UNIVERSIDADE D
COIMBRA

Daniel Cardoso Torres

Low-Frequency Unsupervised Non-Intrusive Load Monitoring For Industrial Loads

Dissertation within the scope of the Master in Electrical and Computer Engineering, specialization in Computers, advised by Professor Jérôme Mendes and co-advised by Professor Cristiano Premebida and presented to the Department of Electrical and Computer Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

September 2023

Faculty of Sciences and Technology
of the University of Coimbra

Low-Frequency Unsupervised Non-Intrusive Load Monitoring for Industrial Loads

Daniel Cardoso Torres

Master's dissertation in the scientific area of Electrical and Computer Engineering advised by Professor Jérôme Mendes and co-advised by Professor Cristiano Premebida and presented to the Department of Electrical and Computer Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

September 2023

Acknowledgements

I would like to express my appreciation to my academic advisor and co-advisor for their help in my educational journey. I am deeply grateful to my family, especially my mother and grandmother, for their selflessness and unwavering support. I want to thank my brother for being there for me. I am grateful to all my friends who were part of this journey, especially José Ramos and Cristiano Oliveira, for all the support over the last year. Lastly, I express my heartfelt gratitude to Beatriz Martinez. I am thankful for her presence in my life.

Abstract

The industrial sector is responsible for a large share of global energy consumption. Lowering energy consumption in the industrial sector can reduce the rate and severity of future climate change impacts on people and ecosystems. Non-Intrusive Load Monitoring (NILM) techniques can disaggregate a facility's power consumption into the individual loads, that is, into the power consumption of each equipment in the facility. NILM methods do not require the presence of a sensor per equipment. These methods provide information that can be used to define strategies for optimal energy usage in a facility and lead to a decrease in operating costs in the industrial sector. This dissertation aimed to develop a NILM algorithm to be part of an intelligent platform for the management of microalgae production within the scope of the InGestAlgae project (reference: CENTRO-01-0247-FEDER-046983) developed at the Institute of Systems and Robotics (ISR) of the University of Coimbra. The requirements defined the method to be an unsupervised and non-event-based method, compliant with low-frequency samples, and deployed in environments with continuously varying equipment. The developed technique is required to estimate the active power consumption of the equipment in an industrial facility. The method can access the values from the Supervisory Control and Data Acquisition (SCADA) system, which includes the aggregate and equipment's ON/OFF state data. Two unsupervised low-frequency NILM methods were proposed, implemented and validated. The first method uses polynomial functions, estimated through a metaheuristic algorithm, to model the active power consumption of the equipment as a function of the aggregate active power. The second technique consists of an Unsupervised Neural Network (UNN) that estimates the active power of the equipment based on the optimization of an objective function and does not require labelled training data. The UNN algorithm was trained and tested with two different architectures and sets of inputs. The first UNN uses the aggregate active power and equipment state samples, and the second network uses the aggregate active power samples passed through a Fourier feature mapping. The High-resolution Industrial Production Energy (HIPE) and the Industrial Machines Dataset for Electrical Load Disaggregation (IMDELD) datasets were preprocessed and used to train and test the proposed methods. The UNN, with the aggregate active power and the equipment state samples as input, estimated the results with the lowest error values, measured with different metrics such as the Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Squared Error (RMSE) for the testing data. The UNN method successfully identified high-consumption equipment.

Keywords: NILM, unsupervised, low-frequency, industrial loads, source separation, non-event-based, optimization, polynomial function, unsupervised neural network.

Resumo

O setor industrial é um dos principais consumidores de energia a nível global. A redução do consumo energético no setor industrial deverá conduzir à diminuição do aquecimento global e à mitigação do seu impacto em populações e ecossistemas. Técnicas de Non-Intrusive Load Monitoring (NILM) permitem desagregar o consumo elétrico de um agregado nos consumos individuais dos equipamentos do agregado e não requerem a presença de um sensor por equipamento. Estas técnicas fornecem informação que pode ser usada para definir estratégias que conduzam à redução do consumo elétrico e consequentemente à diminuição dos custos de operação no setor industrial. A dissertação tem como objetivo o desenvolvimento de um método que realize NILM e que possa integrar uma plataforma de gestão de produção de microalgas, no âmbito do projeto InGestAlgae (referência: CENTRO-01-0247-FEDER-04698) desenvolvido no Instituto de Sistemas e Robótica da Universidade de Coimbra. Os requisitos definem o método como não supervisionado, não baseado em eventos e compatível com amostras de baixa frequência e com equipamentos de tipo III. O método deve estimar o consumo de potência ativa dos equipamentos presentes numa fábrica. O algoritmo tem acesso às amostras de potência ativa do agregado e de estado dos equipamentos (ON/OFF). As amostras são fornecidos pelo sistema SCADA. Dois métodos que realizam NILM não supervisionada para amostra de baixa frequência foram desenvolvidos e validados. O primeiro método, modela o consumo de cada equipamento através de uma função polinomial, estimada a partir de um algoritmo de otimização meta-heurística, que tem como variável a potência ativa do agregado. O segundo método consiste numa rede neuronal não supervisionada que estima a potência ativa consumida por cada equipamento através da otimização de uma função objetivo e não requer dados de treino classificados. Duas redes neuronais não supervisionadas com arquiteturas distintas foram implementas. A primeira rede neuronal tem como entrada as amostras da potência ativa do agregado e o estado dos equipamentos. A segunda rede recebe como entrada um mapeamento de Fourier do agregado da potência ativa. Os métodos propostos foram treinados e validados com recurso a dois datasets públicos pré-processados, o HIPE e o IMDELD. A rede neuronal não supervisionada, que tem como entradas as amostras do agregado da potência ativa e o estado dos equipamentos estimou os resultados com menor erro, segundo as métricas de MAE, MSE e RMSE, para os dados de validação e identificou corretamente os equipamentos de maior consumo.
Palavras-chave: NILM, não supervisionada, baixa frequência, cargas indústriais, separação de fontes, não baseado em eventos, otimização, funções polinomias, rede neuronal não supervisionada.

Contents

List of Acronyms

ACOR Ant Colony Optimization for Continuous Domains.

AFAMAP Additive Factorial Approximate Maximum A Posteriori.

ANN Artificial Neural Networks.

CFHSMM Conditional Factorial Hidden Markov Method.

CI Critical Infrastructure.

CNN Convolutional Neural Network.

CUSUM Cumulative Sum.

CVD Continuously Variable Device.

DBSCAN Density-based Spatial Clustering of Applications with Noise.

DDSC Discriminative Disaggregation via Sparse Coding.

DTW Dynamic Time Warping.

EA Evolutionary Algorithm.

EMI Electromagnetic Interference.

EMUPF Equipment Modelling Using Polynomial Functions.

FHMM Factorial Hidden Markov Method.

FSM Finite State-Machines.

GA Genetic Algorithm.

GLR Generalized Likelihood Ratio.

GSP Graph Signal Processing.

HDP-HSMM Hierarchical Dirichlet Process Hidden Semi Markov Model.

HIPE High-resolution Industrial Production Energy.

HMM Hidden Markov Model.

ICS Industrial Control System.

IEA International Energy Agency.

IED Intelligent Electronic Device.

ILM Intrusive Load Monitoring.

IMDELD Industrial Machines Dataset for Electrical Load Disaggregation.

IPCC Intergovernmental Panel on Climate Change's.

ISIC International Standard Industrial Classification of All Economic Activities.

ISR Institute of Systems and Robotics.

kNN k-Nearest Neighbours.

LSTM Long Short-Term Memory.

LVDB Low-Voltage Distribution Board.

MAE Mean Absolute Error.

MF Matrix Factorization.

ML Machine Learning.

MSE Mean Square Error.

MV/LV Main Medium Voltage/Low Voltage Transformer.

NILM Non-Intrusive Load Monitoring.

NMF Nonnegative Matrix Factorization1.

NN Neural Network.

PLC Programmable Logic Controller.

PSO Particle Swarm Optimization.

ReLU Rectified Linear Unit.

RMSE Root Mean Squared Error.

RNN Recurrent Neural Network.

RTU Remote Telemetry Unit.

SA Simulated annealing.

SC Sparse Coding.

SCADA Supervisory Control and Data Acquisition.

SDS Sustainable Development Scenario.

SI Swarm Intelligence.

STMF Source Separation via Tensor and Matrix Factorization.

SVM Support Vector Machines.

UNN Unsupervised Neural Network.

VFD Variable Frequency Drive.

List of Figures

List of Tables

Chapter 1

Introduction

This chapter comprises five sections. A summary of the background that led to the development of the dissertation is presented. The problem, the significance of the work developed and the gaps in the state of the art are introduced. The goals and the requirements for the methods developed are outlined. The contributions made to the Non-Intrusive Load Monitoring (NILM) field are listed. The structure of the dissertation and an overview of each chapter are defined.

1.1 Context

In the industrial sector, it is essential to know the energy consumption of each equipment present in the industrial facility to identify high-energy consumers and subsequently act on them through strategies such as peak shaving or job scheduling. These strategies lead to a reduction in the energy demand of the facility and a decrease in the costs of operation and greenhouse gas emissions. NILM techniques estimate the electrical consumption of each equipment in an industrial facility that only houses energy meters at the aggregate level [1].

The Supervisory Control and Data Acquisition (SCADA) system [2] provides information, such as active power and equipment state data sampled by energy meters. SCADA information can be used as input for the NILM algorithm. A diagram that exemplifies the expected behaviour of a NILM algorithm is shown in Figure 1.1.

There are three commonly used techniques in the field of NILM. The first technique involves using neural networks, which require labelled data on the consumption of each equipment [3]. The second technique includes variations of Hidden Markov Model (HMM), which are not adequate to estimate the consumption of continuously varying equipment [4]. The third technique consists of solutions inspired by Hart's work [5] with the use of transient data [6].

Almost all of the studies discussed in the literature were conducted using data from household environments. A literature review shows a lack of low-frequency, unsupervised NILM algorithms for industrial loads. The InGestAlgae[1] project called for the development of a NILM method to provide an industrial microalgae production plant with estimates of the active power consumption data of the equipment and to help reduce its energy demand. New algorithms that provide a new approach must be developed to fill the gaps in the existing literature. The algorithms must be developed using different datasets and

[1] https://ingestalgae-p2020.eu

applied and evaluated in a real-world scenario.

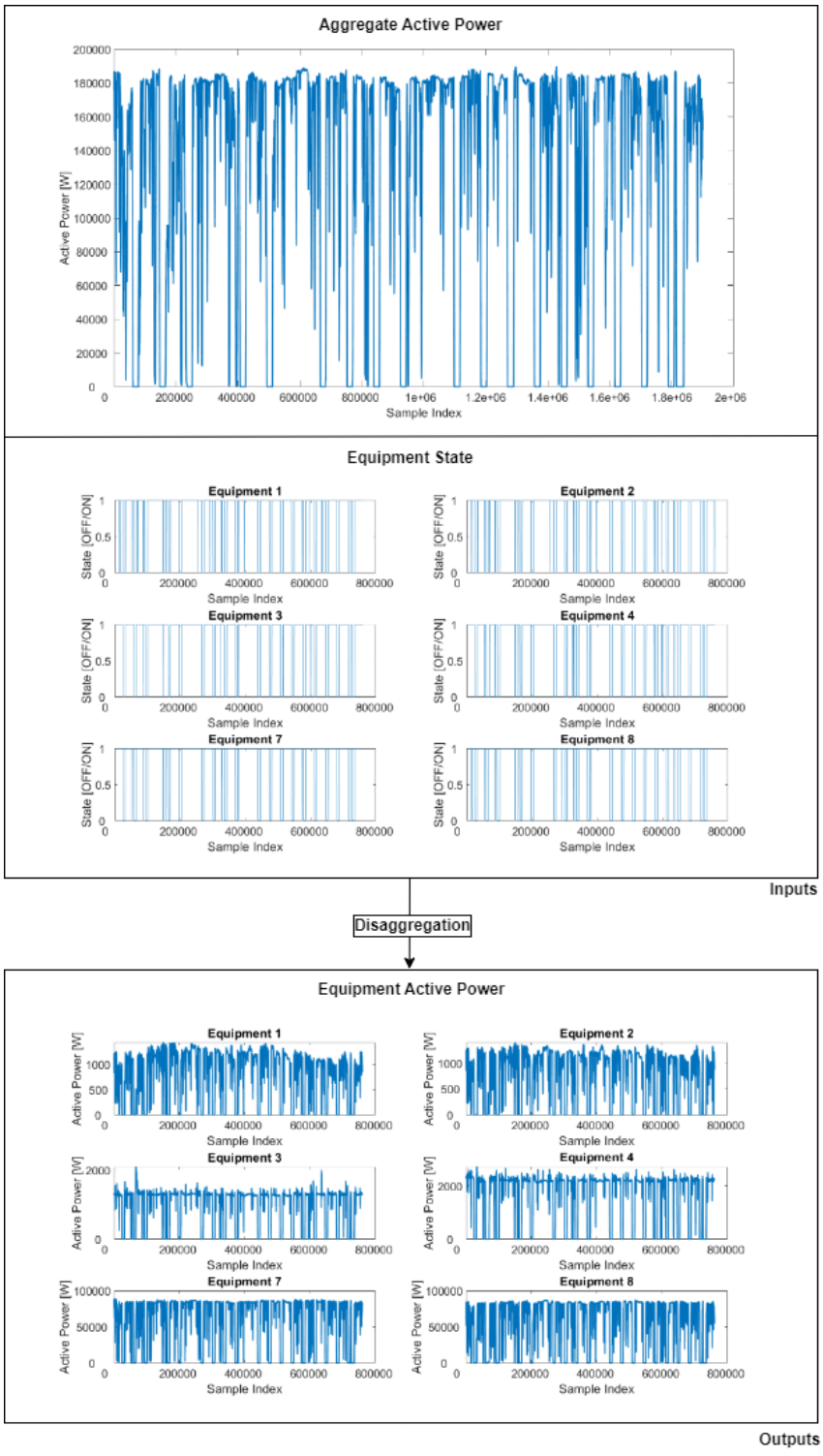

Figure 1.1: Diagram of the expected operation of the developed NILM algorithm, where the algorithm uses the aggregate active power and the ON/OFF state equipment data to estimate the equipment active power consumption.

1.2 Motivation

The severe impact of climate change on both natural and human systems has led to growing concern. Warming of the Earth's climate system negatively affects biodiversity, ecosystems, economic development, livelihoods, food and human security [7]. This is mainly due to rising temperatures, droughts, floods, famines, and economic disruption [8]. Excessive combustion of natural resources leads to high levels of pollutant gas emissions, which exacerbates climate change. Among the significant greenhouse gases, CO_2 is the main heat-trapping gas. Cumulative CO_2 emissions largely determine the mean surface temperature [7]. According to the Fifth Assessment Report, AR5, presented by the Intergovernmental Panel on Climate Change's (IPCC), since the beginning of the industrial revolution, the influence of humans on the climate system has grown [7]. This is due to the increase in greenhouse gas emissions caused by the growth of global and per capita energy consumption [9].

In 2021, the consumption of all fossil fuels increased to meet the growth of the electricity demand. CO_2 emissions of the electricity and heat production sectors increased by more than 900 metric tons, representing 46% of the global growth in CO_2 emissions. Greenhouse gas emissions from the energy sector reached the highest level ever in 2021 [10].

As shown in Figure 1.2, there has been a near-simultaneous increase in CO_2 emissions and electricity generation from 1990 to 2018. Figure 1.3 also shows an increase in worldwide CO_2 emissions originating from energy combustion and industrial processes between 1900 and 2021.

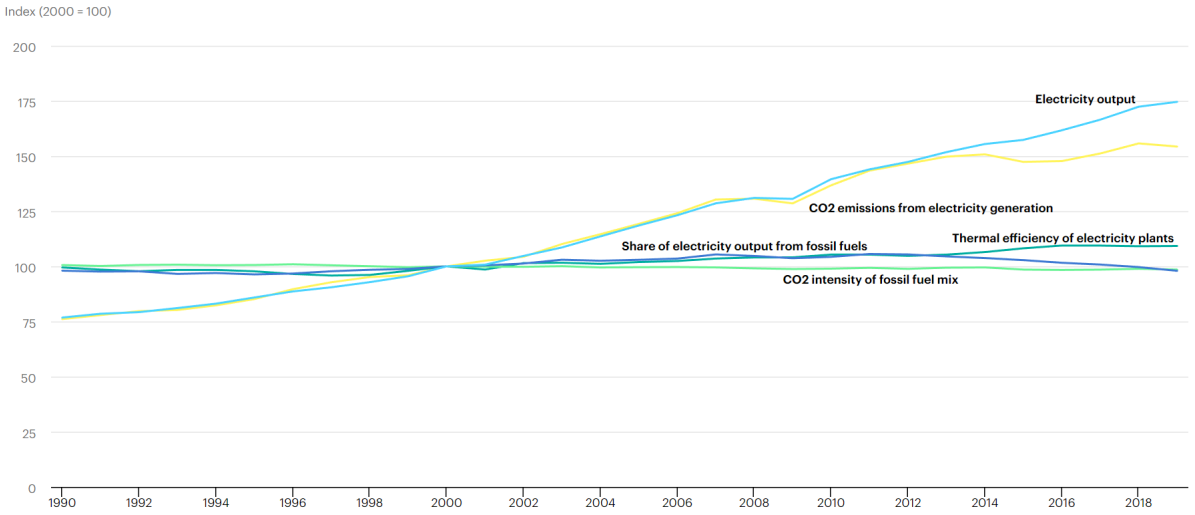

Figure 1.2: Global CO_2 emissions from electricity generation factors, between 1990 and 2019, by the IEA Energy and Carbon Tracker 2020 [11].

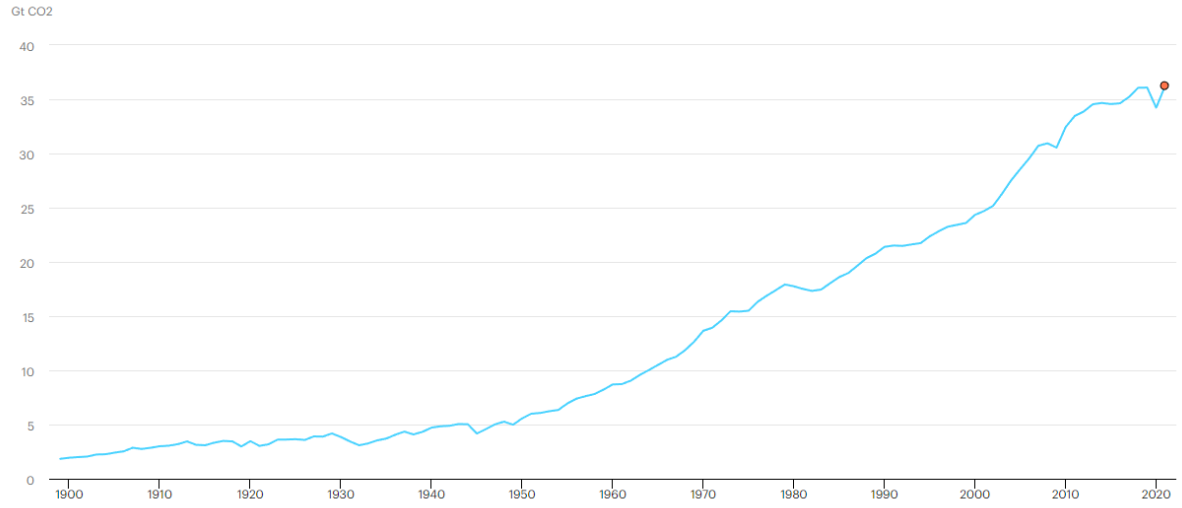

Figure 1.3: Global CO_2 emissions from energy combustion and industrial processes, between 1900 and 2021 [10].

According to the International Energy Agency (IEA) 2021 statistics report [12], in 2019, the industrial sector was the most prominent electricity consumer sector in the world, accounting for 41.9% of the 82 exajoules of electricity consumed, as shown in Figure 1.4.

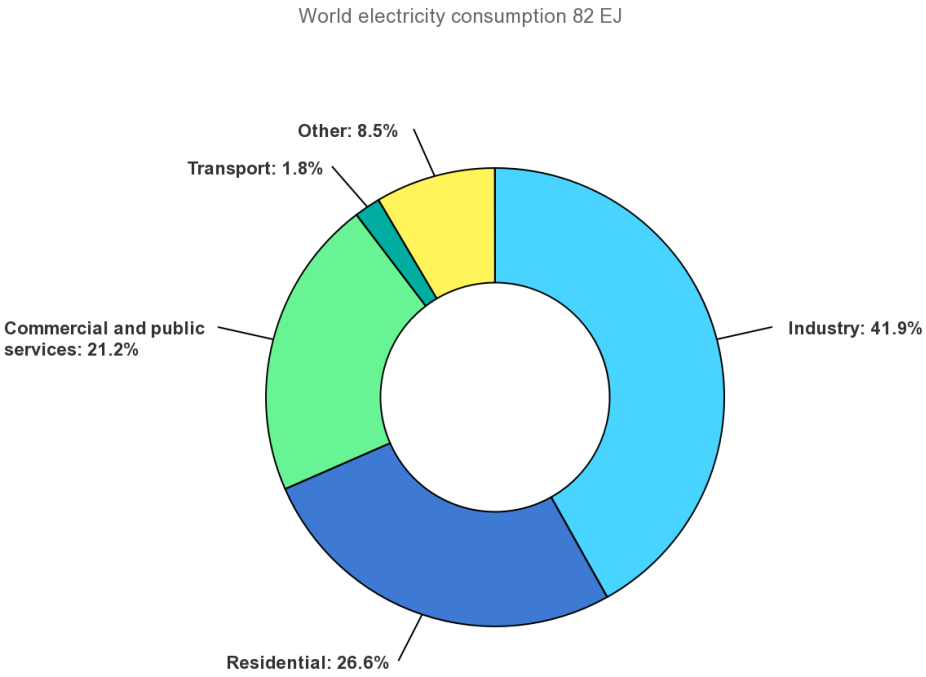

Figure 1.4: 2019 global electricity consumption broken down by sector [12].

It is crucial to implement global-scale strategic actions to manage climate change. Cost-effective measures must be taken to reduce the intensity of net emissions in the

end-use sectors [7]. The expected solutions to reduce global carbon emissions, as described in the Sustainable Development Scenario (SDS) [13] and in the IEA Global Energy Review 2021 report [10], include the spread of renewable energy sources, the reduction of energy demand and the improvement of energy efficiency. Energy efficiency improvements can lead to more than 224 different "non-energy" industrial productivity benefits, including increased profit, safer working conditions and improvement in quality and output [14, 15].

The industrial sector can improve energy efficiency through management, technology, or policy/regulation approaches. Energy management involves strategizing to meet energy demand when and where needed, adjusting and optimizing energy usage [8]. Adopting energy-efficient behaviours through energy management strategies requires a thorough understanding of the electrical consumption of each equipment in a facility. In an industrial setting, only the electrical load data at the aggregate level is available unless expensive and specialized hardware has been installed per equipment. Buggypower's micro-algae production plant, shown in Figure 1.5, does not have individual meters for each equipment.

The total electrical load of Buggypower's plant can be used to estimate the individual equipment loads through a computational technique called energy disaggregation or NILM. NILM techniques performs energy disaggregation and provides feedback that indicates high consumption sources. NILM methods enables subsequent action on sources of high consumption, such as peak shaving or job scheduling [8].

Research has shown that active energy feedback to residential consumers can reduce electricity consumption in homes by 5-20% [16]. The energy savings potential of active energy feedback in industrial facilities has not been studied. Given these considerations, developing a NILM method for the industrial sector and studying its application in real-world scenarios is essential.

Figure 1.5: Photography taken at Buggypower's micro-algae production plant in Porto Santo, Madeira, Portugal [17].

1.3 Objectives

The dissertation aims to develop a NILM method that can estimate the energy consumption of each equipment in an industrial facility using the information provided by the SCADA system. The final objective is to integrate the algorithm into an industrial factory setting to improve energy management and optimize energy usage in microalgae production.

A set of requirements/constraints were defined for the algorithm to meet:

1. Learning should be unsupervised, with no equipment information, except ON/OFF state data;
2. The method should work with low-frequency samples;
3. The algorithm has to perform source separation of multiple equipment;
4. The technique is expected to disaggregate equipment of all types;
5. The user should be able to visualize the results.

The NILM algorithm had the following set of non-functional requirements:

1. Accessibility: The user should be able to access and understand the results easily;
2. Scalability: The algorithm has to scale to accommodate a wide range of scenarios and equipment;
3. Performance: The online stage of the algorithm should have a short response time, suitable for real-time systems;
4. Usability: Users are allowed to derive value from the algorithm.

To achieve this goal, the following objectives must be completed:

1. Analyzing the state of the art in the field of NILM and unsupervised NILM;
2. Surveying and selecting a public NILM dataset;
3. Modelling the problem mathematically;
4. Developing the algorithm;
5. Writing software;
6. Evaluating the results using the selected dataset;
7. Concluding and defining future work.

1.4 Contributions

Applying a low-frequency unsupervised NILM algorithm to the industrial sector is a largely unexplored subject. The state of the art in this area is limited, as most NILM algorithms are supervised, requiring disaggregated training data, or do not perform source separation of continuously varying equipment, or are only applied to domestic environments.

The work is significant because it fills a clear gap in the state of the art. The main contributions and developed work of the dissertation are as follows:

1. A survey of public NILM datasets;

2. A review of the state of the art on unsupervised NILM;
3. An analysis and preprocessing of the HIPE and the IMDELD datasets, developed with MATLAB;
4. The development and testing, using the preprocessed HIPE and IMDELD datasets, of two NILM methods:
   (a) A novel method for the modelling of industrial loads by polynomial functions, with metaheuristic optimization algorithms, developed in C++;
   (b) An unsupervised neural network using Python.

In the literature, no prior method that uses polynomial functions to model the power consumption of the equipment, as a function of the aggregate active power, in an industrial setting, has been found. The proposed method was the first NILM solution to formulate the objective function using matrices to find the coefficients of polynomial functions. No previous study has been found that applies an UNN to solve a NILM solution. The first network architecture and objective function were devised to tackle the NILM problem.

