation outputs. Some details with the relevant analysis in the energy system are ignored such as flexibility. Hence, undertaking the analysis in the further study can better contribute to getting a holistic of the multi-carrier energy system.

- Machine learning algorithms, as one of the popular approaches, are applied to create surrogate models based on the energy system. A total of eight typical machine learning algorithms are selected and used in each network. A more in-depth analysis and study can be undertaken by considering more surrogate models, such as gradient boosting, quantile regression, multivariate adaptative regression splines, in the machine learning algorithms to make detailed comparisons. Furthermore, as mentioned earlier, more works ought to be focused on the LSTM model with finding its appropriate hyperparameters by using advanced methods such as HyperOpt based on Bayesian hyperparameter optimisation. Moreover, a further extension is to create and find the best surrogate models for the whole multi-carrier distribution network even the multi-carrier energy systems, and the conclusion obtained in the thesis work related to the best surrogate models can be treated as a reference. Last but not least, a further investigation related to the surrogate models is to explore their potential in the application of digital twin in energy systems considering the relevant simulation is computationally expensive.

A

Transmission Network Modelling

TSOs perform the function in ensuring the extra high- & high-voltage electricity transmission. Chapter 3 offers the detailed background of Tennet which is the unique TSO in the Netherlands. Due to complicated structures of general transmission networks, a part of the network is only concentrated in the thesis work. Figure A.1 [1] offers the transmission grid map in the Netherlands provided by Tennet, and Figure A.2[2] and Figure A.3 show actual and simplified structure of the study area in the thesis work.

Figure A.1: The transmission grid map in the Netherlands (the study area is highlighted in blue).

[1] Source:https://www.tennet.eu/company/news-and-press/press-room/grid-maps/
[2] Source:https://www.hoogspanningsnet.com/netschema/

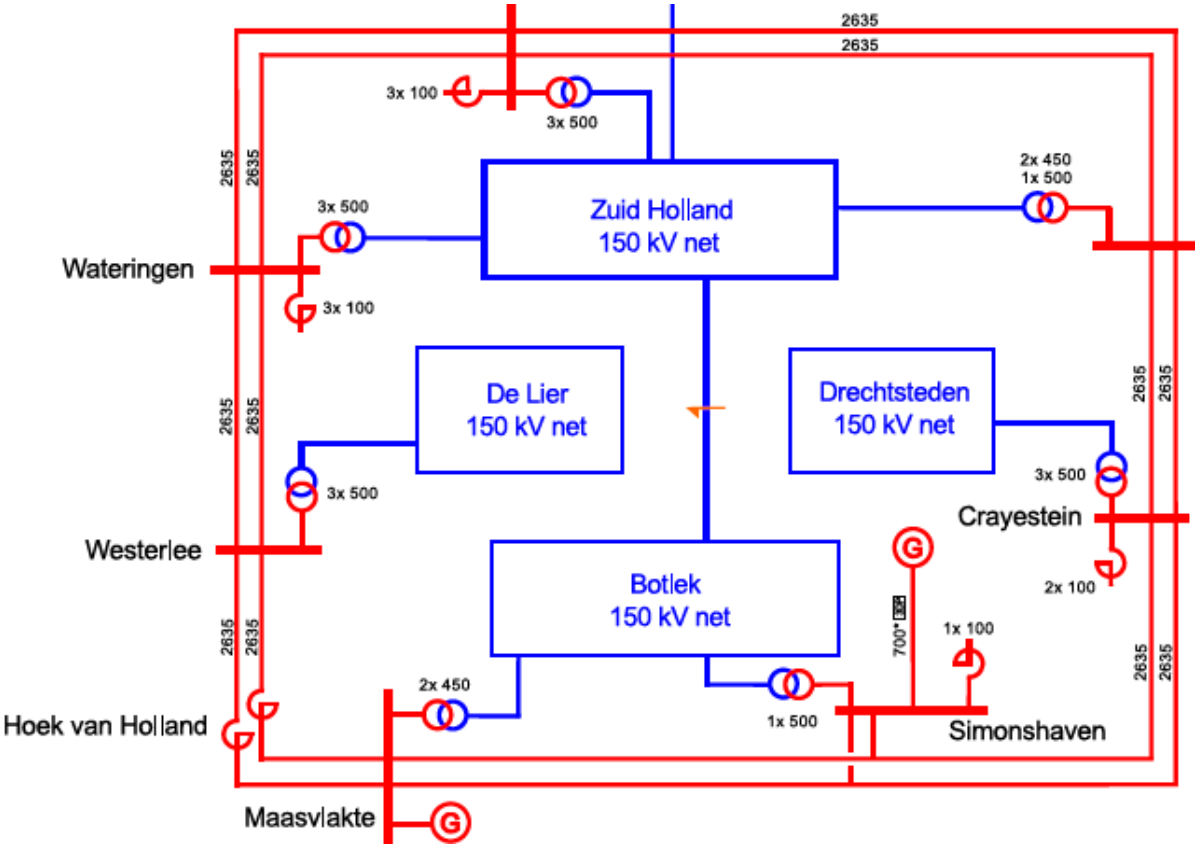


Figure A.2: The zoomed transmission grid map of the study area.