1.5 Dissertation Outline

The dissertation is organized into six chapters:

- Chapter 1 - Introduction: The current chapter introduced the NILM theme, and the work developed and provided motivation and goals for developing and implementing a low-frequency unsupervised NILM method applied to an industrial setting;
- Chapter 2 - Background: An introductory overview of the theoretical foundations is established;
- Chapter 3 - State of the Art: The State of the Art in NILM techniques is presented;
- Chapter 4 - Methodology: The developed work is described, and the implemented NILM methods are detailed;
- Chapter 5 - Results: The results and performance metrics of the proposed NILM methods for the selected industrial datasets are presented;
- Chapter 6 - Discussion and Conclusion: A discussion of the results, a summary, and final remarks are provided. Steps are mentioned to develop further and enhance the method.

Chapter 2

Background

This chapter presents an introductory overview of the key concepts and foundations of the NILM problem and the basic concepts related to the algorithms developed.

2.1 NILM Concepts

2.1.1 Energy Disaggregation

Energy consumption in a facility can be identified and monitored using Intrusive Load Monitoring (ILM) or NILM. ILM requires the installation of individual load meters for each equipment, which is expensive due to the cost of the required hardware, labour and communication infrastructure. On the other hand, NILM techniques use a single meter to monitor the total power consumption of an aggregate of equipment with algorithms to estimate the power consumption of individual loads within the facility [18]. The NILM or energy disaggregation problem can be mathematically formulated as shown by Equation (2.1) [19]:

$$a_t = \sum_{i=1}^{n} p_{ti} + e_t \tag{2.1}$$

i is the equipment index. n is the total number of equipment contributing to the aggregate active power at instant t. p_{ti} corresponds to the active power consumption of the equipment i. e_t is noise or error. a_t is the aggregate active power consumption measured on the meter. The objective of a NILM technique is to estimate p_{ti} from the a_t values. An initial interpretation of the mathematical formulation may suggest that the NILM problem can be solved using combinatorial optimization, which is unfeasible when considering a large number and the different types of equipment [5] and type III equipment changes the problem's domain from discrete to continuous. NILM presents lower costs than ILM but inherently introduces uncertainty in the estimated consumption values. The uncertainty arises from various factors, including noise in the measurements, the complexity of the load signatures of the equipment and the possible presence of multiple and different types of equipment. There are four types of equipment, classified according to their power consumption [20]:

- Type I - ON/OFF equipment: Equipment with only two possible states (ON/OFF);
- Type II - FSM: Equipment's power consumption passes through state transitions;

- Type III - Continuously varying equipment: Equipment where the power consumption values can vary through time in a continuous domain;
- Type IV - Permanent consumer equipment: Equipment with only one state.

Type III, often called Variable Frequency Drive (VFD) or Continuously Variable Device (CVD), is the most challenging type of equipment to disaggregate and is ubiquitous in the industrial sector. Examples include drilling and milling machines, whose power demands vary based on the engine speed [21].

The formulation and estimations of NILM techniques also rely on the sampling rate at which the data is collected. Data acquisition systems can be low-frequency (less than 1Hz) or high-frequency (kHz to MHz). Low-frequency energy acquisition meters are cheaper than high-frequency but do not provide data with as much detail. Low-frequency meters only provide information on steady-state data, and high-frequency energy meters can measure transients and electrical noise [20, 22, 23, 24].

NILM methods can be event-based or non-event-based. Event-based algorithms depend on events. An event corresponds to a significant variation in the aggregate electrical signal and suggests a change in the state of one equipment. An event can provide useful information and is commonly used in the literature by solutions that disaggregate the aggregate active power composed of type I and II equipment. Non-event-based algorithms perform disaggregation at every instant without relying on event detection and are suitable for disaggregating equipment of type III.

NILM algorithms can also be supervised or unsupervised. Supervised NILM methods use a priori knowledge of equipment consumption data, such as labelled consumption data or signature loads, while unsupervised algorithms do not have access to equipment data [20].

NILM techniques can be divided into load classification and source separation. The load classification process identifies the state of the power consumption of each equipment. Source separation estimates the power consumption of the equipment. Most of the NILM literature implements algorithms that follow the same four main steps [23]:

1. Data acquisition and signal preprocessing: In this stage, electrical data is collected and power normalization, filtering and thresholding may take place;
2. Event/edge detection: Events are identified, corresponding to the change in the state of equipment, implied by changes in the aggregate data;
3. Feature extraction: Features that identify the equipment are extracted within the event windows;
4. Learning/inference or classification/load identification: A supervised or unsupervised approach is performed to identify each equipment's power consumption or state based on the extracted features.

In the literature, different algorithms are considered to perform NILM. Still, their expected outcomes differ, resulting in diverse implementations for the last step of the traditional NILM method.

The established requirements prevent the adoption of a traditional approach.

2.2 General Concepts

2.2.1 SCADA Systems

The SCADA system is a complex type of Industrial Control System (ICS) whose purpose is to control and monitor geographically distributed assets, widely used to control industrial processes and Critical Infrastructure (CI) [25]. A SCADA system comprises hardware and software components and a connecting network. The SCADA system is formed by one or more control centres connected by a communication infrastructure to several field physical devices through Intelligent Electronic Devices (IEDs), Programmable Logic Controllers (PLCs) or Remote Telemetry Units (RTUs). PLCs and RTUs acquire data by being connected to physical devices such as sensors and actuators.

The main functionalities of a SCADA system are data logging, performed cyclically or in response to events, alarm handling, and automation. A complex sequence of actions can be automatically executed or triggered by events [26]. Figure 2.1 shows a generic SCADA architecture.

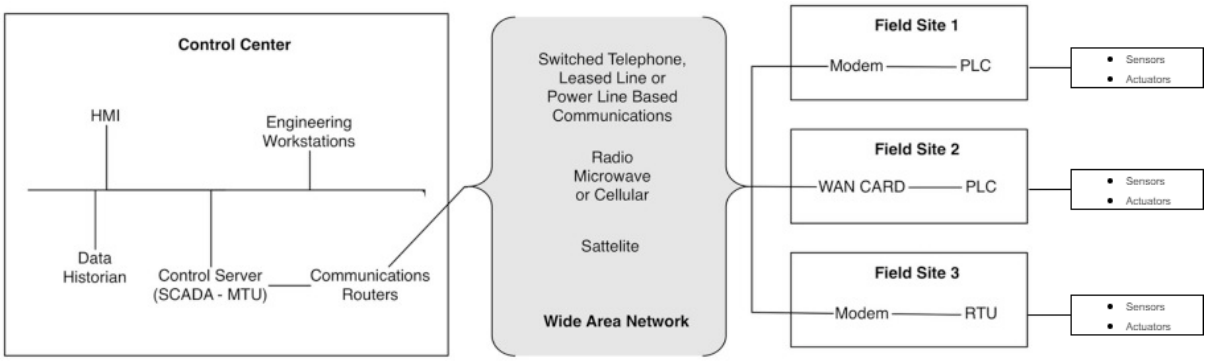

Figure 2.1: Generic SCADA hardware architecture [27].

In the context of energy disaggregation, SCADA systems can provide valuable information. The developed NILM algorithm has access to a unique set of inputs, resulting from integrating the process data from the SCADA system. The system provides information acquired at the energy meters, such as active power and the state of operation of the equipment connected to the meter in the facility.

2.3 Review of NILM Dataset

2.3.1 NILM Datasets

A power disaggregation dataset is required to develop and validate a NILM technique. Multiple datasets differ in various attributes, such as sampling frequency, number and type of equipment, measured units, and environment [28]. Table 2.1 synthesizes the conducted survey of public NILM databases. A dataset was required with active power samples acquired at a low sampling frequency, at the aggregate and equipment level, with samples for multiple equipment in an industrial facility environment. The HIPE[1] and IMDELD[2]

[1]https://www.energystatusdata.kit.edu/hipe.php
[2]https://ieee-dataport.org/open-access/industrial-machines-dataset-electrical-load-disaggregation

datasets were selected as they were the only ones that met the requirements. The selected datasets required preprocessing.

Table 2.1: Survey of public NILM datasets. "agg" stands for aggregate and "eq" for equipment.

Number	Dataset	Year	Citations (Google Scholar, Jan 2023)	Enviroment	Frequency
1	ACS-F1 [29]	2013	61	Household	0.1Hz
2	ACS-F2 [30]	2014	53	Household	0.1Hz
3	AMBAL [31]	2017	29	Household - Synthetic	1Hz
4	AMPds / AMPds2 [32]	2013	217	Household	1Hz / 0.0167Hz
5	BERDS [33]	2013	33	Commercial	0.05Hz
6	BLOND [34]	2018	75	Commercial	BLOND-50: 50kHz agg and 6.4kHz eq BLOND-250: 250kHz agg, 50kHz eq
7	BLUED [35]	2012	398	Household	1Hz current and 60Hz active power
8	COMBED [36]	2014	113	Commercial	2Hz
9	COOLL [37]	2016	87	Laboratory	100kHz
10	CU-BEMS [38]	2020	25	Commercial	0.0167Hz and 1Hz
11	Dataport [39]	2012	54	Household	16.67mHz to 1Hz
12	DRED [40]	2015	121	Household	1Hz
13	ECO [41]	2014	335	Household	1Hz
14	EEUD [42]	2017	38	Household	0.0167Hz
15	ENERTALK [43]	2019	40	Household	15Hz
16	ESHL [44]	2016	2	Household	0.5 to 1Hz
17	GREEND [45]	2014	193	Household	1Hz
18	HELD1 [46]	2018	15	Laboratory	4kHz
19	HFED [47]	2014	66	Household + Laboratory	9kHz to 30MHz
20	HIPE [48]	2018	25	Industry	0.2Hz
21	HES [49]	2012	207	Household	8.33mHz
22	HUE [50]	2019	24	Household	1Hz
23	iAWE [51]	2013	186	Household	1Hz
24	IDEAL [52]	2021	14	Household	1Hz
25	IHEPCDS [53]	2013	12	Household	0.016Hz
26	IMDELD [54]	2020	11	Industry	1Hz
27	I-BLEND [55]	2019	34	Commercial	0.0167Hz
28	LIFTED [56]	2020	12	Household	50Hz
29	LILAC [57]	2019	13	Industrial	50Hz
30	OPLD [58]	2016	3	Commercial	1Hz
31	PLAID I [59]	2014	210	Household	30kHz
32	PlaID II [60]	2017	14	Household	30kHz
33	PlaID III [61]	2020	32	Household	30kHz
34	RAE [62]	2018	63	Household	1Hz
35	RBSA [63]	2014	12	Household	0.0011Hz
36	REDD [64]	2011	1527	Household	15kHz, 0.5Hz and 1Hz
37	REFIT [65]	2017	260	Household	0.0167Hz
38	Sample	2012	54	Household	0.0167Hz
39	SHED [66]	2018	34	Commercial - Synthetic	0.033Hz
40	Smart / Smart* [67]	2017	519	Household	1Hz
41	SmartSim [68]	2016	24	Household - Synthetic	1Hz
42	South Korean factories dataset [69]	2022	1	Industry	0.0167Hz
43	SustData [70]	2014	67	Household	50Hz
44	SustDataED [71]	2016	27	Household	12.8kHz agg and 0.5Hz eq
45	SynD [72]	2020	51	Household	5Hz
46	SPAFID [73]	2021	1	Industry - Synthetic	0.0003Hz
47	Tracebase [74]	2012	303	Household	1Hz
48	UK-DALE [75]	2014	741	Household	16kHz agg and 0.17Hz eq
49	WHITED [76]	2016	123	Household + Industry	44.1kHz

2.3.2 HIPE Dataset

The HIPE dataset [48] contains data on multiple electric quantities, including voltage, current, active, reactive and apparent power, and total harmonic distortion. The samples cover the period from October 23, 2017, to December 1, 2018. The data was sampled at a frequency of 0.2Hz for both the main terminal and the ten equipment listed in Table 2.2. A representation of the electrical installation of the facility is shown in Figure 2.2. The equipment is part of an electronics production plant operated by the Institute of Data Processing and Electronics of Karlsruhe Institute of Technology, in Germany. The plant

produces electronic systems for particle physics, battery systems, and medical applications in batches of less than 1,000 pieces.

Table 2.2: Equipment included in the HIPE dataset.

Equipment Index	Equipment Name	Name Abbreviation
1	Chip press	CP
2	Chip saw	CS
3	High temperature oven	HTO
4	Pick and place unit	PPU
5	Screen printer	SP
6	Soldering oven	SO
7	Vaccum oven	VO
8	Vaccum pump 1	VP1
9	Vaccum pump 2	VP2
10	Washing machine	WM

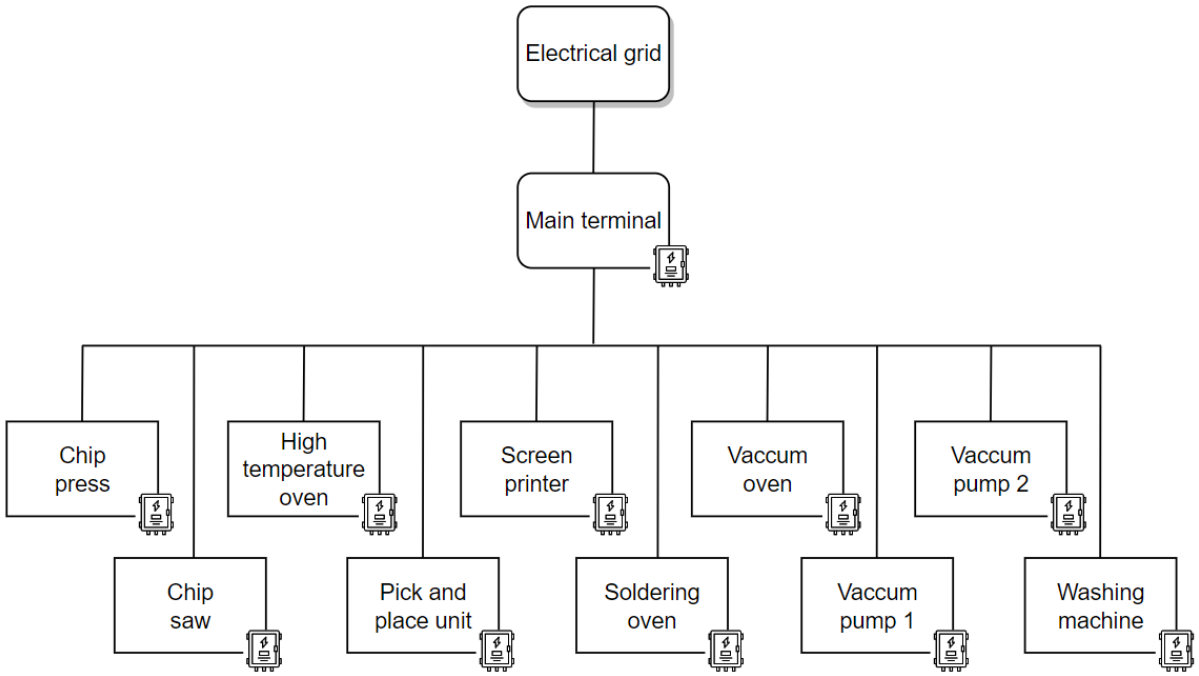

Figure 2.2: Diagram of the factory electrical installation for the HIPE dataset. The rectangles represent the equipment, and the meter illustrations show the locations where the data was sampled.

A one-week period, from October 23, 2017, to October 30, 2017, from the original dataset was used. The aggregate active power, in kW, measured at the main terminal, is shown in Figure 2.3. The active power, in kW, for each equipment during the one-week period is shown in Figure 2.4 and 2.5. Figure 2.4 shows that equipment with indices three and six have the highest active power consumption values and equipment with index one is always in the OFF state. Figure 2.6 displays the histogram of the equipment active power samples and suggests that equipment with indices two, six, eight and nine are of type III and equipment with indices four, five, seven and ten are of type II equipment.

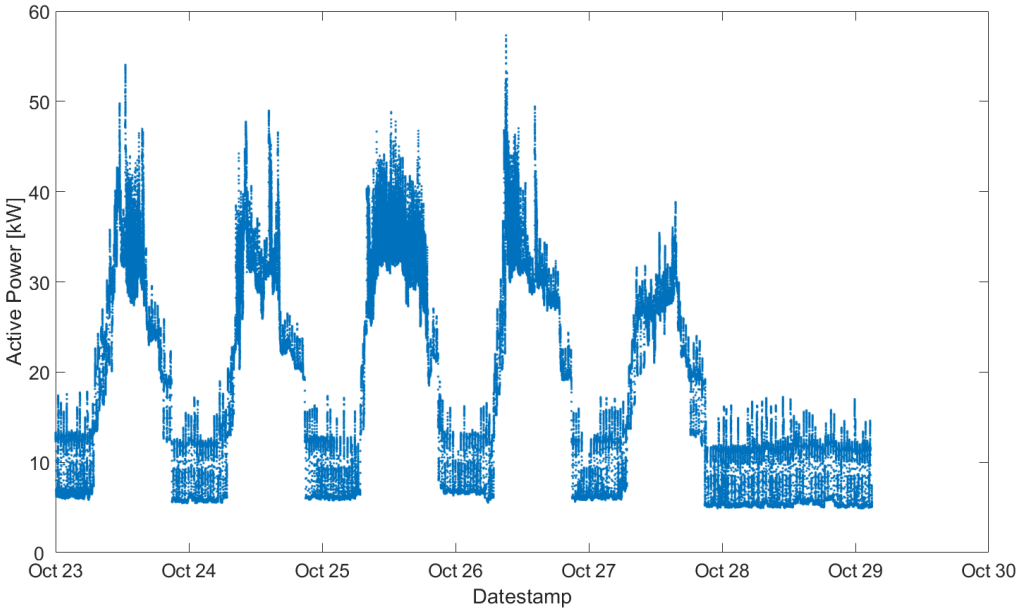

Figure 2.3: HIPE dataset's main terminal's active power, in kW, for a one-week period.

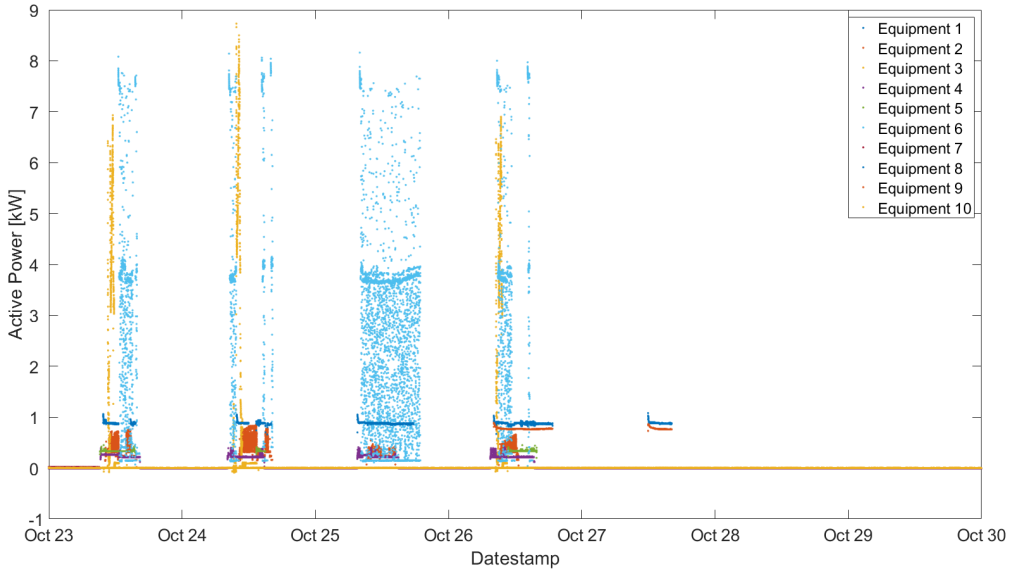

Figure 2.4: Active power, in kW, for the equipment in the HIPE dataset, in a single plot, for a one-week period.

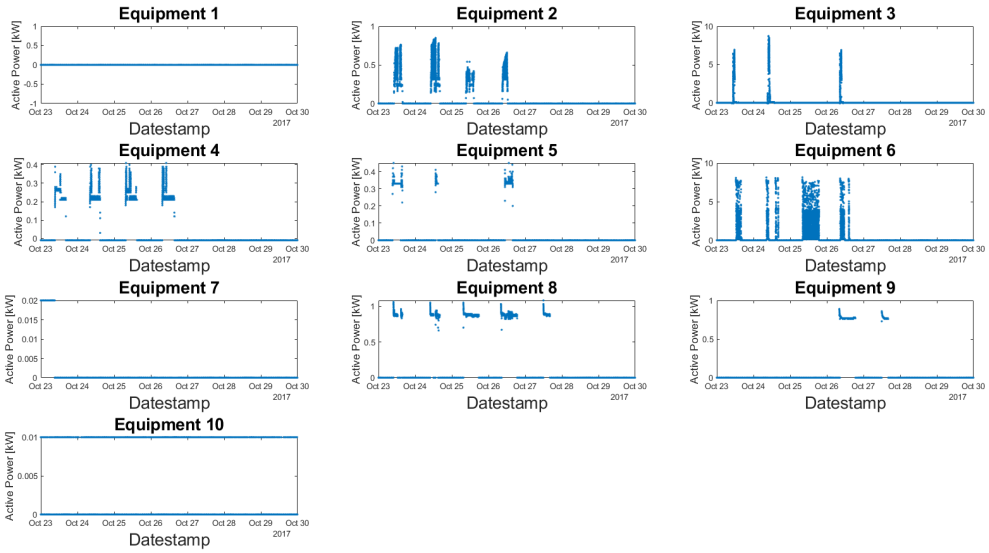

Figure 2.5: Active power, in kW, for the equipment in the HIPE dataset, divided into multiple plots, for a one-week period.

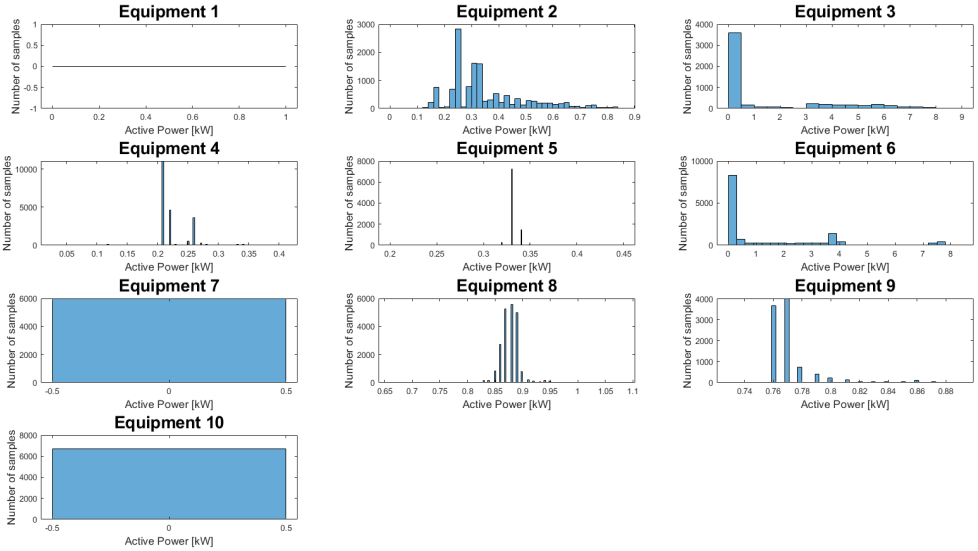

Figure 2.6: Histogram of the active power samples bigger than zero, in kW, for the equipment in the HIPE dataset, divided into multiple plots, for a one-week period.

2.3.3 IMDELD Dataset

The IMDELD dataset, described in the IEEEDataPort [54], contains downsampled low-frequency samples (1Hz) of RMS current and voltage, active, reactive and apparent power readings from a factory located in Minas Gerais, Brazil. The factory produces corn and soybean pellets for poultry from Monday to Friday and occasionally on Saturdays, throughout the day, except from 5:00 PM to 10:00 PM. The samples were collected for 111

days, from December 11, 2017, 18:43:52 UTC to April 1, 2018, 21:33:17 UTC. The milling machines were only sampled for twelve days. Eleven GreenAnt meters were installed, one for each equipment in Table 2.3, one per Low-Voltage Distribution Board (LVDB) and one for the Main Medium Voltage/Low Voltage Transformer (MV/LV). A diagram of the factory electrical substation is shown in Figure 2.7.

Table 2.3: Equipment present on the IMDELD dataset.

Equipment Index	Equipment Name	Name Abbreviation
1	Double-pole Contactor I	DPCI
2	Double-pole Contactor II	DPCII
3	Exhaust Fan I	EFI
4	Exhaust Fan II	EFII
5	Milling Machine I	MI
6	Milling Machine II	MII
7	Pelletizer I	PI
8	Pelletizer II	PII

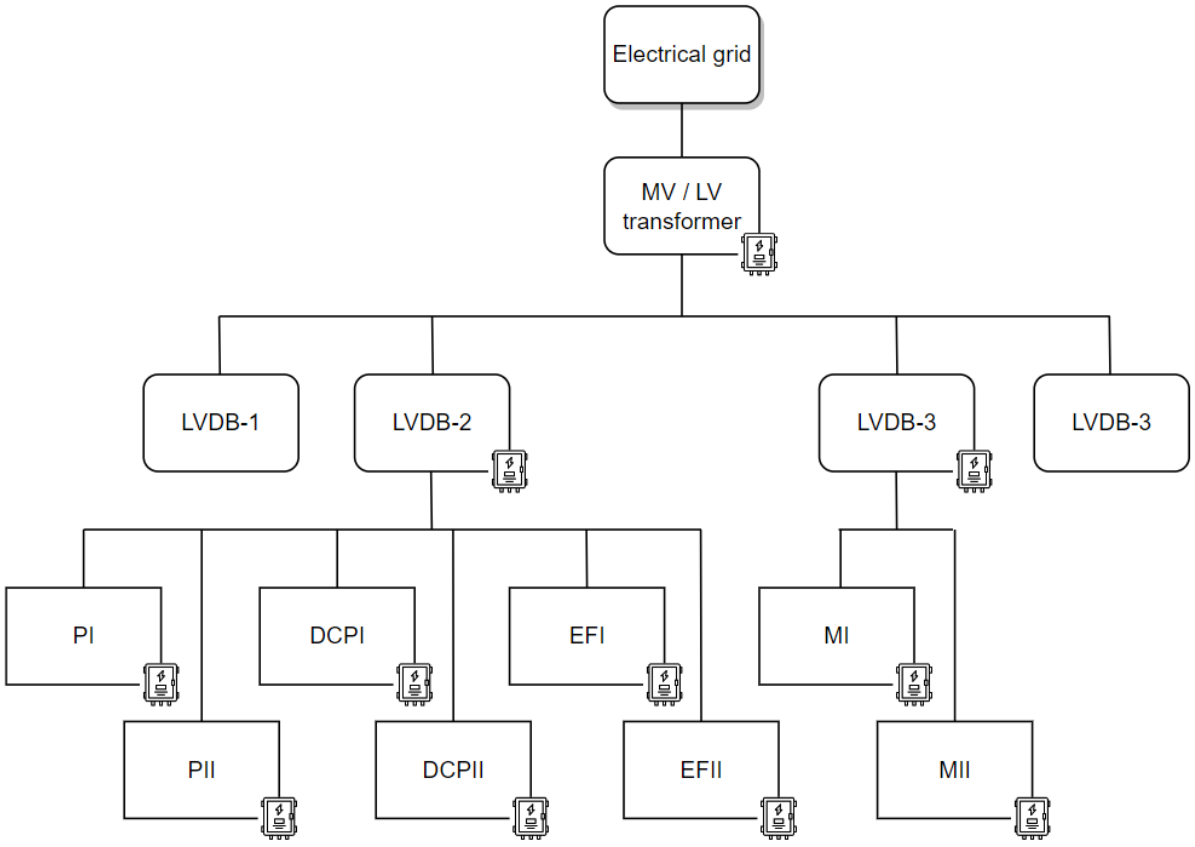

Figure 2.7: Diagram of the factory electrical substation for the IMDELD dataset. The rectangles represent the equipment, and the meter illustrations show the locations where the data was sampled.

The LVDB-2 data was selected over LVDB-3 because it includes a larger number of equipment. LVDB-3 and MI and MII equipment data were discarded. The aggregate active power measurements on the LVDB-2 meter of the IMDELD dataset are shown in

Figure 2.8. Before preprocessing, active power samples from the equipment are shown in Figures 2.9 and 2.10. Figure 2.10 shows that equipment with indices seven and eight has the highest active power consumption values, and Figure 2.8 indicates that these two equipment have the largest influence on the values of the aggregate active power. Figure 2.11 displays the histogram of the equipment active power samples bigger than zero and suggests that the dataset is composed of type III equipment.

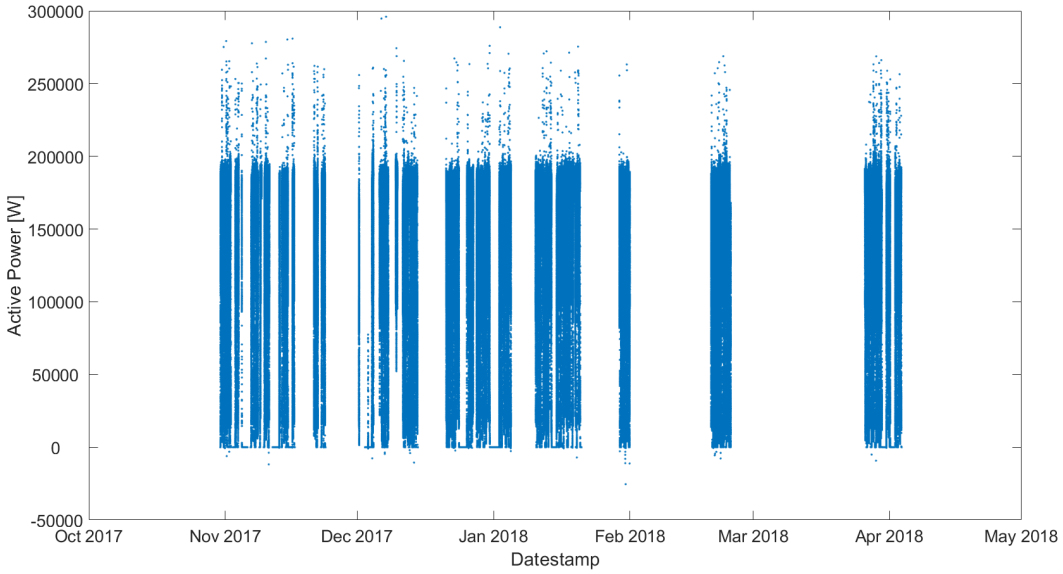

Figure 2.8: Active power, in W, for the aggregate data, measured at the LVDB-2, in the IMDELD dataset.

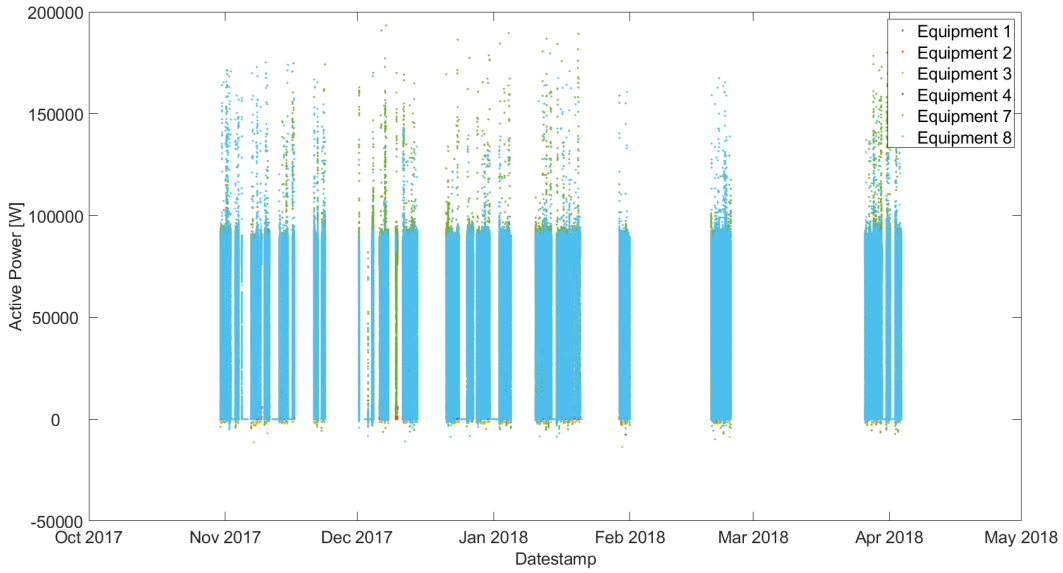

Figure 2.9: Active power, in W, for the equipment in the IMDELD dataset, in a single plot.

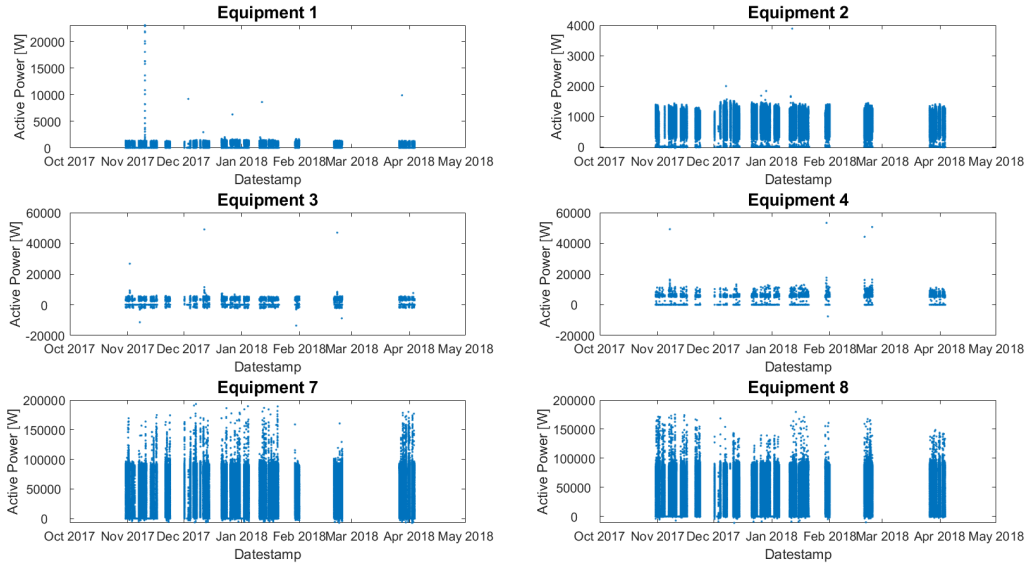

Figure 2.10: Active power, in W, for the equipment in the IMDELD dataset, divided into multiple plots.

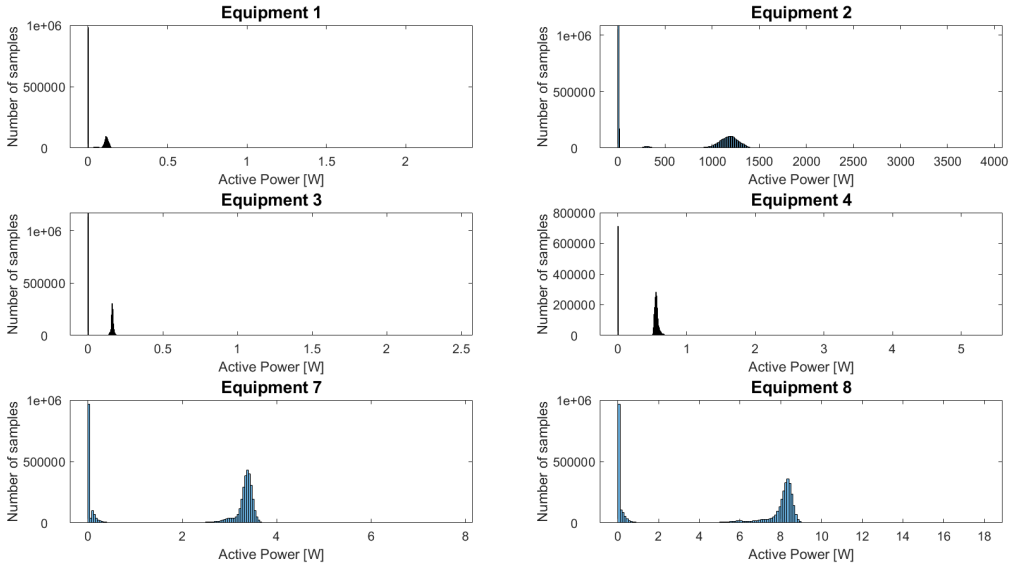

Figure 2.11: Histogram of the active power samples bigger than zero, in W, for the equipment in the IMDELD dataset, divided into multiple plots.

2.4 Mathematical and Computational Concepts

2.4.1 Polynomial Functions

Polynomial functions are compounded by one term or the sum of multiple terms. Each term comprises the product between a constant coefficient and a variable raised to a

non-negative integer exponent [77]. Polynomial functions follow the form:

$$f(x) = \sum_{i=0}^{n} c_i \times x^i \tag{2.2}$$

where i is the index of the term, n is the degree of the polynomial function, c_i corresponds to the coefficient, and x is the variable of the function.

Polynomial functions are useful for approximating complex shapes and are commonly used in curve-fitting problems [77]. In the literature, polynomial functions have been used to model aggregate-level energy consumption as a function of relevant variables [78, 79]. No previous studies have been found in the literature that use polynomial functions to model the equipment's active power consumption, using aggregate active power as the variable.

2.4.2 Optimization

An optimization problem involves finding a given function's maximum or minimum value. Numerical and metaheuristic methods are two approaches to solving optimization problems in continuous domains. Numerical optimization techniques rely on the function's gradient to iteratively approximate the minimum or maximum value. Examples of numerical optimization methods are gradient descent and Newton's method [80].

However, for the cases where the function is a complex search space with multiple local minima or maxima, or for non-differentiable functions, with various saddle points, numerical methods are not suitable. Metaheuristic optimization methods can provide a solution for such cases. Metaheuristic algorithms are computational intelligence techniques that combine two search schemes: exploration and exploitation [81]. The exploitation scheme searches for the best solution within a given search space, and the exploration scheme explores new solution spaces. Although metaheuristic algorithms are flexible and can be applied to various optimization problems, solutions are not guaranteed to correspond to the global optimum. Still, they provide good approximations for complex problems. Metaheuristics techniques can be divided into metaphor-based and non-metaphor-based approaches. The former includes algorithm such as Simulated annealing (SA) [82, 83], Ant Colony Optimization for Continuous Domains (ACOR), PSO [84] and Genetic Algorithm (GA) [85, 86]. ACOR and PSO are examples of algorithms inspired by biological systems that use Swarm Intelligence (SI). SI algorithms simulate the behaviour of a group of agents, where candidate solutions are updated by interaction with the environment and other agents. GA are Evolutionary Algorithm (EA), that model the evolution progression of cells in nature employing mutation, selection, crossover and reproduction schemes [81].

2.4.2.1 Particle Swarm Optimization

PSO is a global search algorithm that employs swarm intelligence [87]. It starts by generating a randomly initialized population of candidate solutions called particles. Each particle is characterized by its position, velocity, and personal best value. The particles are updated at each iteration of the algorithm for a predefined number of times, following the velocity and position Equations (2.3) and (2.4) [88].

$$v_i^{k+1} = wv_i^k + c_1 r_1 (P_b^k - x_i^k) + c_2 r_2 (P_g^k - x_i^k), \tag{2.3}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}, \tag{2.4}$$

where i is the particle index, k is the current algorithm iteration, w is the inertial constant that can gradually decrease with each iteration. v_i is the velocity, which can be limited by a maximum value to prevent swarm explosions. c_1 is the cognitive constant, c_2 is the social constant, r_1 and r_2 are random numbers that follow a normal distribution between zero and one. x is the positions of the particles. P_b is the personal best, which, in a minimization problem, corresponds to the position of the particle with the lowest fitness value from the first to the current iteration. All particles have an associated P_b value. P_g corresponds to the best global position, which is the particle's position with the lowest fitness value across all iterations and all particles. The fitness function evaluates candidate solutions and is described by Equation (2.5).

$$fit(x) = obj(x) + pen(x); \tag{2.5}$$

$fit(x)$ is the fitness function, $obj(x)$ is the objective function and $pen(x)$ is the penalty function. The purpose of the penalty function is to ensure that the solution complies with the constraints. In a minimization problem, a lower fitness value for a candidate solution indicates a better solution.

The PSO method is one of the most effective algorithms in solving optimization problems [89]. The algorithm is suitable for parallel programming, but can be computationally costly and present a long convergence period.

2.4.3 Neural Networks

An Artificial Neural Networks (ANN) is a computational Machine Learning (ML) model, drawing inspiration from the structure and function of the human brain [90]. A NN works as a function approximator and consists of a network of interconnected units known as neurons. In a NN, neurons are organized into layers. A NN has two or more layers, usually one input layer, a set of hidden layers and one output layer. The hidden layers are situated between the input and the output layers. Figure 2.12 illustrates an example of a neural network topology.

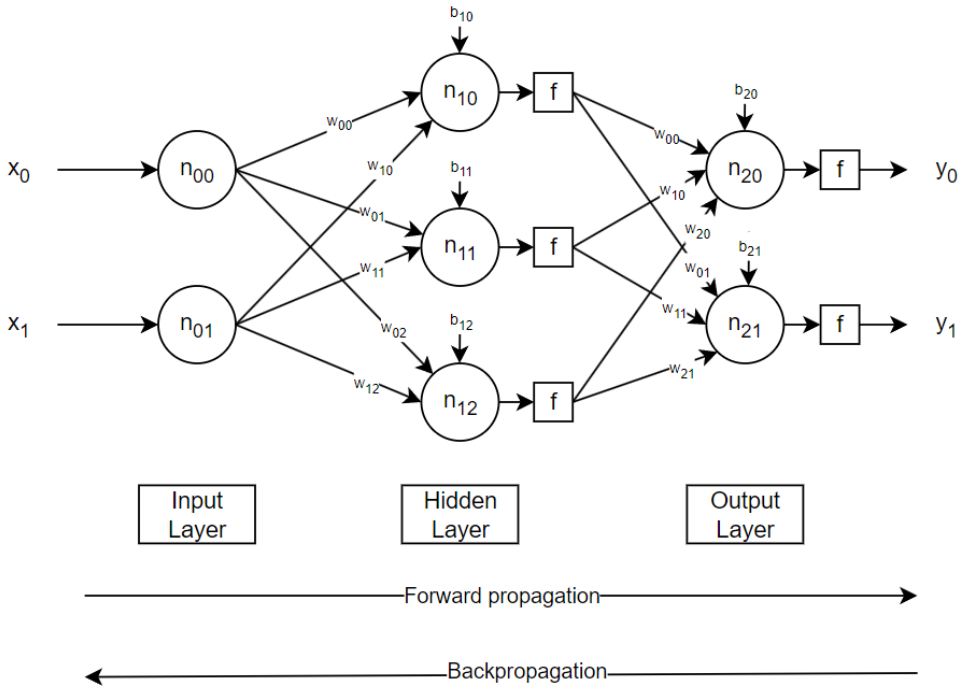

Figure 2.12: Illustration of the topology of a simple NN.

Each neuron has a set of weights, with a weight per input and a bias value. The neurons of a layer are connected to the neurons of the subsequent layer. Equation (2.6) describes a neuron's output.

$$y = f(\sum_{i=0}^{n} x_i \times w_i + b), \tag{2.6}$$

where y is the neuron's output, f is the activation function, n is the number of inputs, x_i is the input, w_i is the neuron's weight and b_i associated with the input i. The neuron's output is calculated by applying the activation function to the sum of the product of the neuron's weights and inputs, plus a bias parameter.

The activation function enables the NN to model non-linear relationships between the inputs and outputs and allows the network to solve complex problems. Examples of commonly used activation functions are the hyperbolic tangent, described by Equation (2.7) and shown in Figure 2.13, the sigmoid and the Rectified Linear Unit (ReLU), shown by Equation (2.8) and Figure 2.14.

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.7}$$

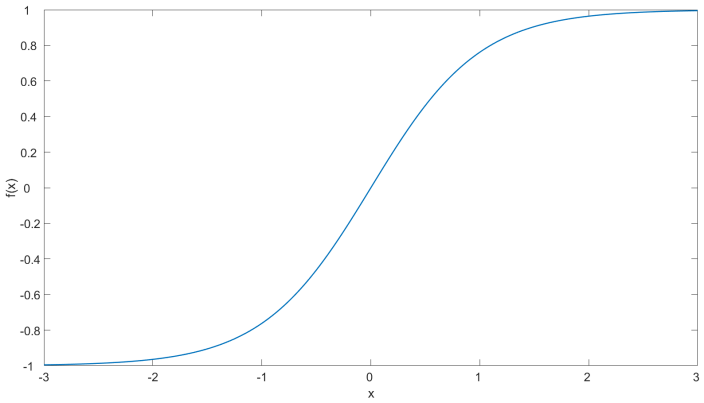

Figure 2.13: Representation of the hyperbolic tangent function.

$$relu(x) = max(0, x) \tag{2.8}$$

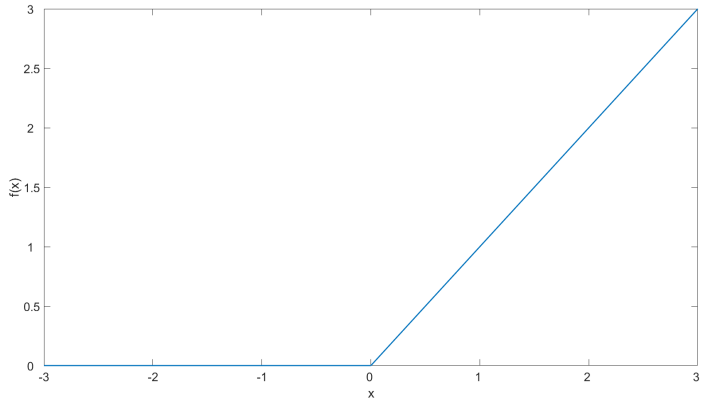

Figure 2.14: Representation of the ReLU function.

Matrix representation is used to reduce the computation complexity of the calculations in a NN. Equation (2.9) describes the calculation in a matrix form, for the output of a neuron.

$$Y = f(X \cdot W + B) \tag{2.9}$$

Where Y is the $1 \times n$ output matrix, f is the activation function applied element-wise, X is the $1 \times n$ vector of inputs, W corresponds to the $n \times i$ weights matrix, and B is the $1 \times i$ bias matrix. n represents the number of inputs and i equals the number of neurons in the layer. The matrices Y, X, W and B are represented in Equations (2.10), (2.11), (2.12) and (2.13), respectively.

$$Y = \begin{bmatrix} y_0 & \cdots & y_n \end{bmatrix} \tag{2.10}$$

$$X = \begin{bmatrix} x_0 & \cdots & x_i \end{bmatrix} \tag{2.11}$$

$$W = \begin{bmatrix} w_{00} & \cdots & w_{0n} \\ \vdots & \ddots & \vdots \\ w_{i0} & \cdots & w_{in} \end{bmatrix} \tag{2.12}$$

$$B = \begin{bmatrix} b_0 & \cdots & b_n \end{bmatrix} \tag{2.13}$$

where n is the number of neurons in the layer, and i is the number of inputs.

The weights and bias values control the behaviour of the network and are adjusted incrementally during the training phase of the NN. The first step of the training phase involves forward propagation, where the input values pass through the layers, and the output of the network is calculated. After the forward propagation phase, backpropagation adjusts the weights and bias parameters based on an optimization algorithm such as the gradient descent method presented in Equation (2.14).

$$x_{t+1} = x_t - \eta \nabla f(x_t) \tag{2.14}$$

where x_{t+1} is the newly updated position, x_t is the current position, η corresponds to the step size and $\nabla f(x_t)$ is the gradient of a function, f in relation to the position x_t.

The loss function measures the discrepancy between the expected and estimated output of the network for the training data. The $L2$ function, displayed in Equation (2.15), is a widely used loss function.

$$L2 = \sum_{i=0}^{n} (y_i - \bar{y_i})^2 \tag{2.15}$$

where y_i is the expected output and $\bar{y_i}$ corresponds to the estimated value.

The NN's training phase aims to minimize the loss function value, as a lower value indicates a better model.

The gradient descent algorithm updates the neuron's parameters by moving down the loss function's combined error surface and updating each weight and bias through the backpropagation algorithm. The backpropagation algorithm adjusts the weights and biases proportionally to the network loss value from changes in each respective weight. The weights and bias values are determined by the partial derivatives of the loss function with respect to each weight, described by Equation (2.16).

$$\frac{\partial L2}{\partial w_i} = \frac{\partial L2}{\partial \bar{y_i}} \frac{\partial \bar{y_i}}{\partial w_i} \tag{2.16}$$

In Equation (2.16), w_i is the weight and the gradient represents the sensitivity of the loss function to changes in the weight parameter.

The chain rule is applied to update the neurons' parameters during the backpropagation algorithm. The calculation can be computationally simplified using matrix form, as shown in Equations (2.17), (2.18) and (2.19). This approach enables the simultaneous update of all neurons in a layer.

$$L_t = \begin{bmatrix} \frac{\partial L2}{\partial w_0} \cdots \frac{\partial L2}{\partial w_n} \end{bmatrix} \tag{2.17}$$

$$W_{t+1} = W_t + \eta X_t^T \cdot L_t \tag{2.18}$$

$$B_{t+1} = B_t + \eta \odot L_t \tag{2.19}$$

W_{t+1} is the updated weight matrix, W_t is the current weight matrix, η is the step size, X_t^T is the transpose of the neuron's inputs and L_t represents the loss matrix. In Equation (2.17), L_t corresponds to the vector of the partial derivatives of the loss function $L2$ with

respect to each weight, for iteration t. B_{t+1} is the updated bias matrix, and B_t is the value of the bias matrix before it was updated.

The backpropagation method starts by updating the neurons of the output layer and then proceeds regressively by recalculating the loss vector for each previous layer, based on Equation (2.20).

$$L_{t+1} = L_t \cdot W_{t+1}^T \tag{2.20}$$

The loss vector, L_{t+1}, is updated by multiplying the previous loss matrix, L_t, by the new values of the transposed weight matrix, W_{t+1}^T. Equations (2.18) and (2.19) are then applied with the new value of L_t. The process continues for all layers of the network.

Training occurs over multiple cycles, called epochs. To improve efficiency during training, the input data can be divided into mini-batches with a specific batch size. The new weights and bias values for each mini-batch element are averaged. NN are commonly used in classification problems, where labelled training data is used to train the network. The objective of a classification problem is to determine the category to which a given input data point belongs. ML algorithms, including NN, are employed to solve problems in high-dimensional spaces that cannot be exhaustively searched.

Bayati et al. proposed a technique for solving constrained continuous optimization problems using unsupervised NN [91]. The NN uses the objective function of the optimization problem as the loss function of the network, which includes a penalty term to discourage solutions that violate the constraints.

2.4.4 Fourier Series

The Fourier series represents a function as a sum of sine and cosine functions with different frequencies. The general form of a Fourier series is given by [92]:

$$f(x) = a_0 + \sum_{i=1}^{\infty} a_i cos(ix) + \sum_{i=1}^{\infty} b_i sin(ix) \tag{2.21}$$

where a_0 is the constant term, a_i and b_i are the coefficients defining the cosine and sine amplitude. i indicates the frequency index. The variable x ranges from $-\pi$ to π.

The Fourier series has multiple applications, including signal analysis and processing, image compression, and some applications in machine learning. Tancik et al. suggested that passing the input of a NN through a Fourier feature mapping could improve the results in a function approximation problem [93].

Chapter 3

State of the Art

The current chapter reviews key studies and research in the field of NILM, with an emphasis on unsupervised and low-frequency NILM algorithms. The literature review followed a semi-systematic methodology [94, 95]. First, review papers were studied, followed by papers that applied specific techniques. The search criteria included the keywords "NILM", "unsupervised", "low-frequency", and "industry" in databases such as *Scopus*[1], *ScienceDirect*[2], *IEEE Xplore*[3], *ArXiv*[4] and *Google Scholar*[5]. Articles were selected based on the number of citations, publication date, abstract section and full text.