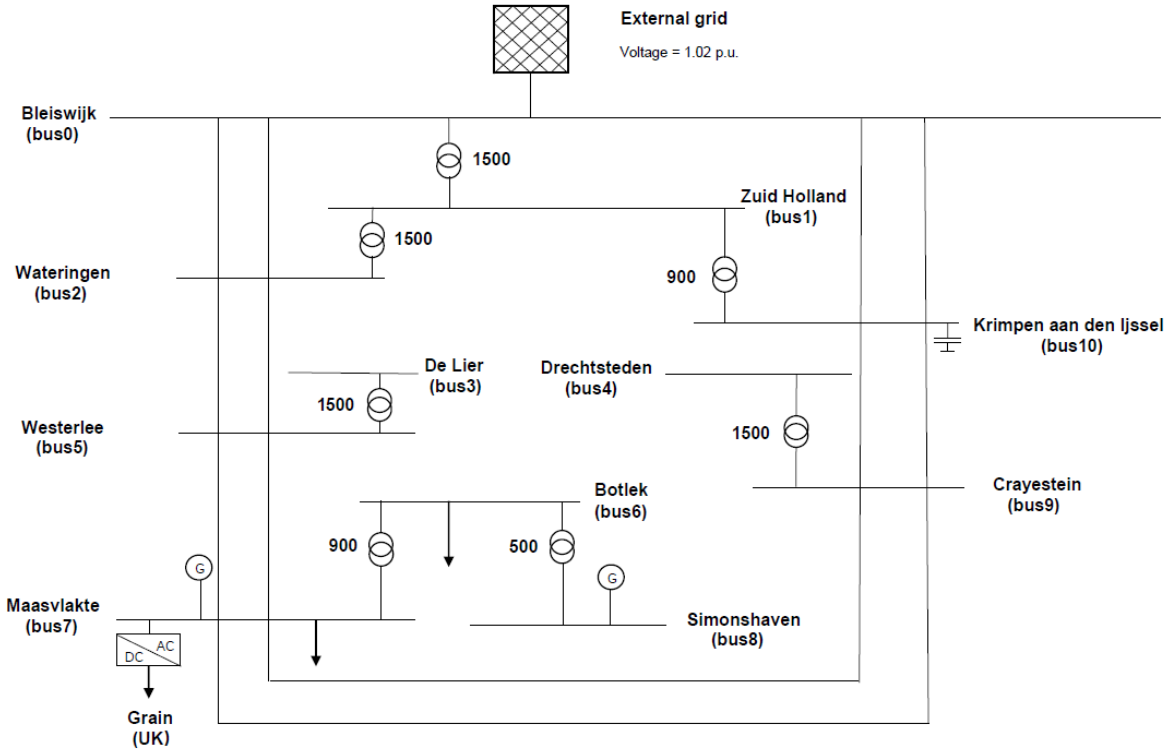


Figure A.3: The designed transmission network used in the thesis work based on the above zoomed area.

Table A.1: The simulation parameters of the external grid in the transmission network.

Index	Bus	vm_pu	va_degree
0	0	1.02	0

Table A.2: The simulation parameters of different buses in the transmission network.

Index	Name	vn (kv)
0	bus0	380
1	bus1	150
2	bus2	380
3	bus3	150
4	bus4	150
5	bus5	380
6	bus6	150
7	bus7	380
8	bus8	380
9	bus9	380
10	bus10	380

Table A.3: The simulation parameters of cable lines in the transmission network.

Index	Name	Type	From	To	l (km)	r (ohm/km)	x (ohm/km)	c (nf/km)	imax (ka)
0	HV Line0.1	679-AL1/86-ST1A 380.0	Bus0	Bus2	22	0.042	0.25	14.6	1.15
1	HV Line0.2	679-AL1/86-ST1A 380.0	Bus0	Bus2	22	0.042	0.25	14.6	1.15
2	LV Line0.3	679-AL1/86-ST1A 380.0	Bus2	Bus5	6.8	0.042	0.25	14.6	1.15
3	LV Line0.4	679-AL1/86-ST1A 380.0	Bus2	Bus5	6.8	0.042	0.25	14.6	1.15
4	LV Line0.5	679-AL1/86-ST1A 380.0	Bus5	Bus7	21.5	0.042	0.25	14.6	1.15
5	LV Line0.6	679-AL1/86-ST1A 380.0	Bus5	Bus7	21.5	0.042	0.25	14.6	1.15
6	LV Line0.7	679-AL1/86-ST1A 380.0	Bus7	Bus8	23	0.042	0.25	14.6	1.15
7	LV Line0.8	679-AL1/86-ST1A 380.0	Bus7	Bus8	23	0.042	0.25	14.6	1.15
8	LV Line0.9	679-AL1/86-ST1A 380.0	Bus8	Bus9	43.5	0.042	0.25	14.6	1.15
9	LV Line1.0	679-AL1/86-ST1A 380.0	Bus8	Bus9	43.5	0.042	0.25	14.6	1.15
10	LV Line1.1	679-AL1/86-ST1A 380.0	Bus9	Bus10	14.8	0.042	0.25	14.6	1.15
11	LV Line1.2	679-AL1/86-ST1A 380.0	Bus9	Bus10	14.8	0.042	0.25	14.6	1.15
12	LV Line1.3	679-AL1/86-ST1A 380.0	Bus0	Bus10	18.6	0.042	0.25	14.6	1.15
13	LV Line1.4	679-AL1/86-ST1A 380.0	Bus0	Bus10	18.6	0.042	0.25	14.6	1.15

Table A.4: The simulation parameters of generators in the transmission network.

Index	Name	Bus	p (mw)	vm (pu)
0	G1	7	3600	1.06
1	G2	8	400	1.06

Section	Type/Rating	Phase wire	Shield wires	Configuration	Line length (km)
Maasvlakte-Westerlee (1)	Line 4000A	st/al 52/591 4 bundle	Hawk 39/242	380 kV FD/T	12.9
Maasvlakte-Westerlee (2)	Cable 4000A	cu 2x1600 mm2		driehoek	2.1
Series Reactor	8 ohm/4000A				
Maasvlakte-Westerlee (3)	Line 4000A	st/al 52/591 4 bundle	Hawk 39/242	380 kV FD/T	6.5
Westerlee-Wateringen	Line 4000A	Bobolink 50/725 3 bundle	Hawk 39/242	380 kV DK/T	6.8
Wateringen-Bleiswijk (1)	Line 4000A	AMS 620 4 bundle	Hawk 39/242	Wintrack 4 circ. Combi	6
Wateringen-Bleiswijk (2)	Cable 4000A	cu 2x2500 mm2		platvlak	10
Wateringen-Bleiswijk (3)	Line 4000A	AMS 620 4 bundle	Hawk 39/242	Wintrack 2 circ.	6
Bleiswijk-Krimpen	Line 4000A	st/al 52/591 4 bundle	Hawk 39/242	380 kV FD/T	18.6
Bleiswijk-Vijfhuizen (1)	Line 2900A	AMS 620 3 bundle	Hawk 39/242	Wintrack 4 circ. Combi	7
Bleiswijk-Vijfhuizen (2)	Line 2900A	AMS 620 3 bundle	Hawk 39/242	Wintrack 2 circ.	3
Bleiswijk-Vijfhuizen (3)	Cable 2900A	cu 2x2500 mm2		platvlak	8
Bleiswijk-Vijfhuizen (4)	Line 2900A	AMS 620 3 bundle	Hawk 39/242	Wintrack 2 circ.	30
Vijfhuizen-Beverwijk (1)	Line 2900A	AMS 620 3 bundle	Hawk 39/242	Wintrack 2 circ.	8
Vijfhuizen-Beverwijk (2)	Cable 2900A	cu 2x2500 mm2		driehoek	2
Vijfhuizen-Beverwijk (3)	Line 2900A	AMS 620 3 bundle	Hawk 39/242	Wintrack 2 circ.	2
Beverwijk-Oostzaan	Line 2900A	st/al 40/457 3 bundle	Hawk 39/242	380 kV S+O/k	17
Oostzaan-Diemen	Line 2900A	st/al 40/457 3 bundle	Hawk 39/242	380 kV S+O/k	15.2
Diemen-Krimpen	Line 2500A	st/al 37/424 3 bundle	Hawk 39/242	380 kV S+O/k	57.6
Krimpen-Crayesteijn	Line 4000A	st/al 52/591 4 bundle	Hawk 39/242	380 kV DO/T	14.8
Crayesteijn-Simonshaven	Line 4000A	st/al 52/591 4 bundle	Hawk 39/242	380 kV DO/T	43.5
Simonshaven-Maasvlakte	Line 4000A	st/al 52/591 4 bundle	Hawk 39/242	380 kV DO/T	23