3.1 Literature Review

George Hart first introduced NILM [5], where an edge detection method was applied to the normalized values of the aggregate active power. Edge detection worked by identifying steady periods, designated as periods, with a minimum length of three samples and without fluctuation in the active power more significant than the tolerance value of 15W. The samples were averaged in each steady period, and the difference between consecutive periods was calculated. A clustering algorithm was applied to the computed power difference values, creating ON/OFF or Finite State-Machines (FSM) models. These models allow for tracking individual equipment through a decoding approach. However, the algorithm presents limitations since it relies on the detection of events, requires periods with slight fluctuations, is unable to disaggregate equipment of type III and can not handle simultaneous events. Hart's method is also prone to errors due to the possibility of a period being created by the events of different equipment and the possible presence of equipment with equal consumption values, which affects the clustering algorithm.

Following the algorithm proposed by Hart, multiple methods have been implemented using a similar structure. Subsequent work has mainly focused on computing harmonicas at events and using transient noise with supervised learning algorithms.

Ruando et al. [96], Gopinath et al. [97], Angelis et al. [98], Zoha et al. [99] and Faustine et al. [19] have conducted literature reviews that present the state of the art of NILM algorithms.

[1]https://www.scopus.com/
[2]https://www.sciencedirect.com/
[3]https://ieeexplore.ieee.org/
[4]https://arxiv.org/
[5]https://scholar.google.com/

Ruando et al. described the approaches used for the first component of the classical NILM algorithm, this is event detection. Event detection can be divided into three approaches: expert heuristics, probabilistic models, and matched filters [96].

Expert heuristics involves defining a set of rules to perform event detection. Fixed thresholding is an example of a typical implementation. Multiple thresholds based on different features can be used to improve the results. An example of a multiple thresholding technique is the multivariate event detection method. An alternative to fixed thresholding is adaptive thresholding, applied in techniques like enveloped-based peak detection [96].

Probabilistic methods require a training process to estimate statistical features for each equipment. Probabilistic models use techniques such as Generalized Likelihood Ratio (GLR), chi-squared, Cumulative Sum (CUSUM) and Bayesian information criterion [96].

Matched filters correlate signal waveforms to known patterns and require high sampling rates and prior knowledge of the equipment load signatures. The correlation is calculated by performing clustering techniques [100, 101].

The feature extraction component [102] of the classical NILM method depends on the sampling rate, where the most employed features are RMS current, RMS voltage, active, reactive and apparent power, total harmonic distortion and power factor. A high sampling rate allows for the capture of harmonics using the Fourier transform. At very high sampling frequencies, two-dimensional voltage-current trajectories, electrical noise and Electromagnetic Interference (EMI) signals can be obtained. Nontraditional features, such as temperature, light sensing and time of day, are also used by some implementations.

Load identification or source separation can be performed using optimization, supervised or unsupervised techniques [96, 103]:

1. Optimization techniques aim to disaggregate the power measurement into combinations of the individual equipment power signals:

    - Genetic Algorithm (GA);
    - Particle Swarm Optimization (PSO);
    - Multi-label classification.

2. Machine learning supervised techniques are methods that use offline training and labelled training data that corresponds to the individual equipment consumption data:

    - Support Vector Machines (SVM);
    - Different types of Neural Network (NN): Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Autoencoders;
    - Bayes classifiers;
    - Decision-tree;
    - Autoregression;
    - k-Nearest Neighbours (kNN).

3. Unsupervised techniques do not require labelled training data. In the literature, the majority of unsupervised NILM algorithms implement variations of the Hidden Markov Model or clustering algorithms:

    - Factorial Hidden Markov Method (FHMM);
    - Conditional Factorial Hidden Markov Method (CFHSMM);

- Additive Factorial Approximate Maximum A Posteriori (AFAMAP);
- Hierarchical Dirichlet Process Hidden Semi Markov Model (HDP-HSMM);
- Clustering techniques, such as Density-based Spatial Clustering of Applications with Noise (DBSCAN).

4. Other methods that can be supervised or unsupervised:

   - Graph Signal Processing (GSP);
   - Matrix Factorization (MF);
   - Dynamic Time Warping (DTW);
   - Sparse Coding (SC).

Barsim et al. [6] proposed a method based on Hart's method that uses transient information. The authors used a sliding window to detect events in the logarithmically transformed active and reactive power signals. The size of the sliding window was set to contain one transient event and two steady-states. A grid-based clustering scheme similar to density-based clustering was applied to the values calculated at each event to identify the equipment. The researchers tested the algorithm using a dataset from a domestic environment. Wang et al. presented another method inspired by Hart's solution [104], to classify individual loads in a household setting. The event detection process was not explained in detail. The available information indicates that power consumption was classified into three categories, namely insignificant fluctuations, fast switching and steady working events. For each event, the start, peak and end times, along with peak values, which include active and reactive power, were acquired. The mean active and reactive power and the variance of active power in steady-state values were calculated. Mean-shift clustering was applied, iteratively, to the acquired values. The clustering results were used along with a knowledge base in a linear discriminate analysis to classify the equipment. Both methods perform load classification. Barsim et al. used an unsupervised approach, whereas Wang et al. used a supervised technique.

Liu et al. reviewed different unsupervised NILM algorithms [4]. These algorithms were mainly variations of HMM or GSP. The reviewed algorithms construct state-equipment models, assuming temporal dependencies and patterns. However, HMM and GSP cannot perform source separation in the presence of type III equipment. HMM and GSP can only perform load classification of type I and II equipment.

Bonfigli et al. [105] conducted an overview study of unsupervised NILM algorithms. The authors divided the unsupervised algorithms into load classification and source separation techniques. The load classification techniques presented are variations of HMM and clustering methods. Bonfigli et al. presented the work by Figueiredo et al. [106] as a supervised source separation method.

Two papers were studied in more detail, the work developed by Kolter et al. [107], and by Figueiredo et al. [106]. Both methods are supervised and construct equipment models using aggregate and equipment data. The models perform the separation of the aggregate signal. Kolter et al. developed a Discriminative Disaggregation via Sparse Coding (DDSC) method [107]. The DDSC method works by training separate models for each equipment and then using the models to disaggregate an aggregate signal. The model consists of a matrix of basis functions called a dictionary and an activation matrix.

The method developed by Kolter et al. can be divided into a training and an estimation phase. During training, the method creates separate models for each equipment. The models are calculated by an optimization method that switches between estimating the

dictionary and the activation matrix. The activation and basis matrices are then calculated with a proposed method developed by Kolter et al. called the augmented regularized disaggregation error. The estimation phase requires solving another optimization problem to calculate a new activation matrix. The gradient descent method and a structure perception-based algorithm are used to optimize the problem.

Figueiredo et al. developed a method called Source Separation via Tensor and Matrix Factorization (STMF) [106], where tensors are composed of three domains: time, day and individual equipment data. Tensors are first decomposed by the PARAFAC method and then further decomposed into Nonnegative Matrix Factorization1 (NMF). According to Figueiredo et al., the STMF implementation presented better results than the DDSC method.

3.2 Chapter Summary

The field of NILM has primarily followed the method proposed by Hart [5]. The only notable innovation was the incorporation of transitory data from high-frequency sampling, the use of supervised algorithms based on labelled training data, and the use of unsupervised HMM that are incompatible with continuous domains. Kolter et al. [107] and Figueiredo et al. [106] proposed equipment modelling methods that require training data. Both methods were not designed to solve non-linear problems, such as NILM problems where type III equipment is preset and are computationally intensive processes. A new unsupervised approach was needed to address the gaps in the literature for low-frequency unsupervised NILM methods for industrial loads.

Chapter 4

Methodology

This chapter outlines the development of the two proposed methods. The chapter presents the preprocessing of the HIPE and IMDELD datasets, details the Equipment Modelling Using Polynomial Functions (EMUPF) and Unsupervised Neural Network (UNN) methods and describes the testing of the models. The testing process compromises the analysis of the results through error metrics and descriptive statistics. The methodology followed is depicted in Figure 4.1.

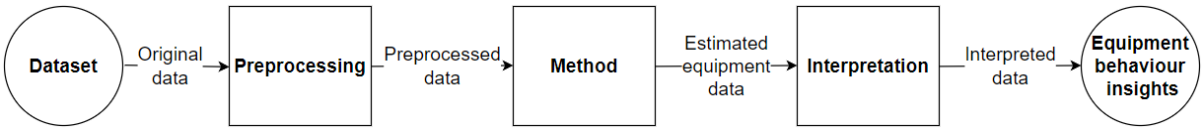

Figure 4.1: Diagram outlining the methodology.

4.1 Dataset Preprocessing

This section addresses the preprocessing stage required by the selected datasets. The preprocessing of the datasets was essential since the algorithm performance is highly dependent on data quality. The HIPE and IMDELD datasets required preprocessing to clean the data and achieve the desired 1Hz sampling frequency. The preprocessing stage was coded in MATLAB.

4.1.1 HIPE Dataset

In the HIPE dataset, equipment one was in the OFF state during the entire sampling period, resulting in the exclusion of its data. Therefore, only the data from nine equipment was used. The dataset had nonconsecutive samples. A sample is nonconsecutive if it has a smaller timestamp than the previous sample's timestamp. All values after the first nonconsecutive timestamp were removed since they were considered faulty samples. Information on the HIPE dataset, before preprocessing, is presented in Table 4.1.

Table 4.1: General information about the HIPE dataset, including the timestamp at which the dates stop being consecutive.

Equipment index	Number of samples	Number of unique samples	Number of missing samples	First timestamp	Last timestamp	Non-consecutive timestamp
CS	110467	110460	494340	23-10-2017 00:00:00	29-10-2017 23:59:56	29-10-2017 02:59:54
HTO	110466	110464	494336	23-10-2017 00:00:02	29-10-2017 23:59:58	29-10-2017 02:59:56
PPU	122773	122771	482029	23-10-2017 00:00:02	29-10-2017 23:59:57	29-10-2017 02:59:56
SP	122742	122741	482059	23-10-2017 00:00:04	29-10-2017 23:59:59	29-10-2017 02:59:58
SO	110492	110490	494310	23-10-2017 00:00:02	29-10-2017 23:59:57	29-10-2017 02:59:56
VO	110465	110459	494341	23-10-2017 00:00:01	29-10-2017 23:59:57	29-10-2017 02:59:55
VP1	110492	110486	494314	23-10-2017 00:00:04	29-10-2017 23:59:57	29-10-2017 02:59:56
VP2	110496	110493	494307	23-10-2017 00:00:01	29-10-2017 23:59:55	29-10-2017 02:59:59
WM	110495	110489	494311	23-10-2017 00:00:03	29-10-2017 23:59:55	29-10-2017 02:59:59

The equipment and aggregate samples were interpolated, ensuring a frequency of 1Hz. It was intended to study the effects of the number of equipment on the results. The aggregate data used is synthetic since the data was generated and does not correspond to the measurements at the main terminal energy meter. The aggregate active power was calculated as the sum of the active power samples of each equipment. In total, eight sums were calculated. The first aggregate is the sum of the first two equipment, the second aggregate is the sum of the first three equipment and so on until the last aggregate corresponds to the sum of all nine equipment. The equipment state data was computed by applying a threshold. A threshold splits the active power data into the ON state if the sample is above the threshold and into the OFF state if the sample is below. The threshold was defined as 0W. Finally, the data was divided into training and testing sets by implementing an adaptive binning algorithm. The algorithm splits the data into six bins based on the standard deviation value of the samples. Data were selected in equal numbers and randomly from each bin. The selected data were divided into a 70-30 ratio for the training and testing data. The adaptive binning technique was used to remove outliers from the training and testing data and to ensure a balanced representation of the data. This step was crucial since some equipment remained in one state for most of the sampling period.

4.1.2 IMDELD Dataset

The IMDELD dataset presented duplicate values, that is, more than one active power sample per meter for the same timestamp. Duplicate samples were averaged. Negative samples were adjusted to zero so that no negative values of the active power were present in the data. The IMDELD dataset had multiple days without valuable data due to missing samples, so days with more than one missing sample per five seconds were discarded. The five-second value was established based on experimentation and on the interpolated results. The IMDELD dataset contained multiple outliers that affected the interpolation, so a moving average sliding window with a size of 1500 samples was applied. Finally, the samples were interpolated so that, for each day, the sampling frequency was equal to 1Hz.

The information related to the original IMDELD dataset's samples is shown in Table 4.2.

Table 4.2: General information about the IMDELD dataset.

Equipment name	Number of samples	Number of unique samples	Number of duplicate samples	Number of missing samples	First timestamp	Last timestamp
DPCI	5504827	5407532	97295	8070868	30-10-2017 20:54:11	03-04-2018 18:48:49
DPCII	5501067	5386100	14890	8092300	30-10-2017 20:54:11	03-04-2018 18:48:49
EFI	5501067	5407195	93872	8071205	30-10-2017 20:54:11	03-04-2018 18:48:48
EFII	5386443	5364580	21863	8113820	30-10-2017 20:54:12	03-04-2018 18:48:49
PI	5474431	5380344	94087	8098056	30-10-2017 20:54:10	03-04-2018 18:48:49
PII	5389689	5362659	27030	8115741	30-10-2017 20:54:11	03-04-2018 18:48:48
LVDB-2	5462407	5405007	57400	8073393	30-10-2017 20:54:11	03-04-2018 18:48:48

The equipment state data was calculated with a threshold equal to 5W. Training and testing data were determined using a 70-30 ratio, randomly selecting results, in equal numbers from each bin, from the adaptive binning technique.

4.2 Methods

In the current section, the developed methods are presented, which include the EMUPF and UNN methods. Both methods create models to estimate the active power values of each equipment. The methods consist of two phases: an initial offline training phase and an online inference phase. In the training phase, the models are calculated using the aggregate active power and the equipment state training data. In the inference phase, the active power consumption values of the equipment are calculated using the estimated models from the training phase.

4.2.1 EMUPF Method

The EMUPF technique was developed in C++ with OpenMP due to a focus on performance. The method models the active power consumption of the equipment using polynomial functions. In the training phase, the coefficients of the polynomial functions are estimated using the PSO method, which minimizes an objective function. In the inference phase, the polynomial functions are used to calculate the active power values of each equipment.

The active power consumption of each equipment is defined as a third-order polynomial function, following Equation (4.1).

$$f(a_t, s_{it}) = s_{it} \cdot (c_{0n} \cdot a_t^0 + c_{1i} \cdot a_t^1 + c_{2i} \cdot a_t^2 + c_{3i} \cdot a_t^3) \tag{4.1}$$

$f(a_t, s_{it})$ is the active power value of equipment with index i and for the sample with index t, for the aggregate active power sample a_t and the equipment state s_{it}.

The degree of the polynomial function was carefully selected. Generally, a higher degree results in a more accurate model but also increases the computational complexity of the optimization problem. It is important to choose a degree that balances between underfitting, overfitting and high computational complexity. Rank three was chosen for its ability to capture the non-linearity of the problem while providing a simplified representation of the equipment's behaviour and allowing for the analysis of the feasibility of the proposed algorithm. Third-order polynomial functions are commonly used in the literature to solve load forecasting problems [108].

4.2.1.1 Training Phase

During the training phase, the objective function, denoted by $f(a, S)$ in Equation (4.2), is minimized and the matrix C is estimated, indicated by Equation (4.3).

$$f(a, S) = (a - JACS)^2 + \lambda p \tag{4.2}$$

$$\min_{C} \quad (a - JACS)^2 + \lambda p \tag{4.3}$$

a is the aggregate active power sample. S is the equipment state matrix, with the state samples for each equipment. $JACS$ represents the sum of the equipment active power values, represented in Equation (4.8). $JACS$ corresponds to the sum of polynomial functions. λ is a regularization parameter multiplied by the penalty value, p, denoted by Equation (4.9). The inputs of the training phase are the aggregate active power and the equipment state samples. Equation (4.3) is minimized for each *input*, this is one aggregate active power sample and the equipment state samples for all the equipment at the same timestamp t. The objective function is minimized for all inputs. Then, the estimated coefficients are averaged.

The objective function calculates the squared difference between the aggregate active power and the sum of the active power values of the equipment in the ON state while penalizing negative values for the estimated equipment's active power.

The matrices J, A, C and S are denoted by Equations (4.4), (4.5), (4.6) and (4.7).

$$J = \begin{bmatrix} 1 \cdots 1 \end{bmatrix} \tag{4.4}$$

$$A = \begin{bmatrix} a^0 & a^1 & \cdots & a^r \\ \vdots & \vdots & \ddots & \vdots \\ a^0 & a^1 & \cdots & a^r \end{bmatrix} \tag{4.5}$$

$$C = \begin{bmatrix} c_{00} & c_{10} & \cdots & c_{n0} \\ c_{01} & c_{11} & \cdots & c_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{0r} & c_{1r} & \cdots & c_{nr} \end{bmatrix} \tag{4.6}$$

$$S = \begin{bmatrix} s_0 \\ \vdots \\ s_n \end{bmatrix} \tag{4.7}$$

n is the total number of equipment in the aggregate. r is the degree of the polynomial functions. J is the unit matrix of size $1 \times n$. A is the aggregate matrix with size $n \times r$. The coefficients matrix C has size $r \times n$. The state matrix S has size $n \times 1$.

$$JACS = \sum_{i=1}^{n} s_i \cdot (c_{0i} \cdot a^0 + c_{1i} \cdot a^1 + \cdots + c_{ri} \cdot a^r) \tag{4.8}$$

where i is the equipment index, s_i is the state of the equipment with index, i, c represents the coefficients of the polynomial function and a is the aggregate active power value. r was defined as three. For the data used from the HIPE dataset, the n value is nine, and for the IMDELD dataset, n is six.

The penalty function, described by Equation (4.9), penalizes negative equipment active power values.

$$p = \sum_{i=1}^{n} p_i \tag{4.9}$$

where p_i is the penalty value for equipment with index i, shown by Equation (4.10).

$$p_i = \begin{cases} |\phi \cdot e_i|^{\gamma}, & \text{if } e_i < 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.10}$$

e_i, denoted by Equation (4.11), corresponds to the estimated active power value of the equipment i. ϕ and γ are positive constants. ϕ was defined as three and γ as two. (4.11)

$$e_i = s_i \cdot (c_{0i} \cdot a^0 + c_{1i} \cdot a^1 + \cdots + c_{ri} \cdot a^r) \tag{4.11}$$

Different optimization algorithms can be used to minimize the objective function and estimate the coefficients of polynomial functions. Metaheuristic algorithms were implemented, including SA [83], PSO [87], GA [109], and ACOR [110]. The PSO method was chosen as it provided the most reasonable estimates for the coefficients. The PSO, defined by Equations (2.3) and (2.4), was initialized with 1000 particles per unknown variable, which is 1000 particles per dimension. Multiplying the number of equipment by four corresponds to the dimension of the search space since a polynomial function with four coefficients represents each equipment. The maximum number of cycles was set at 200. The particle's position is a vector with a size equal to the problem's dimension. The method was trained using all aggregate active power and equipment state samples in the training data. The coefficients were estimated for all training samples in parallel using OpenMP and finally averaged. The optimization of the objective function is not deterministic because the results depend on the random variables used by the PSO method and due to the complex objective function with multiple local optima and high-dimensional search space. The results may differ for each run of the optimization method. Figure 4.2 illustrates a diagram of a single training phase run for the EMUPF method.

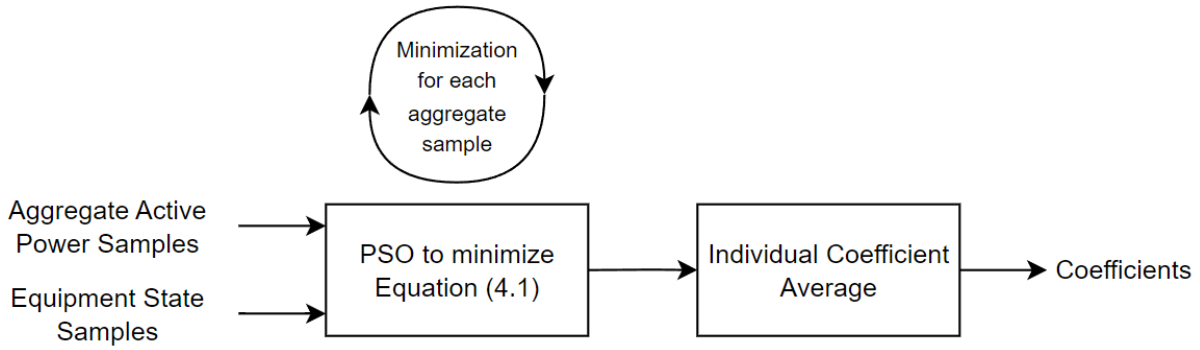

Figure 4.2: Diagram representing a single training phase run for the EMUPF method.

The EMUPF method was run five times and the coefficients of the run with the lowest mean objective function value were selected. The main steps for the method's training phase are presented in Algorithm 1.

Algorithm 1 EMUPF - Training Phase

Input:

1. Aggregate active power (a_train) training data
2. Equipment states (S_train) training data
3. PSO parameters

Output:

1. Coefficients C

Procedure:

Initialize old average results $r_{old} \leftarrow FLOATMAX$
for $i = 0, 1, \cdots, 4$ **do**
    for each $input \in inputs$ **do**
        Run PSO technique to minimize Equation (4.2)
        Get objective function result
        Get estimated coefficients C
    end for
    Average objective function results r_{new}
    if r_{new} is less than r_{old} **then**
        Compute coefficient-wise average C
        Save averaged coefficients C
        Update old average results $r_{old} \leftarrow r_{new}$
    end if
end for

4.2.1.2 Inference Phase

The active power values of the equipment are estimated by calculating the results of polynomial functions. These functions are defined by the estimated coefficients and use the aggregate active power and the equipment state data as input, which is shown in Equation (4.1).

The inference phase is presented in Figure 4.3 and Algorithm 2, where the active power of each equipment is estimated, for a single aggregate active power sample.

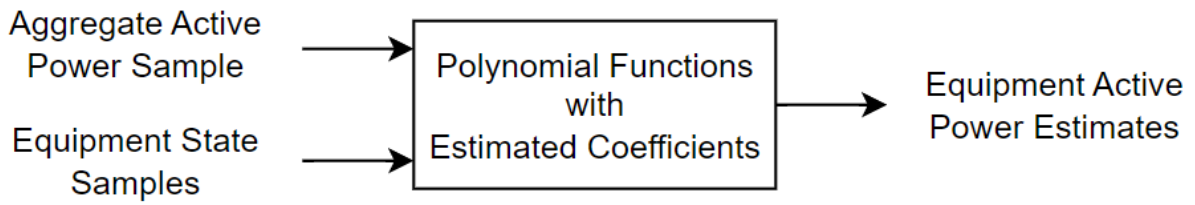

Figure 4.3: Diagram illustrating the inference phase of the EMUPF method for disaggregating an aggregate active power sample into the estimated active power values for each equipment.

Algorithm 2 EMUPF - Inference Phase

Input:

1. Aggregate active power sample a
2. Equipment states samples s
3. Coefficients C

Output:

1. Estimated equipment active power consumption

Procedure:

1: **for each** equipment index i $(i = 1, 2, \cdots, n)$ **do**
2: Compute Equation (4.1)
3: **end for**

The EMUPF method proposes a new approach to solve the NILM problem by modelling the active power consumption of the equipment with polynomial functions. No method that defines an objective function that includes the sum of polynomial functions and uses optimization algorithms to estimate the coefficients of the functions has been found in the literature.

4.2.2 UNN Method

The unsupervised neural network was coded from scratch using Python, without any machine learning library like PyTorch or TensorFlow. The NumPy library was used as it simplifies matrix operations. Python was selected because it streamlines and speeds up the development process. The UNN differs from conventional NN because it does not require labelled training data. The method comprises two phases: training and inference. In the training phase, the network model is defined using the backpropagation algorithm in the inference phase, the network is used to estimate the equipment's active power consumption through forward propagation. The UNN method was implemented with two different architectures. The first network architecture uses the aggregate active power and the equipment state samples as input while the second passes the aggregate active power input through a Fourier feature map.

The Fourier mapping had sixteen features. The matrix of Fourier features is represented by Matrix (4.12).

$$\begin{bmatrix} sin(a) \\ cos(a) \\ sin(2a) \\ cos(3a) \\ \vdots \\ sin(8a) \\ cos(8a) \end{bmatrix} \tag{4.12}$$

where a is the aggregate active power.

Both networks consisted of three layers: one input layer, one hidden layer and one output layer. The number of output neurons is equal to the number of equipment since each output neuron represents the active power consumption of one equipment. The output neuron with index one represents the active power consumption of equipment with index one, the output neuron with index two represents the equipment with index two and so on until the output neuron with index n represents equipment index n. All layers

used the ReLU activation function except the output layer, which used the hyperbolic tangent activation function.

4.2.2.1 Training Phase

The learning rate of the network was set at 0.001 and the mini-batch size was defined as four samples. Training had a duration of 1000 epochs.

The loss function of the network, shown by Equation (4.13), makes the method unsupervised since the function depends only on the inputs and estimated outputs of the network and not on the expected values.

$$l_i = (a - \sum_{i=0}^{n} s_i \cdot o_i)^2 + \sum_{i=0}^{n} p_i \tag{4.13}$$

where l_i is the loss function value for the output neuron i, a is the input aggregate active power sample, s_i is the state of the equipment with index i, o_i is the output value of the neuron and p_i is the value of the penalty function.

Equation (4.14) shows the derivative of the loss function used by the backpropagation algorithm to update the weights and bias of the network's neurons.

$$\frac{\partial l_i}{\partial o_i} = -2 \cdot (a - \sum_{i=0}^{n} s_i \cdot o_i) + \frac{\partial p_i}{\partial o_i} \tag{4.14}$$

The penalty value was calculated for each output neuron following Equation (4.15).

$$p_i = \begin{cases} |\phi \cdot o_i|^\gamma, & \text{if } o_i < 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.15}$$

p_i is the penalty value for the output, o_i. For the current problem, both ϕ and γ were defined as three. The derivative of the penalty function is:

$$\frac{\partial p_i}{\partial o_i} = \begin{cases} \frac{\gamma |o_i|^\gamma |\phi_i|^\gamma}{o_i}, & \text{if } o_i < 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.16}$$

The input aggregate values were normalized based on the maximum and minimum values of the training aggregate active power. The estimates are denormalized at the end. For the Fourier mapping, the inputs must be normalized between $-\pi$ and π.

$$\hat{a}_i = \frac{a_i - min(a)}{max(a) - min(a)} \tag{4.17}$$

$$d_i = o_i \cdot (max(a) + min(a)) \tag{4.18}$$

where i is the sample index, $\hat{a}_i$ is the normalized aggregate active power sample, a_i represents the aggregate active power sample, $min(a)$ is the minimum and $max(a)$ is the maximum value of the aggregate active power. d_i is the denormalized value of the output o_i.

A representation of the training phase for the UNN method with both architectures is shown in Figures 4.4 and 4.5. The main steps of the UNN method training phase are represented in Algorithm 3.

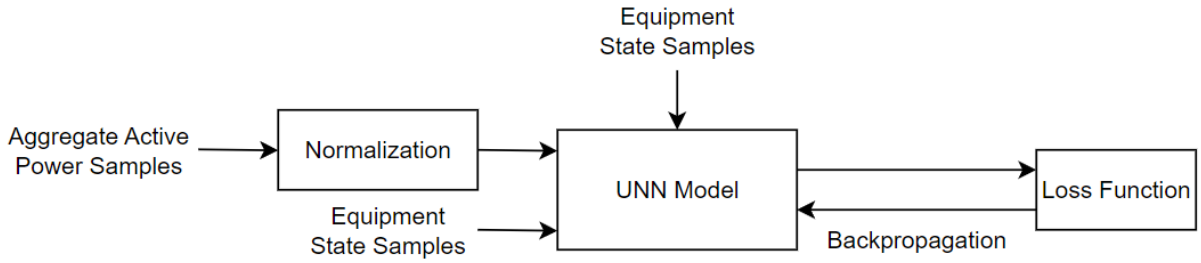

Figure 4.4: Diagram illustrating the training phase for the UNN method with the aggregate active power and the equipment state data as input. Equipment state samples are part of the input layer of the network and are used by the objective function.

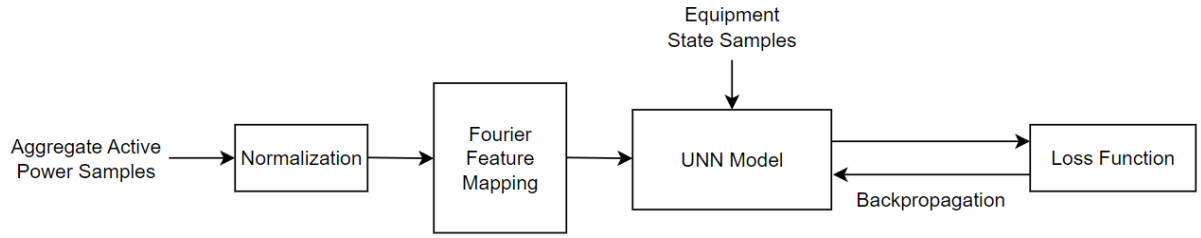

Figure 4.5: Diagram depicting the training phase for the UNN method with Fourier mapping. The objective function uses the equipment state data, but the equipment state samples are not used in the input layer of the network.

Algorithm 3 UNN - Training Phase

Input:

1. Aggregate active power sample a
2. Equipment states samples s
3. Network parameters

Output:

1. Network model

Procedure:

```
Normalize aggregate active power samples
Initialize old averaged loss l_old ← FLOATMAX
for i = 0, 1, ..., 19 do                                   ▷ Train 20 networks
    for epoch = 0, 1, ..., 1000 do
        for each input ∈ inputs do
            for each layer ∈ Layers do
                Forwards propagation
            end for
            Calculate loss value for output layer using Equation (4.13)
            for each layer ∈ Reversed Layers do
                Backwards propagation using loss value
            end for
        end for
    end for
    Average loss values for output layer l_new
    if l_new is less than l_old then
        Save new network model
        Update l_old ← l_new
    end if
end for
```

In Algorithm 3, the *inputs* for the first network architecture are the aggregate active power and the equipment state samples, and for the second network architecture are the feature from the Fourier mapping of the aggregate active power.

Twenty networks with the same architecture were trained. At the end of the training phase, the average loss was calculated for all the outputs and the network model with the lowest mean loss value was selected.

4.2.2.2 Inference Phase

The inference phase for the UNN method and the UNN method with Fourier mapping are depicted in Figures 4.6 and 4.7. In the inference phase, the active power of each equipment is estimated. In this phase, forward propagation occurs for a single aggregate active power sample.

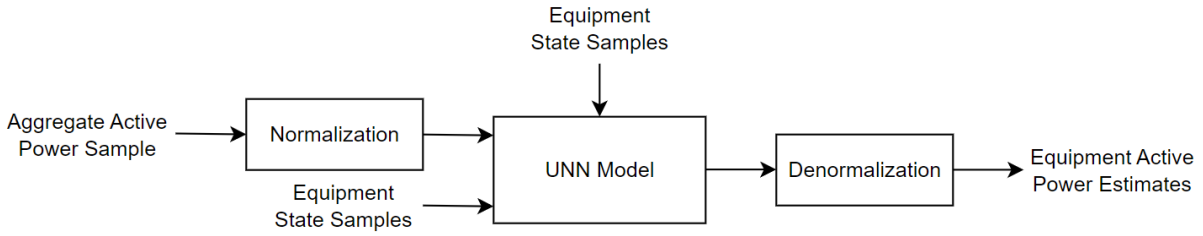

Figure 4.6: Diagram illustrating the UNN method's inference phase for a single aggregate active power sample. The method has as input the aggregate active power and the equipment state data. Equipment state samples are part of the input layer of the network and are used by the objective function.

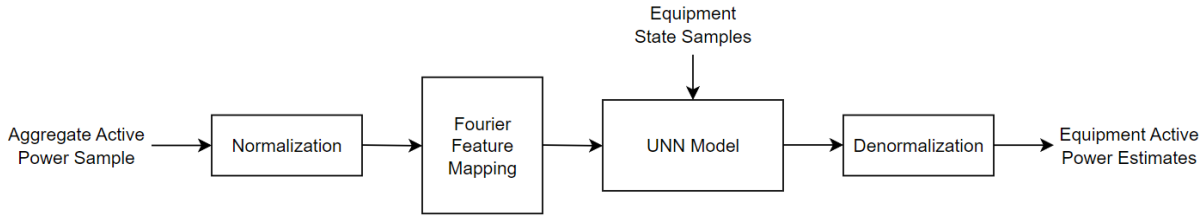

Figure 4.7: Diagram representing the inference phase for the UNN method with Fourier mapping for a single aggregate active power sample. The objective function uses the equipment state data, but the samples are not used in the input layer of the network.

The inference phase is also presented in Algorithm 4.

Algorithm 4 UNN - Inference Phase

Input:

1. Aggregate active power sample a
2. Equipment states samples s
3. Network model

Output:

1. Estimated equipment active power consumption e_i

Procedure:

1: Normalize aggregate active power
2: **for each** $layer \in Layers$ **do**
3: Forwards propagation
4: **end for**
5: Denormalize outputs
6: Save estimates e_i

During the inference phase, the active power consumption values of the equipment were estimated by forward propagation of the inputs for the trained network model.

In the literature, no work was found that used an unsupervised neural network to solve a NILM problem. The first Unsupervised Neural Network architecture, network's parameters and objective function for a NILM problem was defined, and its applications were studied. For the first time, the application of Fourier feature mapping was studied for an UNN.

4.3 Descriptive Statistical Analysis

The final step involved analyzing the results of the EMUPF and UNN methods. The maximum, minimum, mean, median and sum values of the equipment's active power consumption were calculated. The maximum, minimum and median values provide a summarized understanding of the distribution of the estimated values. The mean and sum values give a better insight into the equipment's behaviour and identify the high-active-power consumers.

4.4 Chapter Summary

The HIPE and IMDELD datasets were preprocessed. The preprocessed data was used to validate the developed methods. Two methods were developed, the EMUPF and UNN methods. The EMUPF method uses polynomial functions to model the active power consumption of the equipment. The aggregate active power is the variable of the functions. The polynomial functions were calculated using the PSO algorithm to minimize an objective function. The UNN method is a neural network where an objective function defines the loss function. The number of output neurons corresponds to the number of equipment in the aggregate, and each neuron's output corresponds to the estimated active power consumption for each equipment. Two network architectures were studied. The first UNN uses the aggregate active power and the equipment state data as input. A second UNN architecture was implemented, passing the aggregate active power through a Fourier feature mapping. Fourier mapping was shown to improve results in the literature. A descriptive statistical analysis was performed on the estimated values by the EMUPF and UNN methods.

Chapter 5

Results

In the current chapter, the results of the preprocessing stage and the results of the two proposed methods are presented. The methods were assessed using the MAE, MSE and RMSE error metrics, described in Equations (5.1), (5.2) and (5.3). The error metrics calculate the error between the expected and estimated equipment active power values. The expected values correspond to the equipment active power values from the testing data from both datasets. The estimated equipment active power values were the result of the methods using the testing data as input.

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|y_i - \bar{y_i}| \tag{5.1}$$

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(y_i - \bar{y_i})^2 \tag{5.2}$$

$$RMSE = \sqrt{MSE} \tag{5.3}$$

where y_i is the expected output and $\bar{y_i}$ corresponds to the estimated value. The error metrics were calculated in kW.

5.1 Dataset Preprocessing

This section presents the results from the preprocessing stage. This stage processed the aggregate and the equipment data from the HIPE and the IMDELD dataset. The output of the preprocessing stage was the training and testing data.

5.1.1 HIPE Dataset

The data for eight different aggregates was synthetically calculated. The eight aggregates allow for the study of the impacts of the number of equipment and are displayed in the figures in the Appendix B.1. When considering all eight equipment, the preprocessed aggregate active power and equipment state training data had a total of 1956 samples. The preprocessed data are shown in Figures 5.1, 5.2, 5.3 and 5.4, respectively, for the HIPE dataset. The eight aggregate active power, equipment state and equipment active power testing data had 584 samples each.

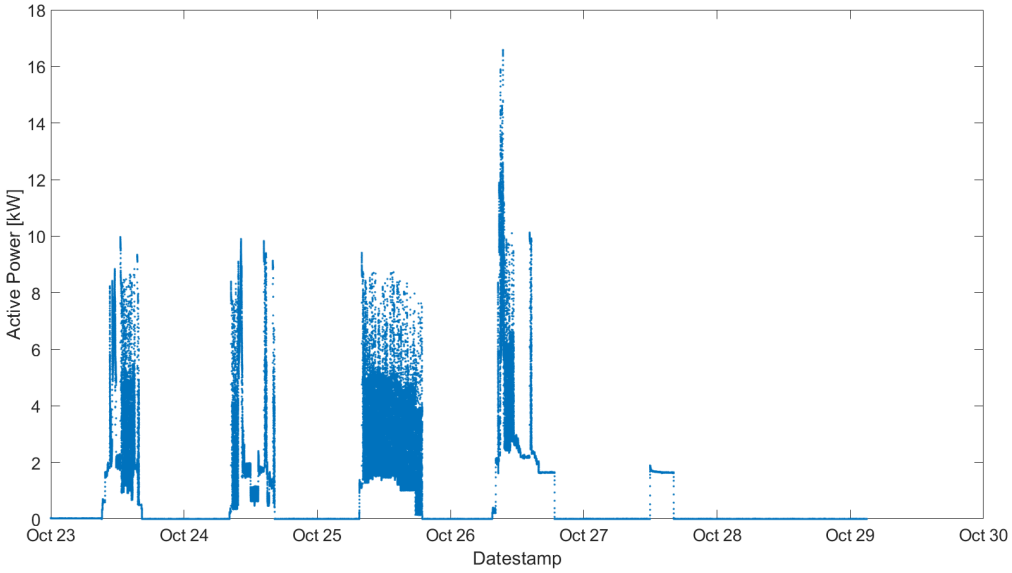

Figure 5.1: Preprocessed aggregate active power that results from the sum of nine equipment in the HIPE dataset.

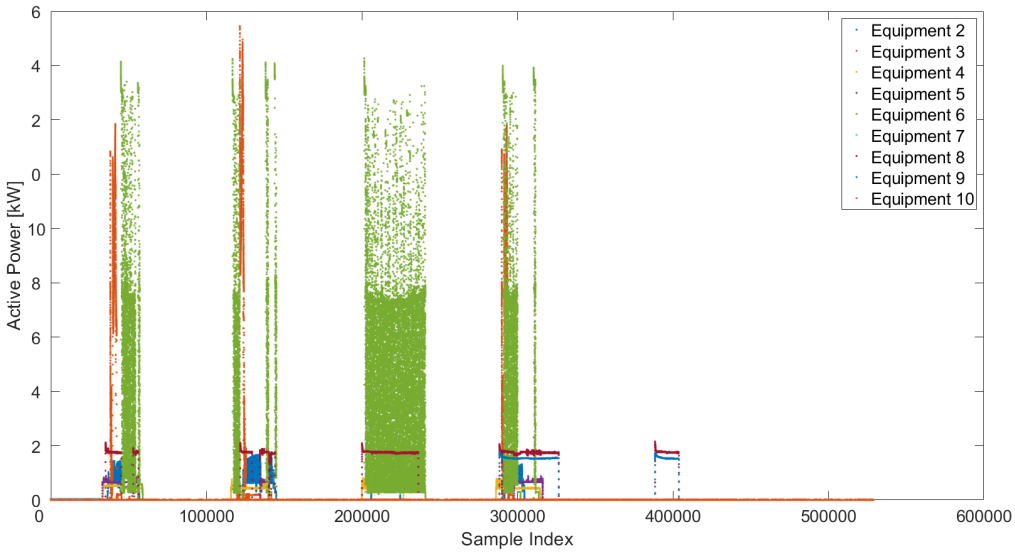

Figure 5.2: Preprocessed equipment active power data for the HIPE dataset in a single plot.

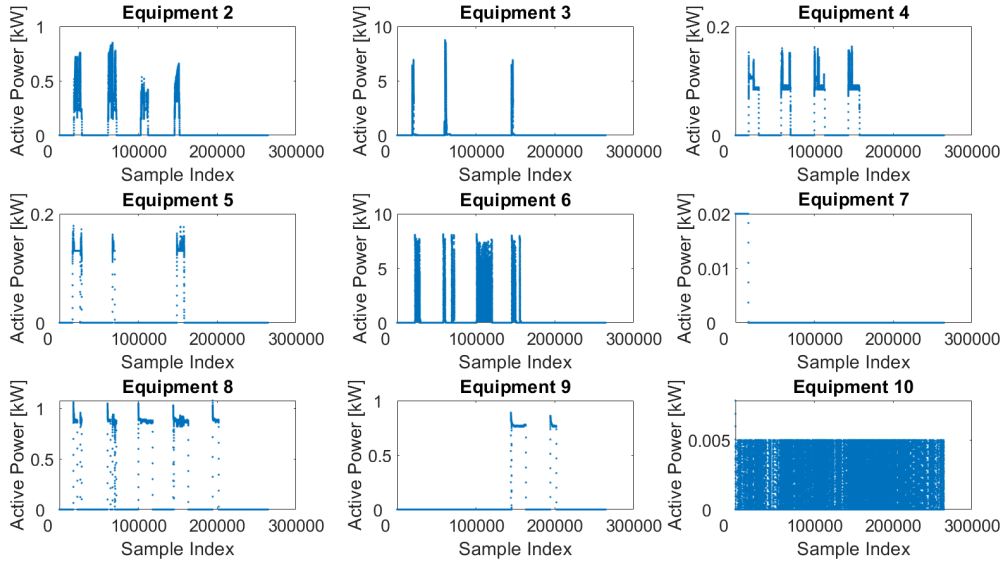

Figure 5.3: Preprocessed equipment active power data for the HIPE dataset, divided into multiple plots.

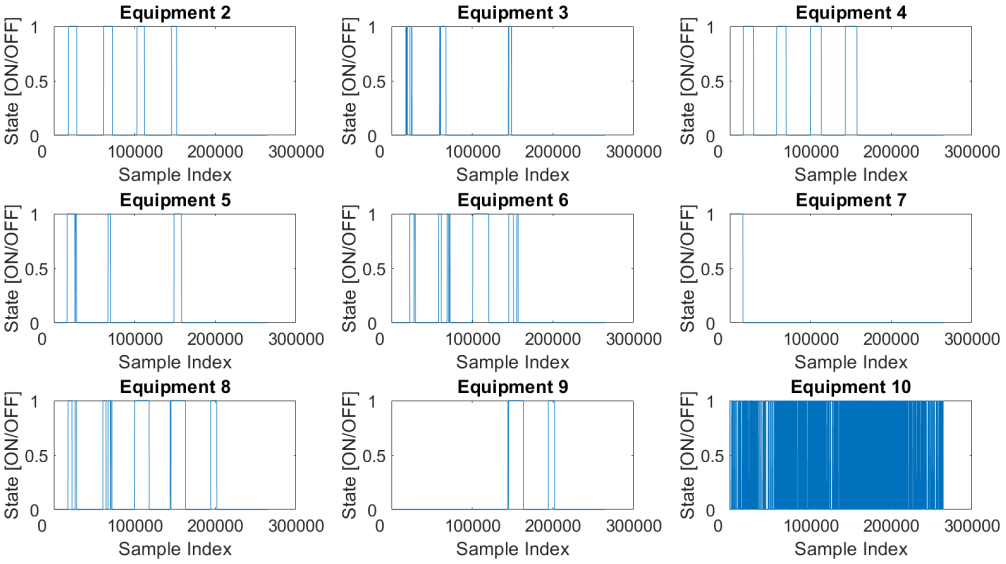

Figure 5.4: Preprocessed equipment states data for the HIPE dataset.

5.1.2 IMDELD Dataset

For the IMDELD dataset, the aggregate active power was not synthetic since the actual active power data sampled on the LVDB-2 meter were used. The training data had 1456 samples, and the testing data had 436 samples. For the IMDELD dataset, the preprocessed aggregate active power, equipment active power and equipment state data are shown in Figures 5.5, 5.6, 5.7 and 5.8.

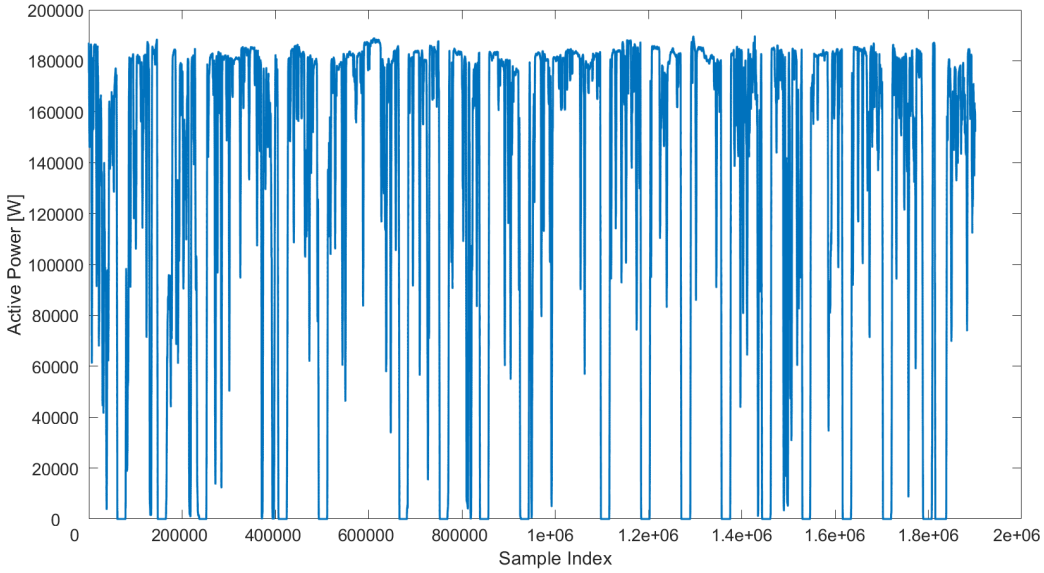

Figure 5.5: Preprocessed aggregate active power for the IMDELD dataset.

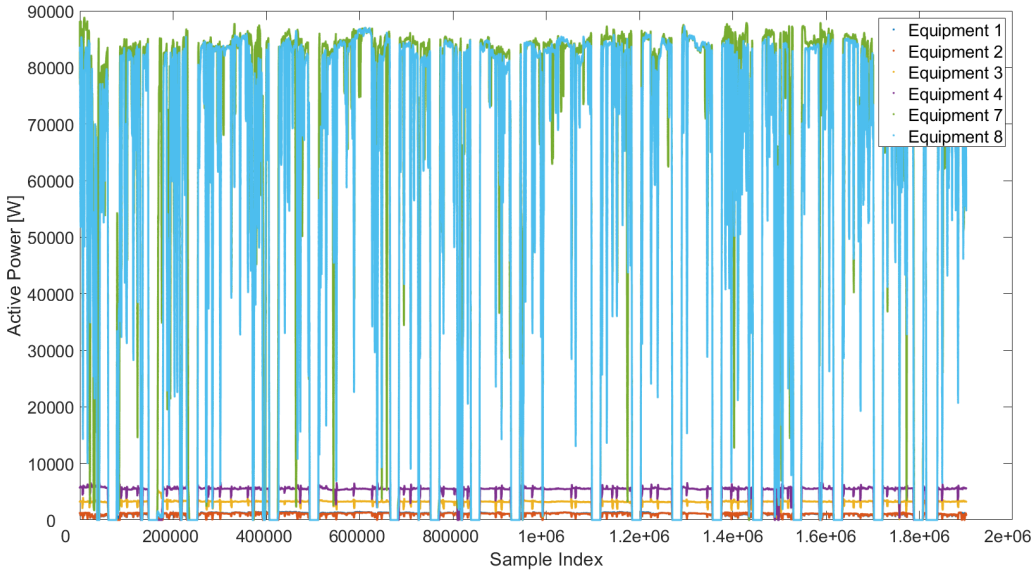

Figure 5.6: Preprocessed equipment active power for the IMDELD dataset in a single plot.

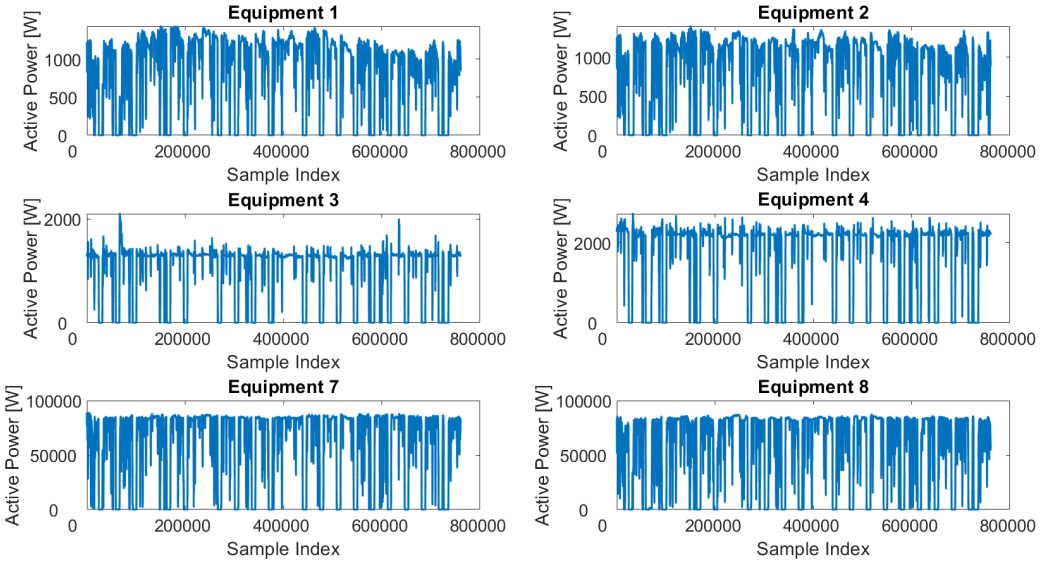

Figure 5.7: Preprocessed equipment active power data for the IMDELD dataset in multiple subplots.

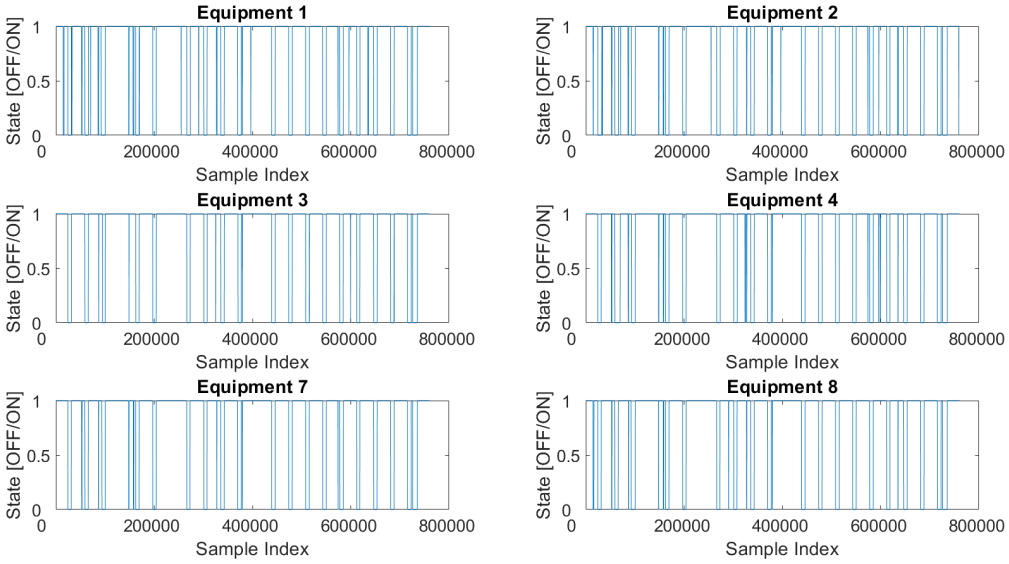

Figure 5.8: Preprocessed equipment states data for the IMDELD dataset.

5.2 Methods Evaluation

The error metrics for the EMUPF and UNN methods estimations were calculated using the HIPE and the IMDELD datasets.