Figure A.4: Specifications 380 kV connections as provided by TenneT (source:Gockel.P.N.M, ”Steady-state voltage profile and reactive power balance for EHV AC cable systems in the Randstad 380 project”, Master Thesis, TU Delft, 2009).

B

Test Results of Designed Controllers

In order to get to know whether designed controllers normally work, they are tested in OpenModelica. The test method is to given initial parameters to observe whether they have different working states (charging state/discharging state) rather than keep constant outputs. As shown in Figure 4.5, temperatures need to be given initially. Here storage temperatures are settled as 65 $°C$, and supply temperatures as 60 $°C$. Figure B.1 illustrates the test results. To show a better description, a shorter time period is zoomed as shown in Figure B.2. To be specific, the ”1” and ”0” of the y-axis in the figures represent the working state and non-working state of the charge valves.

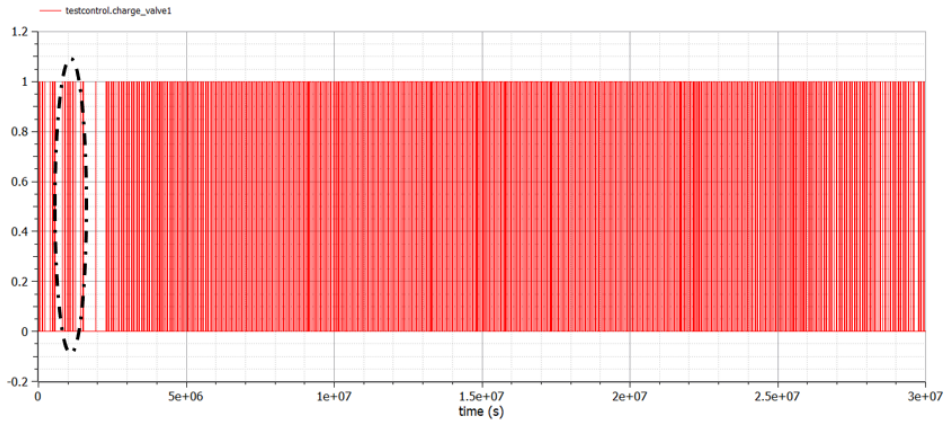


Figure B.1: The simulation result based on the tests results of controllers designed in the thesis work with the zoomed period is highlighted in black.

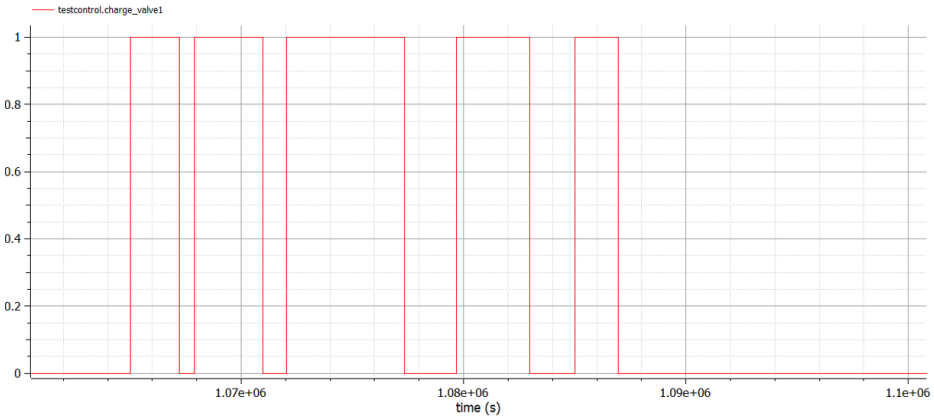


Figure B.2: A zoomed result of Figure B.1.

C

ANOVA Analysis

Table C.1: The ANOVA analysis based on the original model and surrogate models in the electrical network.

	sum (sq)	df	mean (sq)	F-value	p-value
C (treatments)	1.779674e+08	7.0	2.542391e+07	148.4699	**3.664794e-40**
Residual	1.232924e+07	72.0	1.712395e+05		

Table C.2: The ANOVA analysis based on the original model and surrogate models in the heating network.

	sum (sq)	df	mean (sq)	F-value	p-value
C (treatments)	1.253528e+10	7.0	1.790754e+09	1282.036812	**6.969416e-73**
Residual	1.005699e+08	72.0	1.396804e+06		

D

LSTM Model in the Electrical Network

Table D.1 shows a summary based on simulation results of the LSTM model in the electrical network. It should be noted that, when calculating the speed-up factor here, the simulation time of the original model is also obtained from Table 6.5 as the other surrogate models to make the corresponding comparisons.

Table D.1: A summary of simulation time (s), speed-up factor and RMSE of LSTM model in the electrical network.