5.2.1 HIPE Dataset Results

The error metrics were calculated for the results of the methods using the testing data from the HIPE dataset. The testing data was composed of eight different aggregate values. The computed MSE and RMSE metrics for the estimated equipment active power for the HIPE dataset are shown in Appendix B.3.

5.2.1.1 MAE Values for the EMUPF Method

The error metrics evaluated the estimated equipment active power values by the EMUPF method. Each row of Table 5.1, corresponds to an aggregate. The first row corresponds to the first aggregate. The next row corresponds to the second aggregate, and so on, until the last row is the eighth aggregate.

Table 5.1: MAE for the equipment active power samples, estimated by the EMUPF method, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.2216	0.4681	-	-	-	-	-	-	-
Agg. 2	0.2297	0.3739	0.1835	-	-	-	-	-	-
Agg. 3	0.2182	0.1159	0.09	0.0862	-	-	-	-	-
Agg. 4	0.7876	0.2106	0.1911	0.1042	1.2737	-	-	-	-
Agg. 5	0.8154	0.4787	0.352	0.0885	1.1512	0.0036	-	-	-
Agg. 6	0.4012	1.5411	1.8773	0.3673	1.043	0.004	14.688	-	-
Agg. 7	0.0776	0.373	2.3436	0.135	0.7346	0.0027	0.5718	417.47	-
Agg. 8	0.3462	0.1213	1.0858	0.1945	1.7303	0.0113	29.743	1.2669	2589.7

The EMUPF method presented large error values, especially for the estimates of active power for equipment with indexes nine and ten. The large error values indicate the method's inability to estimate the equipment's active power values.

The error metrics for the aggregate active power provide a good overview of the accuracy of the estimations. The error was calculated between the true aggregate active power and the sum of the estimated equipment active power values. The error metric values should be small for an accurate estimation.

The EMUPF shows a significantly high value for the error metrics of the aggregate, shown in Table 5.2.

Table 5.2: Error metrics for the aggregate active power samples, estimated by the EMUPF method, for the testing data from the HIPE dataset.

	Agg. 1	**Agg. 2**	**Agg. 3**	**Agg. 4**	**Agg. 5**	**Agg. 6**	**Agg. 7**	**Agg. 8**
MAE	0.4587	0.4715	0.347	1.172	1.3498	17.199	419.36	2620.6
MSE	0.4532	0.455	0.2867	3.2745	6.8268	502.12	1.0012e+06	9.1477e+07
RMSE	0.6732	0.6745	0.5354	1.8096	2.6128	22.408	1000.6	9564.4

The error increases for a larger number of equipment. The error is the largest for the seventh and eighth aggregates, the aggregates that were calculated with the largest number of equipment. The error values are consistent with the fact that the estimated active power values for the equipment with indexes nine and ten presented larger error values.

5.2.1.2 MAE Values for the UNN Method

The error metrics were calculated for the estimated equipment active power values by the UNN method for the HIPE testing data.

The MAE error metric, shown in Tables 5.3 and 5.4 was small, indicating an accurate estimation of the active power values for the equipment in the HIPE dataset. As the number of equipment increases, the error metrics for the equipment estimated values grow slightly.

Table 5.3: MAE for the equipment active power samples, estimated by the UNN method, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.0991	0.1057	-	-	-	-	-	-	-
Agg. 2	0.1155	0.1456	0.1115	-	-	-	-	-	-
Agg. 3	0.1147	0.1097	0.1146	0.0691	-	-	-	-	-
Agg. 4	0.1023	0.0915	0.184	0.096	0.2649	-	-	-	-
Agg. 5	0.139	0.098	0.0827	0.0568	0.2035	0	-	-	-
Agg. 6	0.2161	0.0558	0.0815	0.0669	0.2055	0.0005	0.1486	-	-
Agg. 7	0.1385	0.1428	0.1198	0.1581	0.3066	0.0003	0.2311	0.0917	-
Agg. 8	0.1253	0.0717	0.2529	0.1115	0.3292	0.0007	0.2844	0.0246	0.0405

Table 5.4: Error metrics for the aggregate active power samples, estimated by the UNN method, for the testing data from the HIPE dataset.

	Agg. 1	**Agg. 2**	**Agg. 3**	**Agg. 4**	**Agg. 5**	**Agg. 6**	**Agg. 7**	**Agg. 8**
MAE	0.0165	0.0077	0.0142	0.0325	0.0493	0.052	0.0919	0.0523
MSE	0.0012	0.0001	0.0008	0.0039	0.0067	0.0137	0.0239	0.01
RMSE	0.0346	0.01	0.0283	0.0624	0.0819	0.117	0.1546	0.1

5.2.1.3 MAE Values for the UNN Method with Fourier Mapping

The error metrics were computed for the UNN method with Fourier feature mapping.

The method shows considerable error values, displayed in Tables 5.5 and 5.6. The error metrics for the aggregate values increase with the number of equipment. The error is larger than for the UNN method without Fourier mapping for all the equipment estimations of the active power.

Table 5.5: MAE for the equipment active power samples, estimated by the UNN method with Fourier mapping, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.3953	0.5121	-	-	-	-	-	-	-
Agg. 2	0.7016	0.2296	0.7884	-	-	-	-	-	-
Agg. 3	0.6498	0.4274	1.192	0.468	-	-	-	-	-
Agg. 4	1.6758	0.4113	1.8618	0.7039	1.3533	-	-	-	-
Agg. 5	1.6043	0.291	1.9477	0.6027	1.3728	0.0035	-	-	-
Agg. 6	1.888	0.5761	2.5625	1.0361	1.5513	0.1355	2.2449	-	-
Agg. 7	2.1269	0.6525	2.851	1.2895	1.6288	0.0408	2.6102	1.1827	-
Agg. 8	2.1421	0.5474	2.7403	1.2297	1.5406	0.0918	2.8044	1.1241	0.7301

Table 5.6: Error metrics for the aggregate active power samples, estimated by the UNN method with Fourier mapping, for the testing data from the HIPE dataset.

	Agg. 1	**Agg. 2**	**Agg. 3**	**Agg. 4**	**Agg. 5**	**Agg. 6**	**Agg. 7**	**Agg. 8**
MAE	0.7292	0.7093	2.1402	4.7807	5.5747	8.8285	11.02	12.493
MSE	0.9288	1.147	7.8018	44.87	56.685	140.7	224.02	258.64
RMSE	0.9637	1.071	2.7932	6.6985	7.529	11.864	14.967	16.082

5.2.2 IMDELD Dataset Results

The error metrics were calculated using the testing data from the IMDELD dataset for the methods' results.

5.2.2.1 Error Measures of the EMUPF Method

The error metrics were calculated for the EMUPF method with the IMDELD testing data. The large values of the error metrics, shown in Tables 5.7 and 5.8, indicate the low accuracy in the estimations of the EMUPF method for the active power values of the equipment of the IMDELD dataset. The equipment with the highest expected active power value is the equipment with an index of six, which has the highest error metrics values.

Table 5.7: Error metrics for the equipment active power samples, estimated by the EMUPF method, for the testing data from the IMDELD dataset.

	Eq. 1	**Eq. 2**	**Eq. 3**	**Eq. 4**	**Eq. 7**	**Eq. 8**
MAE	1.8229e+11	4.9231e+10	8.7258e+10	1.5357e+11	1.1658e+10	1.1383e+43
MSE	7.0815e+22	5.2205e+21	1.6148e+22	5.0103e+22	2.8824e+20	2.7525e+86
RMSE	2.6611e+11	7.2253e+10	1.2708e+11	2.2384e+11	1.6978e+10	1.6591e+43

Table 5.8: Error metrics for the aggregate active power samples, estimated by the EMUPF method, for the testing data from the IMDELD dataset.

MAE	**MSE**	**RMSE**
1.1383e+43	2.7525e+86	1.6591e+43

5.2.2.2 Error Measures of the UNN Method

The error metrics were calculated for the UNN method, shown in Tables 5.9 and 5.10.

The UNN method outperforms the EMUPF method, with much smaller error metrics. However, the error values are still significantly higher than those for the UNN method estimations for the HIPE dataset.

Table 5.9: Error metrics for the equipment active power samples, estimated by the UNN method, for the testing data from the IMDELD dataset.

	Eq. 1	**Eq. 2**	**Eq. 3**	**Eq. 4**	**Eq. 7**	**Eq. 8**
MAE	14.4785	0.9077	6.854	3.5599	15.9762	13.7326
MSE	415.4666	1.2667	83.493	20.501	446.4855	397.8221
RMSE	20.383	1.1255	9.1375	4.5278	21.1302	19.9455

Table 5.10: Error metrics for the aggregate active power samples, estimated by the UNN method, for the testing data from the IMDELD dataset.

MAE	MSE	RMSE
0.6403	6.773	2.6025

5.2.2.3 Error Measures of the UNN Method with Fourier Mapping

The error metrics were calculated for the UNN method with Fourier mapping, displayed in Tables 5.11 and 5.12. The UNN method with Fourier mapping has a higher error value than the UNN method without Fourier mapping for the testing data from the IMDELD dataset.

Table 5.11: Error metrics for the equipment active power samples, estimated by the UNN method with Fourier mapping, for the testing data from the IMDELD dataset.

	Eq. 1	Eq. 2	Eq. 3	Eq. 4	Eq. 7	Eq. 8
MAE	77.9464	73.3989	104.6742	70.5946	68.3065	87.9315
MSE	12952.9571	12348.5452	16654.945	11567.7173	8725.065	11602.1169
RMSE	113.8111	111.1240	129.0540	107.5533	93.4081	107.7131

Table 5.12: Error metrics for the aggregate active power samples, estimated by the UNN method with Fourier mapping, for the testing data from the IMDELD dataset.

MAE	MSE	RMSE
111.7	23047	151.81

5.3 Descriptive Statistical Analysis

Statistical variables provide an overview of the accuracy of the methods. The maximum, minimum, median, mean and sum values of the expected and estimated values were calculated. The maximum, minimum, and median measures are shown in Appendix B.4. The mean and sum measures identify the equipment with the highest active power consumption and are highlighted in yellow in this section tables.

5.3.1 HIPE Dataset Results

The expected and estimated values for the HIPE dataset were used to calculate the statistical variables.

5.3.1.1 EMUPF Method Statistical Analysis

The statistical measures were calculated for the equipment active power estimations of the EMUPF method. The EMUPF method estimated inaccurate values for equipment active power values, as indicated by the large difference between the expected and estimated statistical measures, in Table 5.13. The method could not identify the equipment with the highest active power consumption.

Table 5.13: Mean active power values for each equipment, for the expected and estimated values calculated by the EMUPF method, for the HIPE dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0.26725	0.53221	-	-	-	-	-	-	-
	Estimated	0.25937	0.081432	-	-	-	-	-	-	-
Agg. 2	Expected	0.23658	0.45366	0.16928	-	-	-	-	-	-
	Estimated	0.29402	0.10823	-0.01426	-	-	-	-	-	-
Agg. 3	Expected	0.24671	0.38866	0.1657	0.11395	-	-	-	-	-
	Estimated	0.31315	0.40977	0.13564	0.027939	-	-	-	-	-
Agg. 4	Expected	0.15195	0.090656	0.13843	0.066545	1.2488	-	-	-	-
	Estimated	0.86575	-0.093604	0.2271	0.10483	-0.024894	-	-	-	-
Agg. 5	Expected	0.15599	0.076292	0.14208	0.075591	1.2465	0.00030822	-	-	-
	Estimated	0.89797	0.55498	-0.20995	0.056242	0.097145	0.0038869	-	-	-
Agg. 6	Expected	0.15423	0.057363	0.12941	0.073186	1.1333	0.00054795	0.54527	-	-
	Estimated	0.48079	1.5985	2.0034	-0.29411	0.090579	-0.0034123	15.233	-	-
Agg. 7	Expected	0.16086	0.11537	0.13829	0.089687	0.93818	0.00047945	0.56716	0.20751	-
	Estimated	0.095219	-0.25764	2.4769	0.21941	0.20933	0.0031821	0.62836	417.67	-
Agg. 8	Expected	0.15544	0.092312	0.13225	0.085704	0.99808	0.00044521	0.55303	0.20765	0.0010288
	Estimated	-0.19081	0.2007	1.208	0.27069	-0.73226	-0.010815	30.296	1.4746	2589.7

5.3.1.2 UNN Method Statistical Analysis

The statistical measures were calculated for the equipment active power values estimated by the UNN method. The UNN method resulted in accurate values for the mean and sum of the active power values of the equipment. The method correctly identified the equipment with the highest values of active power consumption, shown in yellow in Table 5.14.

Table 5.14: Mean active power values for each equipment, for the expected and estimated values calculated by the UNN method, for the HIPE dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0.26725	0.53221	-	-	-	-	-	-	-
	Estimated	0.27781	0.51899	-	-	-	-	-	-	-
Agg. 2	Expected	0.23658	0.45366	0.16928	-	-	-	-	-	-
	Estimated	0.24946	0.44518	0.16437	-	-	-	-	-	-
Agg. 3	Expected	0.24671	0.38866	0.1657	0.11395	-	-	-	-	-
	Estimated	0.21746	0.4027	0.18102	0.11587	-	-	-	-	-
Agg. 4	Expected	0.15195	0.090656	0.13843	0.066545	1.2488	-	-	-	-
	Estimated	0.066103	0.095859	0.3129	0.16	1.0511	-	-	-	-
Agg. 5	Expected	0.15599	0.076292	0.14208	0.075591	1.2465	0.00030822	-	-	-
	Estimated	0.20413	0.15235	0.13364	0.091822	1.1274	0.00031438	-	-	-
Agg. 6	Expected	0.15423	0.057363	0.12941	0.073186	1.1333	0.00054795	0.54527	-	-
	Estimated	0.27779	0.043898	0.122	0.1401	0.99185	0.0010164	0.52919	-	-
Agg. 7	Expected	0.16086	0.11537	0.13829	0.089687	0.93818	0.00047945	0.56716	0.20751	-
	Estimated	0.27703	0.11292	0.16861	0.2364	0.71815	0.00079349	0.62767	0.13643	-
Agg. 8	Expected	0.15544	0.092312	0.13225	0.085704	0.99808	0.00044521	0.55303	0.20765	0.0010288
	Estimated	0.25562	0.091772	0.37377	0.19711	0.7605	0.00094469	0.27933	0.22993	0.040142

5.3.1.3 UNN Method with Fourier Mapping Statistical Analysis

The statistical measures were calculated for the equipment active power values estimated by the UNN method with Fourier mapping. The UNN method with Fourier mapping presented worse results than the UNN method without Fourier mapping, which is

in accordance with the error metrics calculated previously. The method failed to identify high-power-consuming equipment. The mean statistical measures are shown in Table 5.15.

Table 5.15: Mean active power values for each equipment, for the expected and estimated values calculated by the UNN method with Fourier mapping, for the HIPE dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0.26725	0.53221	-	-	-	-	-	-	-
	Estimated	0.36388	0.29839	-	-	-	-	-	-	-
Agg. 2	Expected	0.23658	0.45366	0.16928	-	-	-	-	-	-
	Estimated	0.87764	0.52853	-0.19188	-	-	-	-	-	-
Agg. 3	Expected	0.24671	0.38866	0.1657	0.11395	-	-	-	-	-
	Estimated	0.88279	0.58662	1.3491	0.19587	-	-	-	-	-
Agg. 4	Expected	0.15195	0.090656	0.13843	0.066545	1.2488	-	-	-	-
	Estimated	0.72788	0.10651	1.4494	0.74717	2.5621	-	-	-	-
Agg. 5	Expected	0.15599	0.076292	0.14208	0.075591	1.2465	0.00030822	-	-	-
	Estimated	1.7111	0.30063	2.0511	0.48857	2.6193	0.0038127	-	-	-
Agg. 6	Expected	0.15423	0.057363	0.12941	0.073186	1.1333	0.00054795	0.54527	-	-
	Estimated	1.578	0.38572	1.0058	1.1092	2.6846	0.13608	2.4479	-	-
Agg. 7	Expected	0.16086	0.11537	0.13829	0.089687	0.93818	0.00047945	0.56716	0.20751	-
	Estimated	2.2877	0.16125	2.7316	1.3792	2.213	-0.040339	2.7806	0.84668	-
Agg. 8	Expected	0.15544	0.092312	0.13225	0.085704	0.99808	0.00044521	0.55303	0.20765	0.0010288
	Estimated	1.8493	0.56248	2.6314	1.3154	2.2591	0.092278	3.3574	1.3317	0.7311

5.3.2 IMDELD Dataset Results

The expected and estimated values for the IMDELD dataset were used to calculate the statistical measures.

5.3.2.1 EMUPF Method Statistical Analysis

The statistical measures were calculated for the equipment's active power values estimated by the EMUPF method. For the IMDELD dataset, the EMUPF method is unable to identify the equipment with the highest active power consumption, and the statistical values proved the inaccuracy of the method, displayed in Table 5.16.

Table 5.16: Mean and sum active power values for each equipment, for the expected and estimated values calculated by the EMUPF method, for the IMDELD dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 1	Eq. 2	Eq. 3	Eq. 4	Eq. 7	Eq. 8
mean	Expected	711.88	692.52	2523.1	4211.5	54434	48396
	Estimated	1.8229e+14	4.9231e+13	-8.7258e+13	-1.5357e+14	-1.1658e+13	1.1383e+46
sum	Expected	3.1038e+05	3.0194e+05	1.1001e+06	1.8362e+06	2.3733e+07	2.1101e+07
	Estimated	7.948e+16	2.1465e+16	-3.8044e+16	-6.6956e+16	-5.0828e+15	4.9628e+48

5.3.2.2 UNN Method Statistical Analysis

The statistical measures were calculated for the equipment active power values estimated by the UNN method. The UNN method exhibits higher accuracy than the EMUPF method, shown in Table 5.17. Still, it presents some inaccuracies in identifying

the equipment with the highest active power consumption, for the data from the IMDELD dataset.

Table 5.17: Mean and sum active power values for each equipment, for the expected and estimated values calculated by the UNN method, for the IMDELD dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 1	**Eq. 2**	**Eq. 3**	**Eq. 4**	**Eq. 7**	**Eq. 8**
mean	Expected	711.88	692.52	2523.1	4211.5	54434	48396
	Estimated	14886	333.14	7781.9	4864.6	42760	54794
sum	Expected	3.1038e+05	3.0194e+05	1.1001e+06	1.8362e+06	2.3733e+07	2.1101e+07
	Estimated	6.4902e+06	1.4525e+05	3.3929e+06	2.121e+06	1.8643e+07	2.389e+07

5.3.2.3 UNN Method with Fourier Mapping Statistical Analysis

The statistical measures were calculated for the estimated equipment active power values by the UNN method with Fourier mapping. The UNN method with Fourier presented better results than the EMUPF method but worse than the UNN method without Fourier mapping. The method was unable to identify the equipment with the highest active power consumption values, as shown in Figure 5.18.

Table 5.18: Mean and sum active power values for each equipment, for the expected and estimated values calculated by the UNN method with Fourier mapping, for the IMDELD dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 1	**Eq. 2**	**Eq. 3**	**Eq. 4**	**Eq. 7**	**Eq. 8**
mean	Expected	711.88	692.52	2523.1	4211.5	54434	48396
	Estimated	1806.9	-66417	80943	72866	7229.2	1.3325e+05
sum	Expected	844.59	835.01	3227.4	5510	66336	60229
	Estimated	7.8783e+05	-2.8958e+07	3.5291e+07	3.177e+07	3.1519e+06	5.8098e+07

5.4 Chapter Summary

The estimates by the EMUPF method had the highest error metrics values for the HIPE and IMDELD datasets. The EMUPF technique could not identify the equipment with the highest active power consumption values. The UNN method showed the lowest error metrics values and the most accurate descriptive statistical measures. The UNN method accurately estimated values for the HIPE dataset but not for the IMDELD dataset. The UNN method with Fourier mapping had lower error metrics values than the EMUPF method but presented higher values than the UNN method without Fourier mapping. The statistical measures showed an inaccuracy in the equipment active power consumption estimations by the UNN method with Fourier mapping.

Chapter 6

Discussion and Conclusion

The chapter discusses the attained results by analysing the metrics from the proposed methods. Additionally, the chapter presents relevant considerations and final remarks and addresses the possible directions to expand on the developed work.

6.1 Discussion

6.1.1 Results

The EMUPF method was ineffective in estimating the active power of the equipment and had the highest MAE, MSE and RMSE values. The UNN method achieved the lowest MAE, MSE and RMSE values for the HIPE and IMDELD datasets, compared to EMUPF and the UNN with Fourier mapping methods. The UNN method accurately estimated the active power consumption of the equipment. The UNN method proved to be a viable unsupervised NILM technique.

6.1.2 General Considerations

A universal solution with perfect results for the unsupervised NILM problem seems improbable due to the complex objective function with multiple local optima, saddle points and the inherent noise in the samples. Therefore, the best outcome is an approximation that reasonably estimates the equipment behaviour. The two developed methods attempt to approximate the active power consumption of the equipment in an industrial facility, with the UNN providing the most accurate estimates. At the same time, the EMUPF does not meet the scalability requirements. The EMUPF technique suffers from the curse of dimensionality. The higher dimensionality of the problem, by increasing the number of variables in the objective function, increases the size and complexity of the search space and causes an explosion in the computation operations required to optimize the problem. The EMUPF method is not feasible for a considerable number of equipment. The UNN can handle numerous dimensions by changing the architecture to have hundreds of neurons. The UNN algorithm meets the scalability requirement and is suitable for a NILM problem with industrial loads. However, the results show that an increase in the number of equipment in the problem affects the output since the error grows.

6.1.3 Method Considerations

The results of the methods are highly dependent on the data. Equipment seven and eight present much higher active power values than the other equipment in the IMDELD dataset. The high error values are due in part to the large differences in the values of active power in the equipment. The objective function of the unsupervised methods does not consider each equipment's maximum active power values, as there is no prior knowledge of the maximum active power. The IMDELD data makes it challenging to have a reasonable estimate. Another important factor that affects the accuracy of the method results is the small set of training samples. A larger set of training samples could improve the model and the accuracy of the results.

The UNN requires calibration of the model's hyperparameters, architecture and loss function for each NILM problem. Calibration is challenging, requires vast experimentation, and significantly impacts the results. The network architecture used for estimating with the HIPE and IMDELD datasets was the same. The results showed that no single architecture could provide good results for both datasets. The results for the IMDELD dataset could be improved through calibration. Fourier mapping did not improve the results of the UNN method.

6.2 Conclusion

Two novel algorithms that perform low-frequency unsupervised NILM for industrial loads were developed and studied. The algorithms were trained and validated using two public datasets. A survey was conducted on public NILM datasets, and the HIPE and IMDELD datasets were selected. The datasets required preprocessing to clean the aggregate and equipment active power and to calculate the equipment state data. The preprocessing included dividing the data into training and testing data. The EMUPF algorithm was rejected due to the high values of the error metrics of the results and the fact that it is not scalable for a large number of equipment. The UNN method proved to be a viable solution to the NILM problem, with low values for error metrics and accurately identifying the equipment with the highest active power consumption. The UNN algorithm is compatible with type III equipment and is not event-based. The UNN method requires that calibration be performed for each specific problem. The analysis of the results for the IMDELD dataset proved that estimating the equipment active power values from an aggregate can be difficult when there is a significant difference in the active power values of the equipment. The objectives were achieved by creating the first algorithm that performs unsupervised low-frequency NILM for industrial loads.

6.3 Future Work

Exploring different and complex architectures with varying layers, neurons, inputs, step sizes, activation, and loss functions is necessary to improve the UNN method. A comprehensive study must investigate the architecture and hyperparameters' impacts on various datasets. Applications of Fourier mapping should be investigated for different architectures. The UNN method should be trained and validated with a larger dataset. An analysis is required to design a user interface. The user interface should take user input, allow the user to define goals, suggest energy-saving measures, and provide feedback

on the results of the ongoing measures. A study should be conducted to analyze the applications and benefits of the UNN algorithm in a real-world scenario. BuggyPower's production plant should use the UNN method, and the algorithm's performance and outcomes on the facility's electricity consumption should be examined.

Bibliography

[1] J. Leiria, R. Salles, J. Mendes, P. Sousa, Soft sensors for industrial applications: Comparison of variables selection methods and regression models, in: 2023 International Conference on Control, Automation and Diagnosis (ICCAD), 2023, pp. 1–6. doi:10.1109/ICCAD57653.2023.10152323.

[2] M. Gonçalves, P. Sousa, J. Mendes, M. Danishvar, A. Mousavi, Real-time event-driven learning in highly volatile systems: A case for embedded machine learning for scada systems, IEEE Access 10 (2022) 50794–50806. doi:10.1109/ACCESS.2022.3173376.

[3] P. Huber, A. Calatroni, A. Rumsch, A. Paice, Review on deep neural networks applied to low-frequency nilm, Energies 14 (9) (2021). doi:10.3390/en14092390.

[4] Q. Liu, K. M. Kamoto, X. Liu, M. Sun, N. Linge, Low-complexity non-intrusive load monitoring using unsupervised learning and generalized appliance models, IEEE Transactions on Consumer Electronics 65 (2019) 28–37. doi:10.1109/TCE.2019.2891160.

[5] G. W. Hart, Nonintrusive appliance load monitoring, Proceedings of the IEEE 80 (12) (1992) 1870–1891.

[6] K. S. Barsim, R. Streubel, B. Yang, An approach for unsupervised non-intrusive load monitoring of residential appliances, Proceedings of the 48th International Universities' Power Engineering Conference (UPEC) (2013).

[7] J. Y. Leung, B. D. Russell, S. D. Connell, Ar5 synthesis report: Climate change 2014, One Earth 1 (2019).

[8] E. A. Abdelaziz, R. Saidur, S. Mekhilef, A review on energy saving strategies in industrial sector, Renewable and Sustainable Energy Reviews 15 (2011). doi:10.1016/j.rser.2010.09.003.

[9] S. Sorrell, Reducing energy demand: A review of issues, challenges and approaches, Renewable and Sustainable Energy Reviews 47 (2015). doi:10.1016/j.rser.2015.03.002.

[10] IEA, Global energy review 2021 – analysis - iea, International Energy Agency (2021).