	Simulation Time (s)	Speed-up Factor	RMSE
electrical network	2381.586	0.657	8.335E-05

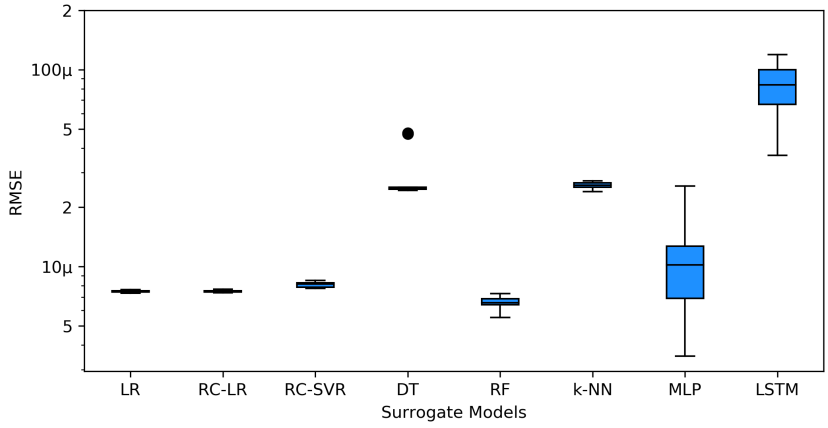


Figure D.1: Performance comparisons of RMSE values for bus 8 between eight surrogate models in the electrical network when the simulated days are one year.

In terms of the metric 1, its performance is not satisfactory because its speed-up factor value (0.657) is less than 1. It means the simulation time is even longer than the original model. In terms of the metric 2, Figure D.1 depicts the performance comparisons of RMSE values for various surrogate models after considering the LSTM model in the network. The performance of the LSTM model (8.335E-05) is not satisfactory with the highest RMSE value. Figure D.2 illustrates the prediction results of the LSTM model based on one simulated day in the electrical network.

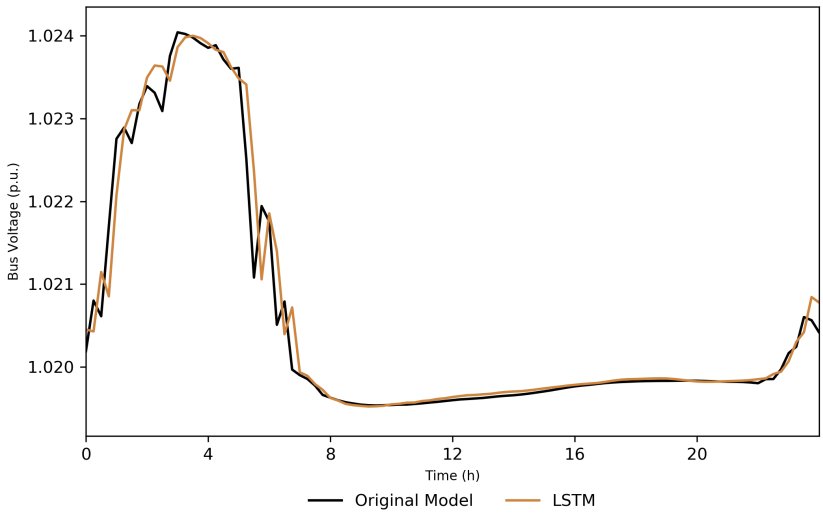


Figure D.2: The simulation results of the outputs based on Bus 8 in the LSTM model of the electrical network. The result of the original model is illustrated as a reference.

After considering the LSTM model, a performance summary can be finally made based on eight surrogate models in the electrical network, and the conclusion in the main text keeps constant. LR model has the best performance in all the surrogate models, used in the thesis work, of the electrical network.

Bibliography

[1] *Global climate report - september 2020 | state of the climate | national centers for environmental information (ncei)*, https://www.ncdc.noaa.gov/sotc/global/202009, (Accessed on 03/31/2021).

[2] *Global climate report - october 2020 | state of the climate | national centers for environmental information (ncei)*, https://www.ncdc.noaa.gov/sotc/global/202010, (Accessed on 03/31/2021).

[3] *Climate change: Global sea level | noaa climate.gov*, https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level, (Accessed on 03/31/2021).

[4] *Celex:52020dc0562:en:txt.pdf*, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0562&from=EN, (Accessed on 03/31/2021).

[5] *Paris agreement | climate action*, https://ec.europa.eu/clima/policies/internati/onal/negotiations/paris_en, (Accessed on 03/31/2021).

[6] K. Kulovesi and S. Oberthür, "Assessing the eu's 2030 climate and energy policy framework: Incremental change toward radical transformation?" *Review of European, Comparative & International Environmental Law*, vol. 29, no. 2, pp. 151–166, 2020.

[7] L. E. Jones, *Renewable energy integration: practical management of variability, uncertainty, and flexibility in power grids*. Academic press, 2017.

[8] R. Daiyan, I. MacGill, and R. Amal, *Opportunities and challenges for renewable power-to-x*, 2020.

[9] P. J. Agrell and P. Bogetoft, "International benchmarking of electricity transmission system operators," in *11th International Conference on the European Energy Market (EEM14)*, IEEE, 2014, pp. 1–5.

[10] S. Papathanassiou, N. Hatziargyriou, K. Strunz, *et al.*, "A benchmark low voltage microgrid network," in *Proceedings of the CIGRE symposium: power systems with dispersed generation*, CIGRE, 2005, pp. 1–8.

[11] M. J. O'Malley, M. B. Anwar, S. Heinen, T. Kober, J. McCalley, M. McPherson, M. Muratori, A. Orths, M. Ruth, T. J. Schmidt, *et al.*, "Multicarrier energy systems: Shaping our energy future," *Proceedings of the IEEE*, vol. 108, no. 9, pp. 1437–1456, 2020.

[12] P. Mancarella, "Multi-energy systems: The smart grid beyond electricity," *LBL, Berkeley*, 2012.

[13] ——, "Mes (multi-energy systems): An overview of concepts and evaluation models," *Energy*, vol. 65, pp. 1–17, 2014.

[14] H. Lund, P. A. Østergaard, D. Connolly, and B. V. Mathiesen, "Smart energy and smart energy systems," *Energy*, vol. 137, pp. 556–565, 2017.

[15] E. Widl, T. Jacobs, D. Schwabeneder, S. Nicolas, D. Basciotti, S. Henein, T.-G. Noh, O. Terreros, A. Schuelke, and H. Auer, "Studying the potential of multi-carrier energy distribution grids: A holistic approach," *Energy*, vol. 153, pp. 519–529, 2018.

[16] Q. Wu, A. H. Nielsen, J. Østergaard, S. T. Cha, and Y. Ding, "Impact study of electric vehicle (ev) integration on medium voltage (mv) grids," in *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, IEEE, 2011, pp. 1–7.

[17] N. G. Paterakis, O. Erdinç, and J. P. Catalão, "An overview of demand response: Key-elements and international experience," *Renewable and Sustainable Energy Reviews*, vol. 69, pp. 871–891, 2017.

[18] D. Z. Fitiwi, F. De Cuadra, L. Olmos, and M. Rivier, "A new approach of clustering operational states for power network expansion planning problems dealing with res (renewable energy source) generation operational variability and uncertainty," *Energy*, vol. 90, pp. 1360–1376, 2015.

[19] G. Prettico, M. Flammini, N. Andreadou, S. Vitiello, G. Fulli, and M. Masera, "Distribution system operators observatory 2018," *Publications Office of the European Union*, 2019.

[20] M. Visser, "Renewable energy in the netherlands: August 2019," 2019.

[21] L. Thurner, A. Scheidler, F. Schäfer, J.-H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, "Pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, 2018.

[22] P. Fritzson, P. Aronsson, A. Pop, H. Lundvall, K. Nystrom, L. Saldamli, D. Broman, and A. Sandholm, "Openmodelica-a free open-source environment for system modeling, simulation, and teaching," in *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*, IEEE, 2006, pp. 1588–1595.

[23] D. Gusain, M. Cvetković, and P. Palensky, "Energy flexibility analysis using fmuworld," in *2019 IEEE Milan PowerTech*, IEEE, 2019, pp. 1–6.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[25] *Tensorflow/tensorflow: An open source machine learning framework for everyone*, https://github.com/tensorflow/tensorflow, (Accessed on 03/31/2021).

[26] F.Chollet, *Github - keras-team/keras: Deep learning for humans*, https://github.com/keras-team/keras, (Accessed on 03/31/2021), 2015.

[27] *Verbruiksprofielen - nedu*, https://www.nedu.nl/documenten/verbruiksprofielen/, (Accessed on 03/31/2021).

[28] E. Widl, B. Leitner, D. Basciotti, S. Henein, T. Ferhatbegovic, and R. Hofmann, "Combined optimal design and control of hybrid thermal-electrical distribution grids using co-simulation," *Energies*, vol. 13, no. 8, p. 1945, 2020.

[29] *Eu science hub | the european commission's science and knowledge service*, https://ec.europa.eu/jrc/en, (Accessed on 03/31/2021).

[30] G. Prettico, F. Gangale, A. Mengolini, A. Lucas, and G. Fulli, "Distribution system operators observatory," *European Commission Joint Research Centre: Ispra, Italy*, 2016.

[31] M. Grzanic, M. G. Flammini, and G. Prettico, "Distribution network model platform: A first case study," *Energies*, vol. 12, no. 21, p. 4079, 2019.

[32] G. Aghajani, H. Shayanfar, and H. Shayeghi, "Demand side management in a smart micro-grid in the presence of renewable generation and demand response," *Energy*, vol. 126, pp. 622–637, 2017.

[33] C. Zhang, J. Wu, Y. Zhou, M. Cheng, and C. Long, "Peer-to-peer energy trading in a microgrid," *Applied Energy*, vol. 220, pp. 1–12, 2018.

[34] M. S. Ayaz, R. Azizipanah-Abarghooee, and V. Terzija, "European lv microgrid benchmark network: Development and frequency response analysis," in *2018 IEEE International Energy Conference (ENERGYCON)*, IEEE, 2018, pp. 1–6.

[35] M. Grond, B. Schepers, E. Veldman, J. Slootweg, and M. Gibescu, "Impact of future residential loads on medium voltage networks," in *2011 46th International Universities' Power Engineering Conference (UPEC)*, VDE, 2011, pp. 1–6.

[36] L. De Vos, N. Leemput, N. Refa, R. Bernards, H. Fidder, and F. de Rijke, "Optimal integration of electric vehicles, pv, heat pumps in existing distribution grids in the netherlands," 2019.

[37] M. Chertkov and G. Andersson, “Multienergy systems,” *Proceedings of the IEEE*, vol. 108, no. 9, pp. 1387–1391, 2020.

[38] A. S. R. Subramanian, T. Gundersen, and T. A. Adams, “Modeling and simulation of energy systems: A review,” *Processes*, vol. 6, no. 12, p. 238, 2018.

[39] A. Díaz-Manríquez, G. Toscano-Pulido, and W. Gómez-Flores, “On the selection of surrogate models in evolutionary optimization algorithms,” in *2011 IEEE congress of evolutionary computation (CEC)*, IEEE, 2011, pp. 2155–2162.

[40] J. Straus and S. Skogestad, “Surrogate model generation using self-optimizing variables,” *Computers & Chemical Engineering*, vol. 119, pp. 143–151, 2018.

[41] L. Riboldi and L. O. Nord, “Offshore power plants integrating a wind farm: Design optimisation and techno-economic assessment based on surrogate modelling,” *Processes*, vol. 6, no. 12, p. 249, 2018.

[42] M. Geidl and G. Andersson, “Optimal power dispatch and conversion in systems with multiple energy carriers,” in *Proc. 15th Power Systems Computation Conference (PSCC)*, 2005.

[43] M. Mohammadi, Y. Noorollahi, B. Mohammadi-Ivatloo, and H. Yousefi, “Energy hub: From a model to a concept–a review,” *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1512–1527, 2017.

[44] A. Perera, S. Coccolo, J.-L. Scartezzini, and D. Mauree, “Quantifying the impact of urban climate by extending the boundaries of urban energy system modeling,” *Applied Energy*, vol. 222, pp. 847–860, 2018.

[45] M. Arnaudo, M. Topel, and B. Laumert, “Techno-economic analysis of demand side flexibility to enable the integration of distributed heat pumps within a swedish neighborhood,” *Energy*, vol. 195, p. 117 012, 2020.

[46] M. Arnaudo, M. Topel, P. Puerto, E. Widl, and B. Laumert, “Heat demand peak shaving in urban integrated energy systems by demand side management-a techno-economic and environmental approach,” *Energy*, vol. 186, p. 115 887, 2019.

[47] B. Leitner, E. Widl, W. Gawlik, and R. Hofmann, “A method for technical assessment of power-to-heat use cases to couple local district heating and electrical distribution grids,” *Energy*, vol. 182, pp. 729–738, 2019.