[11] IEA, Iea energy and carbon tracker 2020, https://www.iea.org/data-and-statistics/data-product/iea-energy-and-carbon-tracker-2020, accessed: 10.09.2022 (2020).

[12] IEA, Statistics report: Key world energy statistics 2021 (2021).

[13] IEA, Sustainable development scenario – world energy model – analysis - iea, https://www.iea.org/reports/world-energy-model/sustainable-development-scenario (2020).

[14] N. Campbell, C. Forbes, L. Ryan, Spreading the net: Evaluating the multiple benefits delivered by energy efficiency policy, 2012 International Energy Program Evaluation Conference (2012).

[15] E. Worrell, J. A. Laitner, M. Ruth, H. Finman, Productivity benefits of industrial energy efficiency measures, Energy 28 (2003). doi:10.1016/S0360-5442(03)00091-4.

[16] D. Vine, L. Buys, P. Morris, The effectiveness of energy feedback for conservation and peak demand: A literature review, Open Journal of Energy Efficiency 02 (2013). doi:10.4236/ojee.2013.21002.

[17] E. de Monaco, Porto santo, buggypower, repas d'algues : Jour 15, https://www.monacoexplorations.org/porto-santo-buggypower-repas-dalgues-jour-15 (2023).

[18] S. Rastegar, R. Araújo, M. Malekzadeh, Álvaro Gomes, H. Jorge, A new nialm system design based on neural network architecture and adaptive springy particle swarm optimization algorithm, Energy Efficiency 16 (6) (2023) 52. doi:10.1007/s12053-023-10125-5.

[19] A. Faustine, N. H. Mvungi, S. Kaijage, K. Michael, A survey on non-intrusive load monitoring methodologies and techniques for energy disaggregation problem (2017). arXiv:1703.00785.

[20] Hernández, A. Ruano, J. Ureña, M. G. Ruano, J. J. Garcia, Applications of nilm techniques to energy management and assisted living, IFAC-PapersOnLine 52 (2019). doi:10.1016/j.ifacol.2019.09.135.

[21] F. Kalinke, P. Bielski, S. Singh, E. Fouché, K. Böhm, An evaluation of nilm approaches on industrial energy-consumption data, e-Energy 2021 - Proceedings of the 2021 12th ACM International Conference on Future Energy Systems (2021). doi:10.1145/3447555.3464863.

[22] E. J. Aladesanmi, K. A. Folly, Overview of non-intrusive load monitoring and identification techniques, IFAC-PapersOnLine 48 (2015). doi:10.1016/j.ifacol.2015.12.414.

[23] M. Zhuang, M. Shahidehpour, Z. Li, An overview of non-intrusive load monitoring: Approaches, business applications, and challenges, 2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings (2019) 4291–4299doi:10.1109/POWERCON.2018.8601534.

[24] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, J. Garcia, Nilm techniques for intelligent home energy management and ambient assisted living: A review, Energies 12 (2019). doi:10.3390/en12112203.

[25] G. Yadav, K. Paul, Architecture and security of scada systems: A review, International Journal of Critical Infrastructure Protection 34 (2021). doi:10.1016/j.ijcip.2021.100433.

[26] A. Daneels, W. Salter, What is scada ?, International Conference on Accelerator and Large Experimental Physics Control Systems, Trieste, Italy (1999).

[27] Y. Cherdantseva, P. Burnap, A. Blyth, P. Eden, K. Jones, H. Soulsby, K. Stoddart, A review of cyber security risk assessment methods for scada systems, Computers and Security 56 (2016). doi:10.1016/j.cose.2015.09.009.

[28] H. K. Iqbal, F. H. Malik, A. Muhammad, M. A. Qureshi, M. N. Abbasi, A. R. Chishti, A critical review of state-of-the-art non-intrusive load monitoring datasets, Electric Power Systems Research 192 (3 2021). doi:10.1016/j.epsr.2020.106921.

[29] C. Gisler, A. Ridi, D. Zujferey, O. A. Khaled, J. Hennebert, Appliance consumption signature database and recognition test protocols, 2013 8th International Workshop on Systems, Signal Processing and Their Applications, WoSSPA 2013 (2013). doi:10.1109/WoSSPA.2013.6602387.

[30] A. Ridi, C. Gisler, J. Hennebert, Acs-f2 - a new database of appliance consumption signatures, 6th International Conference on Soft Computing and Pattern Recognition, SoCPaR 2014 (2014). doi:10.1109/SOCPAR.2014.7007996.

[31] N. Buneeva, A. Reinhardt, Ambal: Realistic load signature generation for load disaggregation performance evaluation, 2017 IEEE International Conference on Smart Grid Communications, SmartGridComm 2017 2018-January (2018). doi:10.1109/SmartGridComm.2017.8340657.

[32] S. Makonin, B. Ellert, I. V. Bajić, F. Popowich, Electricity, water, and natural gas consumption of a residential house in canada from 2012 to 2014, Scientific Data 3 (2016). doi:10.1038/sdata.2016.37.

[33] M. Maasoumy, B. M. Sanandaji, K. Poolla, A. S. Vincentelli, Berds - berkeley energy disaggregation data set, Proceedings of the Workshop on Big Learning at the Conference on Neural Information Processing Systems (NIPS) (2013).

[34] T. Kriechbaumer, H. A. Jacobsen, Blond, a building-level office environment dataset of typical electrical appliances, Scientific Data 5 (2018). doi:10.1038/sdata.2018.48.

[35] K. Anderson, A. F. Ocneanu, D. Benitez, D. Carlson, A. Rowe, M. Bergés, Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research, Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD) (2012).

[36] N. Batra, O. Parson, M. Berges, A. Singh, A. Rogers, A comparison of non-intrusive load monitoring methods for commercial and residential buildings (2014). doi:10.48550/ARXIV.1408.6595.
URL https://arxiv.org/abs/1408.6595

[37] T. Picon, M. N. Meziane, P. Ravier, G. Lamarque, C. Novello, J.-C. L. Bunetel, Y. Raingeaud, Cooll: Controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification (2016). doi:10.48550/ARXIV.1611.05803.
URL https://arxiv.org/abs/1611.05803

[38] M. Pipattanasomporn, G. Chitalia, J. Songsiri, C. Aswakul, W. Pora, S. Suwankawin, K. Audomvongseree, N. Hoonchareon, Cu-bems, smart building electricity consumption and indoor environmental sensor datasets, Scientific Data 7 (2020). doi:10.1038/s41597-020-00582-3.

[39] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, A. Rogers, Dataport and nilmtk: A building data set designed for non-intrusive load monitoring, 2015 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2015 (2016). doi:10.1109/GlobalSIP.2015.7418187.

[40] S. N. A. U. Nambi, A. R. Lua, R. V. Prasad, Loced: Location-aware energy disaggregation framework, BuildSys 2015 - Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built (2015) 45–54doi:10.1145/2821650.2821659.

[41] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, S. Santini, The eco data set and the performance of non-intrusive load monitoring algorithms, BuildSys 2014 - Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (2014). doi:10.1145/2674061.2674064.

[42] G. Johnson, I. Beausoleil-Morrison, Electrical-end-use data from 23 houses sampled each minute for simulating micro-generation systems, Applied Thermal Engineering 114 (2017). doi:10.1016/j.applthermaleng.2016.07.133.

[43] C. Shin, E. Lee, J. Han, J. Yim, W. Rhee, H. Lee, The enertalk dataset, 15 hz electricity consumption data from 22 houses in korea, Scientific Data 6 (2019). doi:10.1038/s41597-019-0212-5.

[44] H. Xu, L. König, D. Cáliz, H. Schmeck, A generic user interface for energy management in smart homes, Energy Informatics 1 (2018). doi:10.1186/s42162-018-0060-0.

[45] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, A. M. Tonello, Greend: An energy consumption dataset of households in italy and austria, 2014 IEEE International Conference on Smart Grid Communications, SmartGridComm 2014 (2015). doi:10.1109/SmartGridComm.2014.7007698.

[46] P. Held, S. Mauch, A. Saleh, D. Benyoucef, Held1 : Home equipment laboratory dataset for non-intrusive load monitoring, The Third International Conference on Advances in Signal, Image and Video Processing (Signal 2018) (2018).

[47] M. Gulati, S. S. Ram, A. Singh, An in depth study into using emi signatures for appliance identification, BuildSys 2014 - Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (2014). doi:10.1145/2674061.2674070.

[48] S. Bischof, H. Trittenbach, M. Vollmer, D. Werle, T. Blank, K. Böhm, Hipe – an energy-status-data set from industrial production, e-Energy 2018 - Proceedings of the 9th ACM International Conference on Future Energy Systems (2018). doi:10.1145/3208903.3210278.

[49] J.-P. Zimmermann, M. Evans, T. Lineham, J. Griggs, G. Surveys, L. Harding, N. King, P. Roberts, Household electricity survey: A study of domestic electrical product usage, Intertek (2012).

[50] S. Makonin, Hue: The hourly usage of energy dataset for buildings in british columbia, Data in Brief 23 (2019). doi:10.1016/j.dib.2019.103744.

[51] N. Batra, M. Gulati, A. Singh, M. B. Srivastava, It's different: Insights into home energy consumption in india, BuildSys 2013 - Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings (2013). doi:10.1145/2528282.2528293.

[52] M. Pullinger, J. Kilgour, N. Goddard, N. Berliner, L. Webb, M. Dzikovska, H. Lovell, J. Mann, C. Sutton, J. Webb, M. Zhong, The ideal household energy dataset, electricity, gas, contextual sensor data and survey data for 255 uk homes, Scientific Data 8 (2021). doi:10.1038/s41597-021-00921-y.

[53] G. Hebrail, A. Barard, Individual household electric power consumption data set, UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science 1 (2012).

[54] P. Bandeira de Mello Martins, V. Barbosa Nascimento, A. R. de Freitas, P. Bittencourt e Silva, R. Guimarães Duarte Pinto, Industrial machines dataset for electrical load disaggregation (2018). doi:10.21227/cg5v-dk02.

[55] H. Rashid, P. Singh, A. Singh, Data descriptor: I-blend, a campus-scale commercial and residential buildings electrical energy dataset, Scientific Data 6 (2019). doi:10.1038/sdata.2019.15.

[56] L. Yan, J. Han, R. Xu, Z. Li, Lifted: Household appliance-level load dataset and data compression with lossless coding considering precision, IEEE Power and Energy Society General Meeting 2020-August (2020). doi:10.1109/PESGM41954.2020.9282138.

[57] M. Kahl, V. Krause, R. Hackenberg, A. U. Haq, A. Horn, H. A. Jacobsen, T. Kriechbaumer, M. Petzenhauser, M. Shamonin, A. Udalzow, Measurement system and dataset for in-depth analysis of appliance energy consumption in industrial environment, Technisches Messen 86 (2019). doi:10.1515/teme-2018-0038.

[58] B. Kalluri, S. Kondepudi, K. H. Wei, T. K. Wai, A. Kamilaris, Opld: Towards improved non-intrusive office plug load disaggregation, 2015 IEEE International Conference on Building Energy Efficiency and Sustainable Technologies, ICBEST 2015 (2016). doi:10.1109/ICBEST.2015.7435865.

[59] J. Gao, S. Giri, E. C. Kara, M. Bergés, Plaid: A public dataset of high-resoultion electrical appliance measurements for load identification research: Demo abstract, Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (2014) 198–199doi:10.1145/2674061.2675032.

[60] L. D. Baets, C. Develder, T. Dhaene, D. Deschrijver, J. Gao, M. Berges, Handling imbalance in an extended plaid, 5th IFIP Conference on Sustainable Internet and ICT for Sustainability, SustainIT 2017 (2018). doi:10.23919/SustainIT.2017.8379795.

[61] R. Medico, L. D. Baets, J. Gao, S. Giri, E. Kara, T. Dhaene, C. Develder, M. Bergés, D. Deschrijver, A voltage and current measurement dataset for plug load appliance identification in households, Scientific Data 7 (2020). doi:10.1038/s41597-020-0389-7.

[62] S. Makonin, Z. J. Wang, C. Tumpach, Rae: The rainforest automation energy dataset for smart grid meter data analysis, Data 3 (2018). doi:10.3390/data3010008.

[63] B. Larson, L. Gilman, R. Davis, M. Logsdon, J. Uslan, B. Hannas, D. Baylon, P. Storm, V. Mugford, N. Kvaltine, Residential building stock assessment: Metering study, Northwest Energy Efficiency Alliance (2014).

[64] J. Z. Kolter, M. J. Johnson, Redd: A public data set for energy disaggregation research, SustKDD workshop (2011).

[65] D. Murray, L. Stankovic, V. Stankovic, An electrical load measurements dataset of united kingdom households from a two-year longitudinal study, Scientific Data 4 (2017). doi:10.1038/sdata.2016.122.

[66] S. Henriet, U. Şimşekli, B. Fuentes, G. Richard, A generative model for non-intrusive load monitoring in commercial buildings, Energy and Buildings 177 (2018). doi:10.1016/j.enbuild.2018.07.060.

[67] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, J. Albrecht, Smart*: An open data set and tools for enabling research in sustainable homes, SustKDD (2012).

[68] D. Chen, D. Irwin, P. Shenoy, Smartsim: A device-accurate smart home simulator for energy analytics, 2016 IEEE International Conference on Smart Grid Communications, SmartGridComm 2016 (2016). doi:10.1109/SmartGridComm.2016.7778841.

[69] E. Lee, K. Baek, J. Kim, Datasets on south korean manufacturing factories' electricity consumption and demand response participation, Sci Data 9, 227 (2022). doi:10.1038/s41597-022-01357-8.

[70] L. Pereira, F. Quintal, R. Gonçalves, N. J. Nunes, Sustdata: A public dataset for ict4s electric energy research, ICT for Sustainability 2014, ICT4S 2014 (2014). doi:10.2991/ict4s-14.2014.44.

[71] M. Ribeiro, L. Pereira, F. Quintal, N. Nunes, Sustdataed: A public dataset for electric energy disaggregation research, ICT for Sustainability 2016 (2016). doi:10.2991/ict4s-16.2016.36.

[72] C. Klemenjak, C. Kovatsch, M. Herold, W. Elmenreich, A synthetic energy dataset for non-intrusive load monitoring in households, Scientific Data 7 (2020). doi:10.1038/s41597-020-0434-6.

[73] J. Valdes, L. R. Camargo, Synthetic hourly electricity load data for the paper and food industries, Data in Brief 35 (2021). doi:10.1016/j.dib.2021.106903.

[74] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, R. Steinmetz, On the accuracy of appliance identification based on distributed load metering data, 2012 Sustainable Internet and ICT for Sustainability, SustainIT 2012 (2012).

[75] J. Kelly, W. Knottenbelt, The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes, Scientific Data 2 (2015). doi:10.1038/sdata.2015.7.

[76] M. Kahl, A. U. Haq, T. Kriechbaumer, H. arno Jacobsen, Whited - a worldwide household and industry transient energy data set, 3rd International Workshop on Non-Intrusive Load Monitoring (NILM2016) (2016).

[77] E. W. Weisstein, Polynomial., https://mathworld.wolfram.com/Polynomial.html (2023).

[78] N. Fumo, M. A. R. Biswas, Regression analysis for prediction of residential energy consumption, Renewable and Sustainable Energy Reviews 47 (2015). doi:10.1016/j.rser.2015.03.035.

[79] T. Hong, S. Fan, Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting 32 (2016). doi:10.1016/j.ijforecast.2015.11.011.

[80] J. Nocedal, S. J. Wright, Numerical optimization, Springer Series in Operations Research and Financial Engineering (2006). doi:10.1201/b19115-11.

[81] M. Abdel-Basset, L. Abdel-Fatah, A. K. Sangaiah, Metaheuristic algorithms: A comprehensive review, Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications (2018). doi:10.1016/B978-0-12-813314-9.00010-4.

[82] J. Pereira, J. Mendes, J. S. S. Júnior, C. Viegas, J. R. Paulo, Wildfire spread prediction model calibration using metaheuristic algorithms, in: IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society, 2022, pp. 1–6. doi:10.1109/IECON49645.2022.9968435.

[83] D. Bertsimas, J. Tsitsiklis, Simulated annealing, Statistical science 8 (1) (1993) 10–15.

[84] R. Salles, J. Mendes, C. Henggeler Antunes, P. Moura, J. Dias, Dynamic setpoint optimization using metaheuristic algorithms for wastewater treatment plants, in: IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society, 2022, pp. 1–6. doi:10.1109/IECON49645.2022.9968617.

[85] R. Maia, J. Mendes, R. Araújo, Electric vehicle physical parameters identification, in: IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society, 2022, pp. 1–7. doi:10.1109/IECON49645.2022.9968543.

[86] J. Mendes, R. Seco, R. Araújo, Automatic extraction of the fuzzy control system for industrial processes, in: ETFA2011, 2011, pp. 1–8. doi:10.1109/ETFA.2011.6059063.

[87] J. Kennedy, R. Eberhart, Particle swarm optimization, Proceedings of ICNN'95 - International Conference on Neural Networks 4 (1995) 1942–1948 vol.4. doi:10.1109/ICNN.1995.488968.

[88] L. Laím, J. Mendes, H. D. Craveiro, A. Santiago, C. Melo, Structural optimization of closed built-up cold-formed steel columns, Journal of Constructional Steel Research 193 (2022) 107266. doi:10.1016/j.jcsr.2022.107266.

[89] A. E. Ezugwu, O. J. Adeleke, A. A. Akinyelu, S. Viriri, A conceptual comparison of several metaheuristic algorithms on continuous optimisation problems, Neural Computing and Applications 32 (2020) 6207–6251.

[90] J. D. Kelleher, Deep Learning, MIT PRESS, 2020. doi:10.7551/mitpress/11171.003.0006.

[91] S. Bayati, F. Jabbarvaziri, Learning to optimize under constraints with unsupervised deep neural networks (2021). arXiv:2101.00744.

[92] G. P. Tolstov, Fourier series, Courier Corporation, 2012.

[93] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, R. Ng, Fourier features let networks learn high frequency functions in low dimensional domains (2020). arXiv:2006.10739.

[94] H. Snyder, Literature review as a research methodology: An overview and guidelines, Journal of Business Research 104 (2019). doi:10.1016/j.jbusres.2019.07.039.

[95] B. de Matos, R. Salles, J. Mendes, J. R. Gouveia, A. J. Baptista, P. Moura, A review of energy and sustainability kpi-based monitoring and control methodologies on wwtps, Mathematics 11 (1) (2023). doi:10.3390/math11010173.

[96] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, J. Garcia, Nilm techniques for intelligent home energy management and ambient assisted living: A review, Energies 12 (6 2019). doi:10.3390/en12112203.

[97] R. Gopinath, M. Kumar, C. P. C. Joshua, K. Srinivas, Energy management using non-intrusive load monitoring techniques – state-of-the-art and future research directions, Sustainable Cities and Society 62 (2020). doi:10.1016/j.scs.2020.102411.

[98] G. F. Angelis, C. Timplalexis, S. Krinidis, D. Ioannidis, D. Tzovaras, Nilm applications: Literature review of learning approaches, recent developments and challenges, Energy and Buildings 261 (2022). doi:10.1016/j.enbuild.2022.111951.

[99] A. Zoha, A. Gluhak, M. A. Imran, S. Rajasegarar, Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey, Sensors (Switzerland) 12 (2012). doi:10.3390/s121216838.

[100] J. S. S. Júnior, J. Pãulo, J. Mendes, D. Alves, L. M. Ribeiro, Automatic calibration of forest fire weather index for independent customizable regions based on historical records, in: 2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 2020, pp. 1–8. doi:10.1109/AIKE48582.2020.00011.

[101] J. Mendes, S. Pinto, R. Araújo, F. Souza, Evolutionary fuzzy models for nonlinear identification, in: Proceedings of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA 2012), 2012, pp. 1–8. doi:10.1109/ETFA.2012.6489621.

[102] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, C. Develder, Comprehensive feature selection for appliance classification in nilm, Energy and Buildings 151 (2017) 98–106. doi:10.1016/j.enbuild.2017.06.042.

[103] P. A. Schirmer, I. Mporas, Non-intrusive load monitoring: A review, IEEE Transactions on Smart Grid 14 (1) (2023) 769–784. doi:10.1109/TSG.2022.3189598.

[104] Z. Wang, G. Zheng, Residential appliances identification and monitoring by a nonintrusive method, IEEE Transactions on Smart Grid 3 (1) (2012) 80–92. doi:10.1109/TSG.2011.2163950.

[105] R. Bonfigli, S. Squartini, M. Fagiani, F. Piazza, Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview, 2015 IEEE 15th International Conference on Environment and Electrical Engineering, EEEIC 2015 - Conference Proceedings (2015) 1175–1180doi:10.1109/EEEIC.2015.7165334.

[106] M. Figueiredo, B. Ribeiro, A. D. Almeida, Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home, IEEE Transactions on Instrumentation and Measurement 63 (2014). doi:10.1109/TIM.2013.2278596.

[107] J. Z. Kolter, S. Batra, A. Y. Ng, Energy disaggregation via discriminative sparse coding, Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010 (2010).

[108] N. Mohan, K. Soman, S. Sachin Kumar, A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model, Applied Energy 232 (2018) 229–244. doi:https://doi.org/10.1016/j.apenergy.2018.09.190. URL https://www.sciencedirect.com/science/article/pii/S0306261918315009

[109] M. Thakur, A new genetic algorithm for global optimization of multimodal continuous functions, Journal of Computational Science 5 (2014). doi:10.1016/j.jocs.2013.05.005.

[110] K. Socha, M. Dorigo, Ant colony optimization for continuous domains, European Journal of Operational Research 185 (3) (2008) 1155–1173. doi:10.1016/j.ejor.2006.06.046.

[111] U. Nations, International standard industrial classification of all economic activities (isic) rev. 4, Statistical Papers 1 (2008) 307.

Appendix A

Definitions

A.1 Equipment Type

- Type I - ON/OFF equipment: Equipment with only two possible states (ON/OFF);
- Type II - FSM: Equipment's power consumption passes through state transitions;
- Type III - Continuously varying equipment: Equipment where the power consumption values can vary through time in a continuous domain;
- Type IV - Permanent consumer equipment: Equipment with only one state.

A.2 Event

An event corresponds to a significant variation in the aggregate electrical signal and suggests a change in the state of one equipment.

A.3 Industrial Sector

The industrial sector is defined according to the International Standard Industrial Classification of All Economic Activities (ISIC) [111]. It encompasses various industries, such as iron and steel, chemicals, cement, aluminium, pulp and paper, and light industry. Light industry refers to a group of sectors with a lower energy usage, which includes food production, timber, machinery, vehicles, textiles and other consumer goods, construction and mining.

A.4 Load Classification

Load classification is the process of identifying the state of the power consumption of each equipment in an aggregate.

A.5 Low-Frequency

Low-frequency refers to frequencies equal to or less than 1Hz.

A.6 State

The equipment state refers to the discrete operating modes of the equipment, which can be ON or OFF.

A.7 Source Separation

Source separation refers to the process of estimating the power consumption of the equipment in an aggregate.

A.8 Unsupervised

Supervised NILM algorithms use a priori knowledge of equipment consumption data, such as labelled consumption data or signature loads, while unsupervised algorithms do not have access to equipment data.

Appendix B

Results

B.1 Results from the Preprocessing of the HIPE Dataset

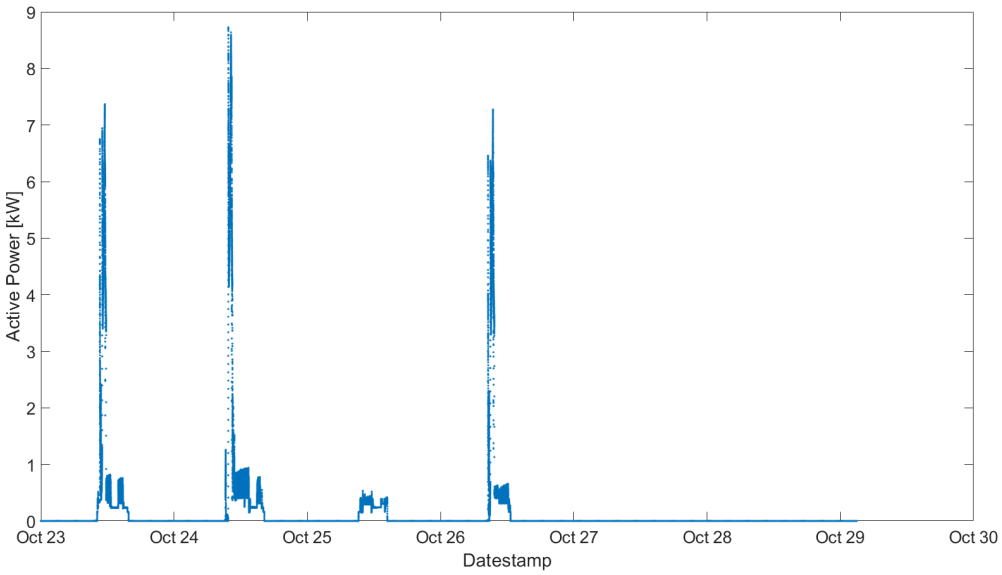

Figure B.1: Preprocessed aggregate active power data for the HIPE dataset for the sum of the equipment two and three.

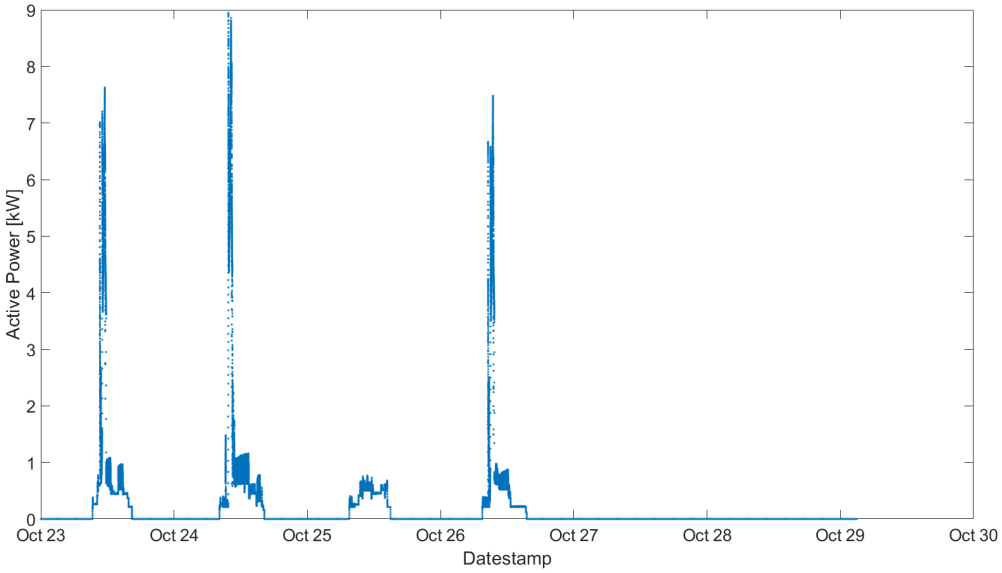

Figure B.2: Preprocessed aggregate active power data for the HIPE dataset for the sum of the equipment with indexes two through four.