[48] C. D. López, M. Cvetković, A. van der Meer, and P. Palensky, “Co-simulation of intelligent power systems,” in *Intelligent Integrated Energy Systems*, Springer, 2019, pp. 99–119.

[49] S. Schütte, S. Scherfke, and M. Tröschel, “Mosaik: A framework for modular simulation of active components in smart grids,” in *2011 IEEE First International Workshop on Smart Grid Modeling and Simulation (SGMS)*, IEEE, 2011, pp. 55–60.

[50] S. Rotger-Griful, S. Chatzivasileiadis, R. H. Jacobsen, E. M. Stewart, J. M. Domingo, and M. Wetter, “Hardware-in-the-loop co-simulation of distribution grid for demand response,” in *2016 Power Systems Computation Conference (PSCC)*, IEEE, 2016, pp. 1–7.

[51] A. Nicolai and A. Paepcke, “Co-simulation between detailed building energy performance simulation and modelica hvac component models,” in *Proceedings of the 12th International Modelica Conference, Prague, Czech Republic, May 15-17, 2017*, Linköping University Electronic Press, 2017, pp. 63–72.

[52] K. Anderson, J. Du, A. Narayan, and A. El Gamal, “Gridspice: A distributed simulation platform for the smart grid,” *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2354–2363, 2014.

[53] R. Bottura, D. Babazadeh, K. Zhu, A. Borghetti, L. Nordström, and C. A. Nucci, “Sitl and hla co-simulation platforms: Tools for analysis of the integrated ict and electric power system,” in *Eurocon 2013*, IEEE, 2013, pp. 918–925.

[54] E. Widl, W. Müller, D. Basciotti, S. Henein, S. Hauer, and K. Eder, “Simulation of multi-domain energy systems based on the functional mock-up interface specification,” in *2015 International symposium on smart electric distribution systems and technologies (EDST)*, IEEE, 2015, pp. 510–515.

[55] M. Bacic, “On hardware-in-the-loop simulation,” in *Proceedings of the 44th IEEE Conference on Decision and Control*, IEEE, 2005, pp. 3194–3198.

[56] S. Balduin, “Surrogate models for composed simulation models in energy systems,” *Energy Informatics*, vol. 1, no. 1, pp. 403–410, 2018.

[57] F. Cremona, M. Lohstroh, D. Broman, E. A. Lee, M. Masin, and S. Tripakis, “Hybrid co-simulation: It’s about time,” *Software & Systems Modeling*, vol. 18, no. 3, pp. 1655–1679, 2019.

[58] *Simulation solutions for systems engineering and virtual testing - claytex*, https://www.claytex.com/, (Accessed on 03/31/2021).

[59] P. Fritzson, *Principles of Object-Oriented Modeling and Simulation with Modelica 3.3: A Cyber-Physical Approach*, 2nd ed. Wiley-IEEE Press, Apr. 2015, 1256 pp., ISBN: 978-1-118-85912-4.

[60] M. Sjölund, M. Gebremedhin, and P. Fritzson, “Parallelizing equation-based models for simulation on multi-core platforms by utilizing model structure,” in *17th International Workshop on Compilers for Parallel Computing (CPC 2013), Lyon, France, July 3-5, 2013*, 2013.

[61] C. Sun, Y. Jin, R. Cheng, J. Ding, and J. Zeng, “Surrogate-assisted cooperative swarm optimization of high-dimensional expensive problems,” *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 4, pp. 644–660, 2017.

[62] C. Safta, R. L.-Y. Chen, H. N. Najm, A. Pinar, and J.-P. Watson, “Toward using surrogates to accelerate solution of stochastic electricity grid operations problems,” in *2014 North American Power Symposium (NAPS)*, IEEE, 2014, pp. 1–6.

[63] R. Bornatico, J. Hüssy, A. Witzig, and L. Guzzella, “Surrogate modeling for the fast optimization of energy systems,” *Energy*, vol. 57, pp. 653–662, 2013.

[64] R. Pinto, R. J. Bessa, and M. A. Matos, “Multi-period flexibility forecast for low voltage prosumers,” *Energy*, vol. 141, pp. 2251–2263, 2017.

[65] A. Perera, P. Wickramasinghe, V. M. Nik, and J.-L. Scartezzini, “Machine learning methods to assist energy system optimization,” *Applied energy*, vol. 243, pp. 191–205, 2019.

[66] D. Gusain, J. L. Rueda, J. C. Boemer, and P. Palensky, “Identification of dynamic equivalents of active distribution networks through mvmo,” *IFAC-PapersOnLine*, vol. 49, no. 27, pp. 262–267, 2016.

[67] S. Balduin, M. Tröschel, and S. Lehnhoff, “Towards domain-specific surrogate models for smart grid co-simulation,” *Energy Informatics*, vol. 2, no. 1, pp. 1–19, 2019.

[68] S. Balduin, T. Westermann, and E. Puiutta, “Evaluating different machine learning techniques as surrogate for low voltage grids,” *Energy Informatics*, vol. 3, no. 1, pp. 1–12, 2020.

[69] S. Balduin, F. Oest, M. Blank-Babazadeh, A. Nieße, and S. Lehnhoff, “Tool-assisted surrogate selection for simulation models in energy systems,” in *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, 2019, pp. 185–192.

[70] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, “Efficient processing of deep neural networks: A tutorial and survey,” *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.

[71] M. Kezunovic, P. Pinson, Z. Obradovic, S. Grijalva, T. Hong, and R. Bessa, “Big data analytics for future electricity grids,” *Electric Power Systems Research*, vol. 189, p. 106 788, 2020.

[72] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, “Energy forecasting: A review and outlook,” *IEEE Open Access Journal of Power and Energy*, 2020.

[73] J. Brownlee, *Difference between classification and regression in machine learning*, https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/, (Accessed on 03/31/2021), 2017.

[74] P. Lison, “An introduction to machine learning,” *Language Technology Group (LTG)*, vol. 1, no. 35, 2015.

[75] R. Halder, R. Chatterjee, D. K. Sanyal, and P. K. Mallick, “Deep learning-based smart attendance monitoring system,” in *Proceedings of the Global AI Congress 2019*, Springer, 2020, pp. 101–115.

[76] A. Oppermann, *What is deep learning and how does it work?* https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac, (Accessed on 03/31/2021), 2019.

[77] J. Brownlee, *What is deep learning?* https://machinelearningmastery.com/what-is-deep-learning/, (Accessed on 03/31/2021), 2019.

[78] A. Bronshtein, *Train/test split and cross validation in python*, https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6, (Accessed on 03/31/2021), 2017.

[79] R. Sharma, A. V. Nori, and A. Aiken, "Bias-variance tradeoffs in program analysis," *ACM SIGPLAN Notices*, vol. 49, no. 1, pp. 127–137, 2014.

[80] M. Fedotenkova, "Extraction of multivariate components in brain signals obtained during general anesthesia," Ph.D. dissertation, Université de Lorraine, 2016.

[81] A. Mosavi, M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies*, vol. 12, no. 7, p. 1301, 2019.

[82] D. Yang, J.-e. Guo, J. Li, S. Wang, and S. Sun, "Knowledge mapping in electricity demand forecasting: A scientometric insight," *arXiv preprint arXiv:2003.10055*, 2020.

[83] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International journal of forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.

[84] N. Papakonstantinou, J. Savolainen, J. Koistinen, A. Aikala, and V. Vyatkin, "District heating temperature control algorithm based on short term weather forecast and consumption predictions," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, 2016, pp. 1–8.

[85] F. Dalipi, S. Yildirim Yayilgan, and A. Gebremedhin, "Data-driven machine-learning model in district heating system for heat load prediction: A comparison study," *Applied Computational Intelligence and Soft Computing*, vol. 2016, 2016.

[86] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 9, no. 2, e365, 2020.

[87] G. M. Yagli, D. Yang, and D. Srinivasan, "Automatic hourly solar forecasting using machine learning models," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 487–498, 2019.

[88] *Entso-e member companies*, https://www.entsoe.eu/about/inside-entsoe/members/, (Accessed on 03/31/2021).

[89] *About tennet - tennet*, https://www.tennet.eu/company/profile/about-tennet/, (Accessed on 03/31/2021).

[90] *Power distribution in europe - facts & figures*, https://www3.eurelectric.org/powerdi//stributionineurope/, (Accessed on 03/31/2021).

[91] *Future role of distribution system operators: Innovation landscape brief*, https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/Feb/IRENA_Landscape_Future_DSOs_2019.PDF?la=en&hash=EDEBEDD537DE4ED1D716F4342F2//D55D890EA5B9A&hash=EDEBEDD537DE4ED1D716F4342F2D55D890EA5B9A/, (Accessed on 03/31/2021), 2019.

[92] *E.dso*, https://www.edsoforsmartgrids.eu/, (Accessed on 03/31/2021).

[93] N. Hatziargyriou, *Microgrids: architectures and control*. John Wiley & Sons, 2014.

[94] H. Erhorn and H. Erhorn-Kluttig, "Selected examples of nearly zero-energy buildings," *Report of the Concerted Action EPBD*, 2014.

[95] *About the campaign – renovate europe*, https://www.renovate-europe.eu/about-the-campaign/, (Accessed on 03/31/2021).

[96] *Open power system data – a platform for open data of the european power system.* https://open-power-system-data.org/, (Accessed on 03/31/2021).

[97] *Electrical Engineering Handbook*. New Age International Publishers, 1998, ISBN: 9780852268858. [Online]. Available: https://books.google.nl/books?id=wxDjhgTabmQC.

[98] *Prakab: Home*, https://www.prakab.cz/, (Accessed on 03/31/2021).

[99] D. Fiems and W. Knottenbelt, *Analytical and Stochastic Modeling Techniques and Applications*. Springer, 2010.

[100] J. Heres, W. van Westering, G. van der Lubbe, and D. Janssen, "Stochastic effects of customer behaviour on bottom-up load estimations," *CIRED-Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2543–2547, 2017.

[101] M. Biech, T. Bigdon, C. Dielitz, G. Fromme, and A. Remke, "A smart neighbourhood simulation tool for shared energy storage and exchange," in *International Conference on Analytical and Stochastic Modeling Techniques and Applications*, Springer, 2016, pp. 76–91.

[102] *Github - ibpsa/modelica-ibpsa: Modelica library for building and district energy systems developed within ibpsa project 1*, https://github.com/ibpsa/modelica-ibpsa, (Accessed on 03/31/2021).

[103] *Github - ait-ies/disheatlib: Modelica library for district heating network modelling using ibpsa library as core*, https://github.com/AIT-IES/DisHeatLib, (Accessed on 03/31/2021).

[104] D. Müller, M. Lauster, A. Constantin, M. Fuchs, and P. Remmen, "Aixlib-an open-source modelica library within the iea-ebc annex 60 framework," *BauSIM*, vol. 20162016, pp. 3–9, 2016.

[105] H. Haes Alhelou, M. E. Hamedani-Golshan, T. C. Njenda, and P. Siano, "A survey on power system blackout and cascading events: Research motivations and challenges," *Energies*, vol. 12, no. 4, p. 682, 2019.

[106] M. Vesterlund, J. Sandberg, B. Lindblom, and J. Dahl, "Evaluation of losses in district heating system, a case study," in *International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems: 16/07/2013-19/07/2013*, 2013.

[107] J. Van Der Herten, T. Van Steenkiste, I. Couckuyt, and T. Dhaene, "Surrogate modelling with sequential design for expensive simulation applications," *Computer Simulation*, p. 173, 2017.

[108] S. Koziel, L. Leifsson, and X.-S. Yang, *Solving computationally expensive engineering problems: methods and applications*. Springer, 2014, vol. 97.

[109] W. M. Hartmann, "Computationally expensive methods in statistics: An introduction," in *International Workshop on Applied Parallel Computing*, Springer, 2004, pp. 928–930.

[110] P. Jiang, Q. Zhou, and X. Shao, "Surrogate-model-based design and optimization," in *Surrogate Model-Based Engineering Design and Optimization*, Springer, 2020, pp. 135–236.

[111] Y. S. Ong, P. B. Nair, and A. J. Keane, "Evolutionary optimization of computationally expensive problems via surrogate modeling," *AIAA journal*, vol. 41, no. 4, pp. 687–696, 2003.

[112] R. Christensen, *Analysis of variance, design, and regression: applied statistical methods*. CRC Press, 1996.

[113] W. H. Greene, "The econometric approach to efficiency analysis," *The measurement of productive efficiency and productivity growth*, vol. 1, no. 1, pp. 92–250, 2008.

[114] S. Hartshorn, "Linear regression and correlation: A beginner's guide," *United States: Amazon digital services LLC*, 2017.

[115] H. Deng and X. Song, "The theory and practice of linear regression," *World Transactions on Engineering and Technology Education*, vol. 11, no. 4, pp. 382–387, 2013.

[116] *Ridge and lasso regression: L1 and l2 regularization | by saptashwa bhattacharyya | towards data science*, https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b, (Accessed on 03/31/2021).

[117] H. Borchani, G. Varando, C. Bielza, and P. Larranaga, "A survey on multi-output regression," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, no. 5, pp. 216–233, 2015.

[118] K. C. Demirel, A. Sahin, and E. Albey, “Ensemble learning based on regressor chains: A case on quality prediction.,” in *DATA*, 2019, pp. 267–274.

[119] *Understanding support vector machine regression - matlab & simulink*, https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html, (Accessed on 03/31/2021).

[120] *Sklearn.svm.linearsvc — scikit-learn 0.24.1 documentation*, https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html, (Accessed on 03/31/2021).

[121] R. Choudhary and H. K. Gianey, “Comprehensive review on supervised machine learning algorithms,” in *2017 International Conference on Machine Learning and Data Science (MLDS)*, IEEE, 2017, pp. 37–43.

[122] J. Brownlee, *How to develop multi-output regression models with python*, https://machinele//arningmastery.com/multi-output-regression-models-with-python/, (Accessed on 03/31/2021).

[123] P. Geurts, A. Irrthum, and L. Wehenkel, “Supervised learning with decision tree-based methods in computational and systems biology,” *Molecular Biosystems*, vol. 5, no. 12, pp. 1593–1605, 2009.

[124] W.-Y. Loh, “Classification and regression trees,” *Wiley interdisciplinary reviews: data mining and knowledge discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[125] M.-H. Chiu, Y.-R. Yu, H. L. Liaw, and L. Chun-Hao, “The use of facial micro-expression state and tree-forest model for predicting conceptual-conflict based conceptual change,” *Chapter Title & Authors Page*, vol. 184, 2016.

[126] S. R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology,” *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[127] S. Fletcher and M. Z. Islam, “Decision tree classification with differential privacy: A survey,” *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–33, 2019.

[128] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, “Random forests and decision trees,” *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, p. 272, 2012.

[129] A. Verikas, E. Vaiciukynas, A. Gelzinis, J. Parker, and M. C. Olsson, “Electromyographic patterns during golf swing: Activation sequence profiling and prediction of shot effectiveness,” *Sensors*, vol. 16, no. 4, p. 592, 2016.

[130] M. R. Segal, “Machine learning benchmarks and random forest regression,” 2004.

[131] D. Bzdok, M. Krzywinski, and N. Altman, *Machine learning: Supervised methods*, 2018.

[132] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk, *A distribution-free theory of nonparametric regression*. Springer Science & Business Media, 2006.

[133] O. Theobald, *Machine learning for absolute beginners: a plain English introduction*. Scatterplot Press, 2017.

[134] L.-Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, “The distance function effect on k-nearest neighbor classification for medical datasets,” *SpringerPlus*, vol. 5, no. 1, pp. 1–9, 2016.

[135] N. M. Yusof, “Time series modeling and designing and artificial neural network (ann) for revenue forecasting,” Ph.D. dissertation, Universiti Teknologi Malaysia, 2006.

[136] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain.,” *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[137] P. Le and W. Zuidema, “Week 3: Perceptron and multi-layer perceptron,” 2013.

[138] R. C. Staudemeyer and E. R. Morris, “Understanding lstm–a tutorial into long short-term memory recurrent neural networks,” *arXiv preprint arXiv:1909.09586*, 2019.

[139] M. W. Gardner and S. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences,” *Atmospheric environment*, vol. 32, no. 14-15, pp. 2627–2636, 1998.

[140] S. Sheehan and Y. S. Song, “Deep learning for population genetic inference,” *PLoS computational biology*, vol. 12, no. 3, e1004845, 2016.

[141] X. Yuan, L. Li, and Y. Wang, “Nonlinear dynamic soft sensor modeling with supervised long short-term memory network,” *IEEE transactions on industrial informatics*, vol. 16, no. 5, pp. 3168–3176, 2019.

[142] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, “A dual-stage attention-based recurrent neural network for time series prediction,” *arXiv preprint arXiv:1704.02971*, 2017.

[143] X.-H. Le, H. V. Ho, G. Lee, and S. Jung, “Application of long short-term memory (lstm) neural network for flood forecasting,” *Water*, vol. 11, no. 7, p. 1387, 2019.

[144] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with lstm,” 1999.

[145] *How to build a compelling data science portfolio? | by srivatsan srinivasan | medium*, https://medium.com/@srivatsan88/how-to-build-a-compelling-data-science-portfolio-d92d3ca1644c, (Accessed on 03/31/2021).

[146] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, *et al.*, “Array programming with numpy,” *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[147] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, “Scipy 1.0: Fundamental algorithms for scientific computing in python,” *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[148] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn, and K. Smith, “Cython: The best of both worlds,” *Computing in Science & Engineering*, vol. 13, no. 2, pp. 31–39, 2010.

[149] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, *et al.*, “Theano: A python framework for fast computation of mathematical expressions,” *arXiv e-prints*, arXiv–1605, 2016.

[150] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, 2019.

[151] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, “Tensorflow: A system for large-scale machine learning,” in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.

[152] G. Luo, “A review of automatic selection methods for machine learning algorithms and hyper-parameter values,” *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 5, no. 1, pp. 1–16, 2016.

[153] J. Brownlee, *Hyperparameter optimization with random search and grid search*, https://machinelearni//ngmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/, (Accessed on 03/31/2021).

[154] ——, *How to tune lstm hyperparameters with keras for time series forecasting*, https://machinele//arningmastery.com/tune-lstm-hyperparameters-keras-time-series-foreca//sting/, (Accessed on 03/31/2021).

[155] P. Q. A. Guide, “Voltage disturbances,” *Standard EN*, vol. 50160, 2004.

[156] J. Bergstra, D. Yamins, D. D. Cox, *et al.*, “Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms,” in *Proceedings of the 12th Python in science conference*, Citeseer, vol. 13, 2013, p. 20.