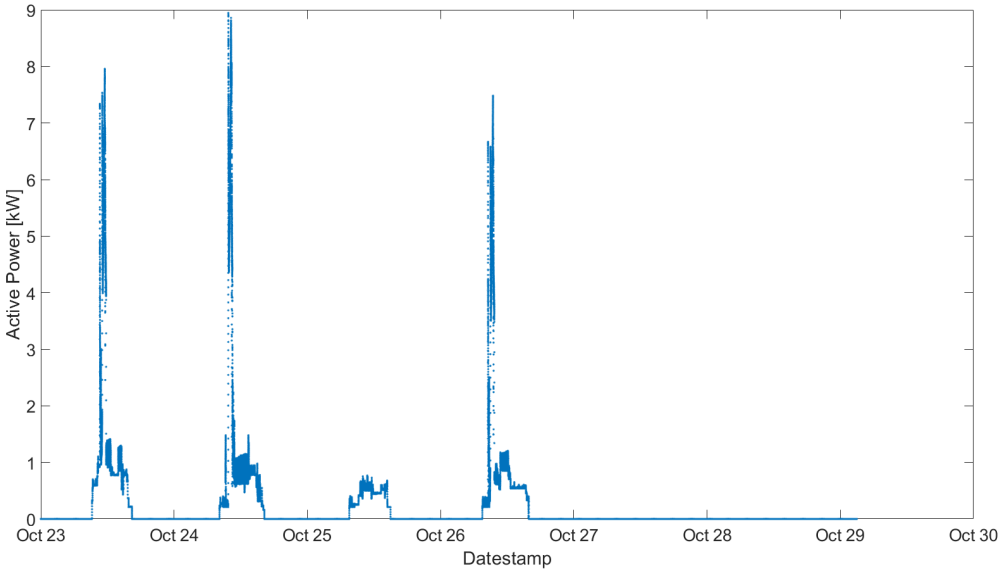

Figure B.3: Preprocessed aggregate active power data for the HIPE dataset for the sum of the equipment with indexes two through five.

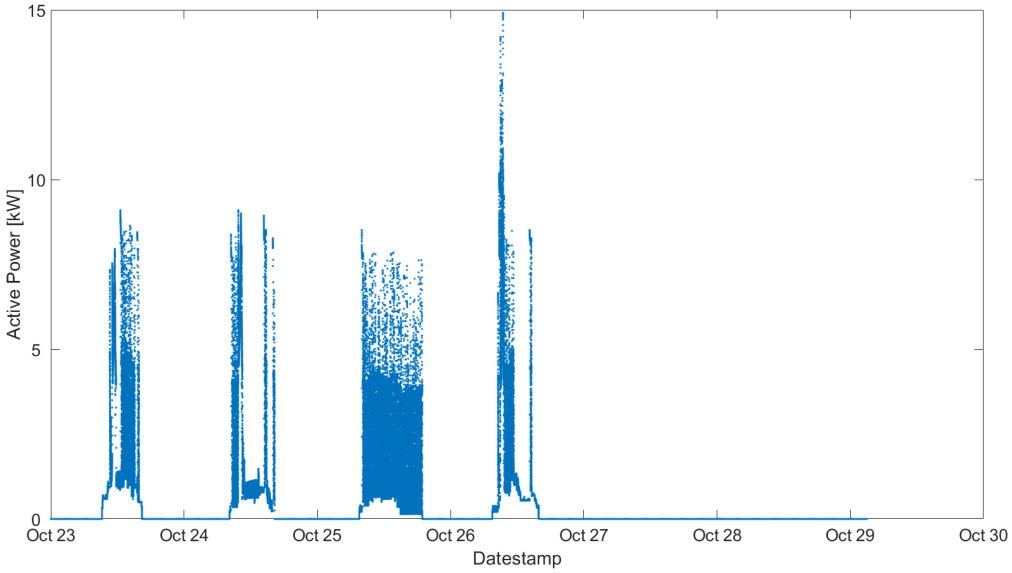

Figure B.4: Preprocessed aggregate active power data for the HIPE dataset for the sum of the equipment with indexes two through six.

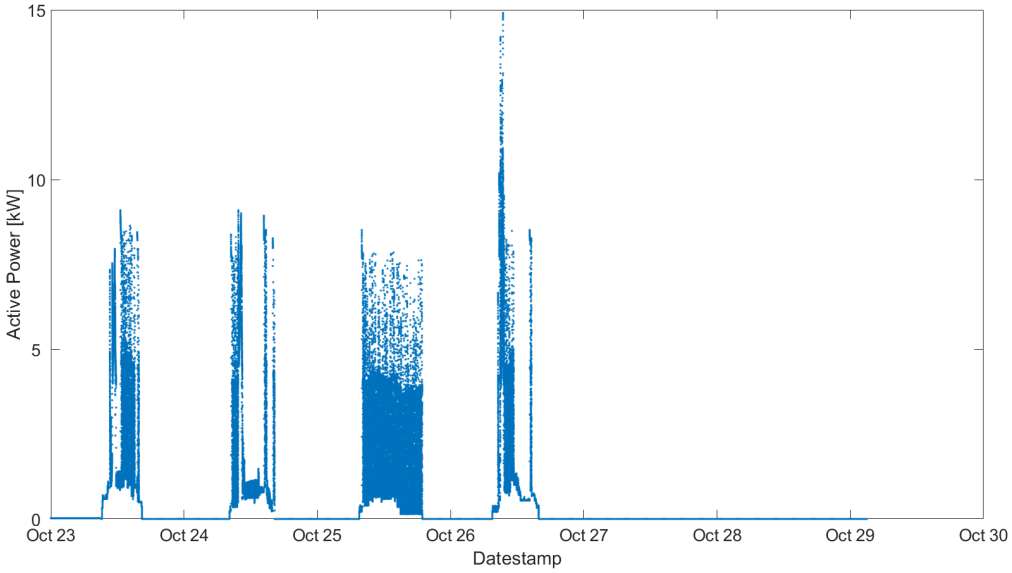

Figure B.5: Preprocessed aggregate active power data for the HIPE dataset for the sum of the equipment with indexes two through seven.

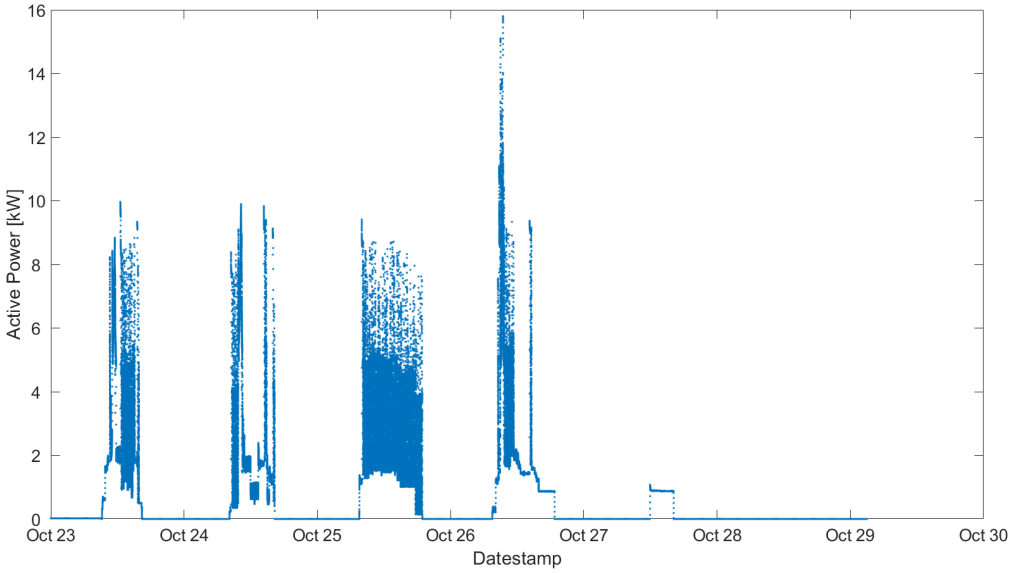

Figure B.6: Preprocessed aggregate active power data for the HIPE dataset for the sum of the equipment with indexes two through eight.

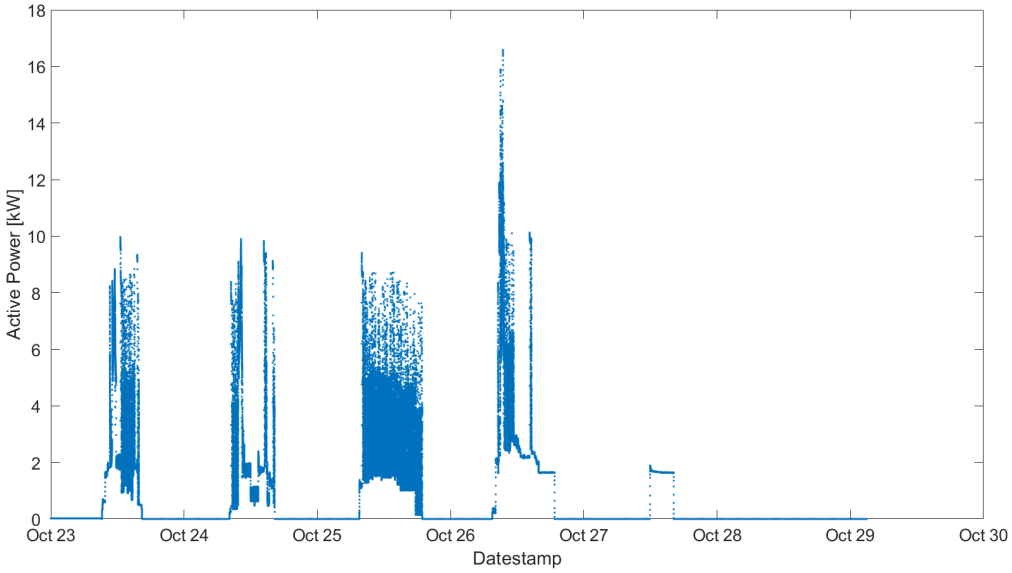

Figure B.7: Preprocessed aggregate active power data for the HIPE dataset for the sum of the equipment with indexes two through nine.

B.2 Estimated Equipment Active Power Values

B.2.1 Estimations for the HIPE Dataset

B.2.1.1 Estimations by the EMUPF Method

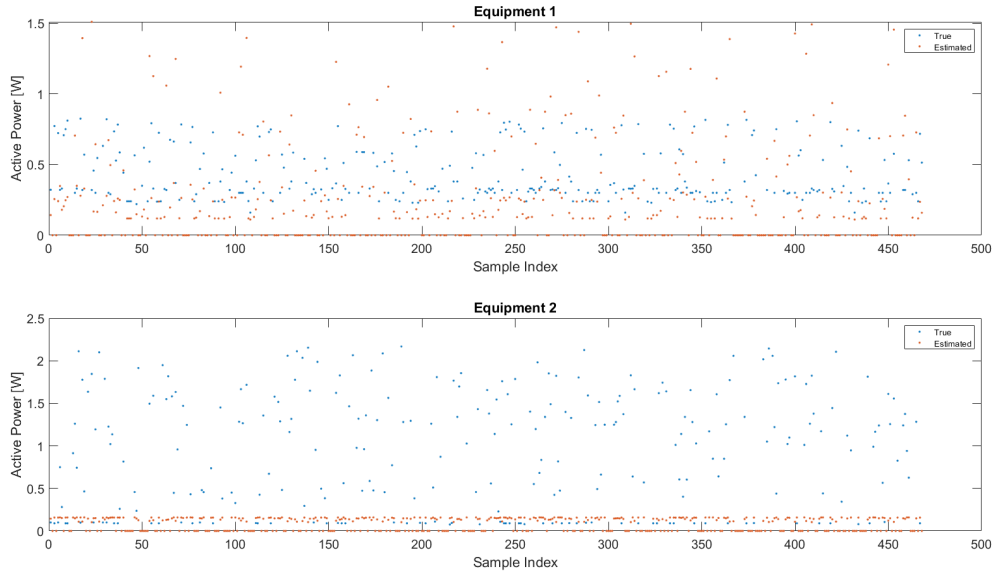

Figure B.8: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two and three.

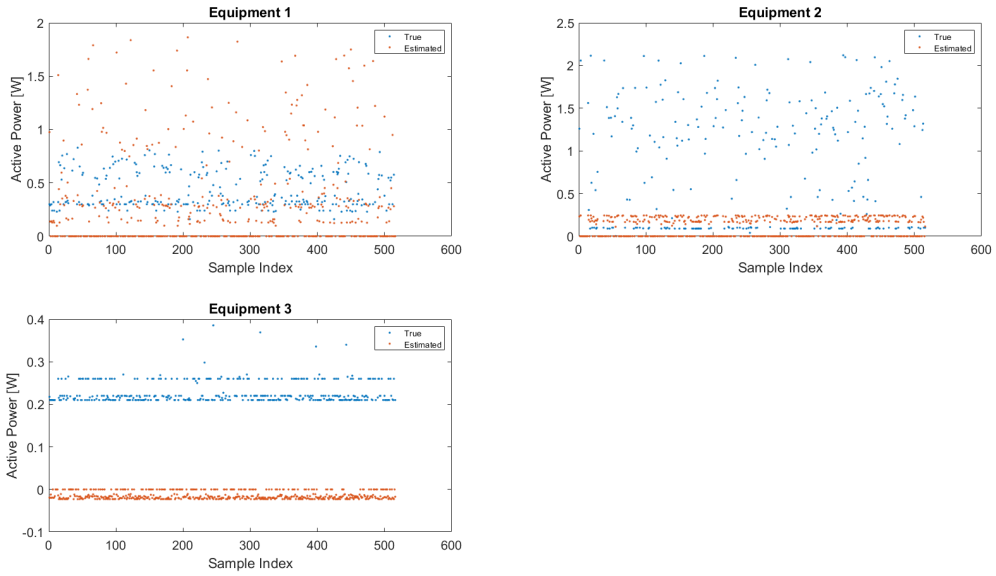

Figure B.9: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two through four.

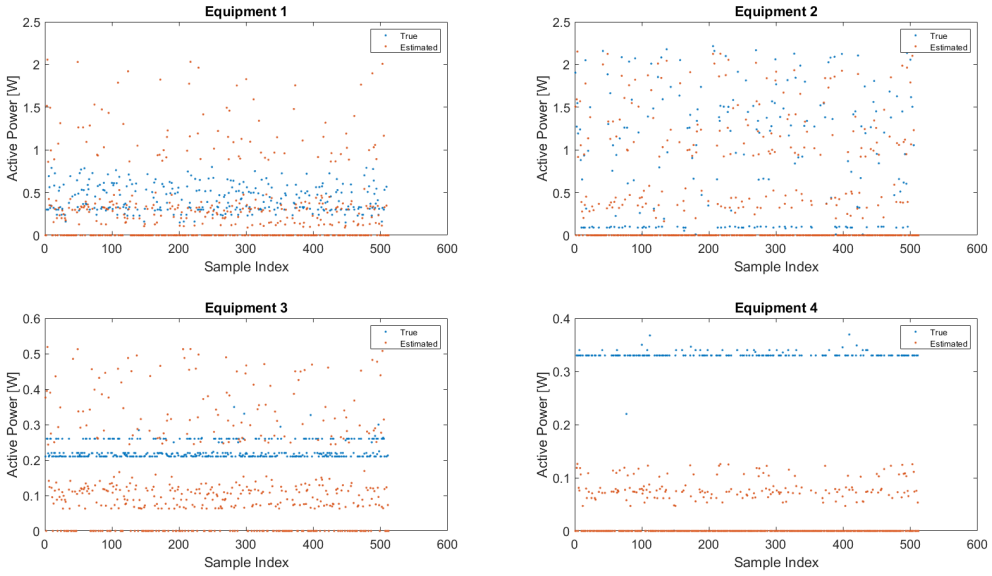

Figure B.10: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two through five.

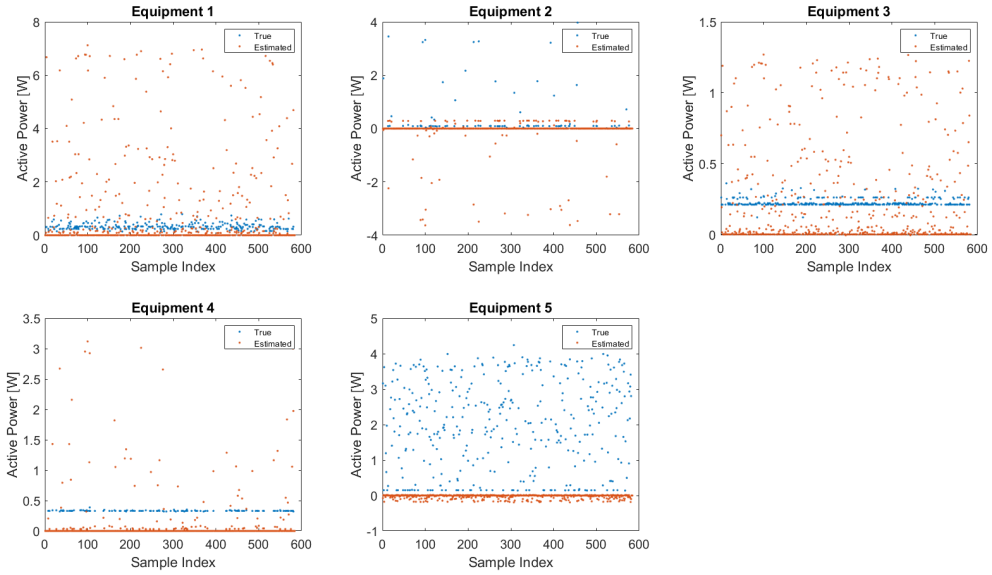

Figure B.11: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two through six.

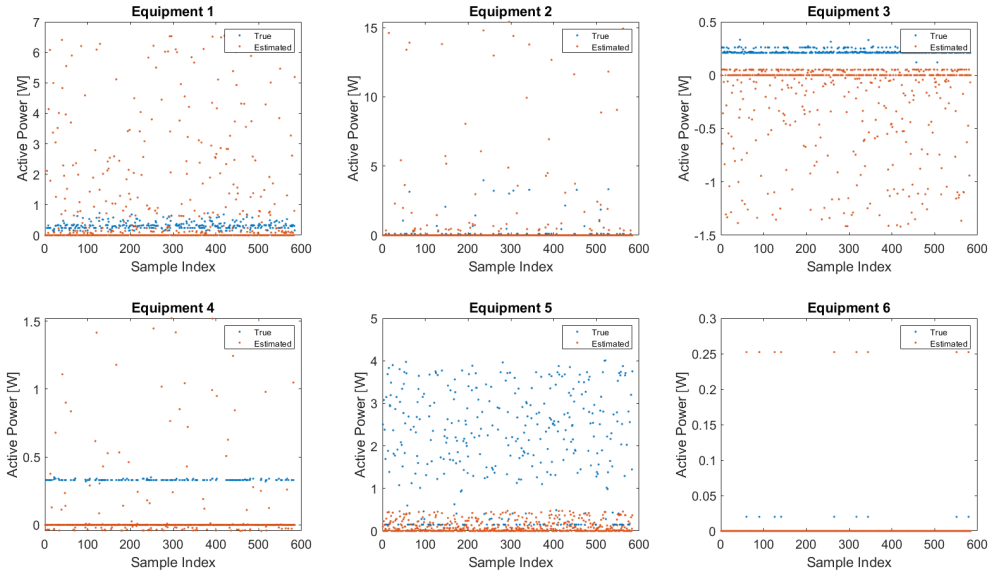

Figure B.12: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two through seven.

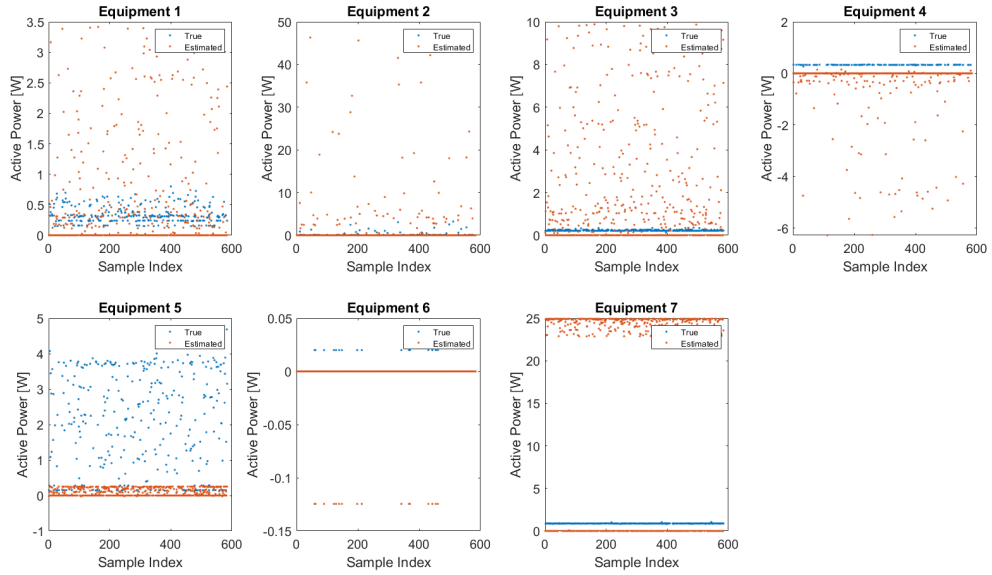

Figure B.13: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two through eight.

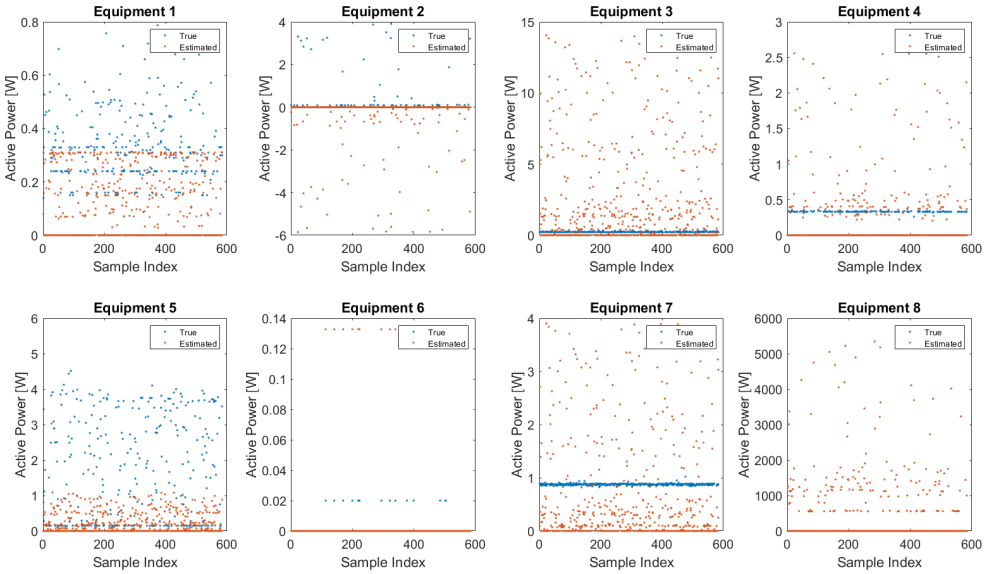

Figure B.14: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two through nine.

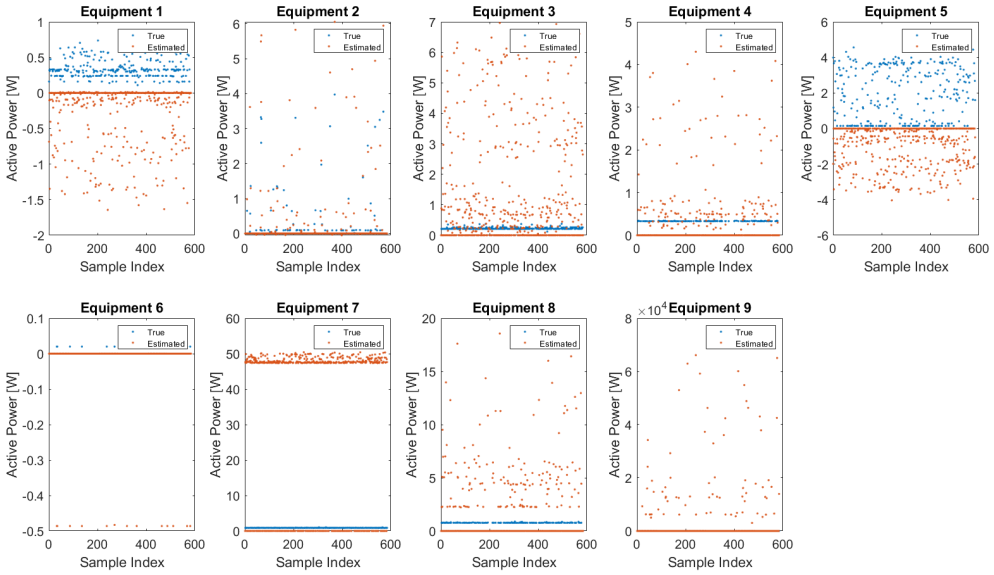

Figure B.15: Expected and estimated equipment active power samples, estimated by the EMUPF method for the HIPE dataset. The aggregate was calculated as the sum of the equipment with indexes two through ten.

B.2.1.2 Estimations by the UNN Method

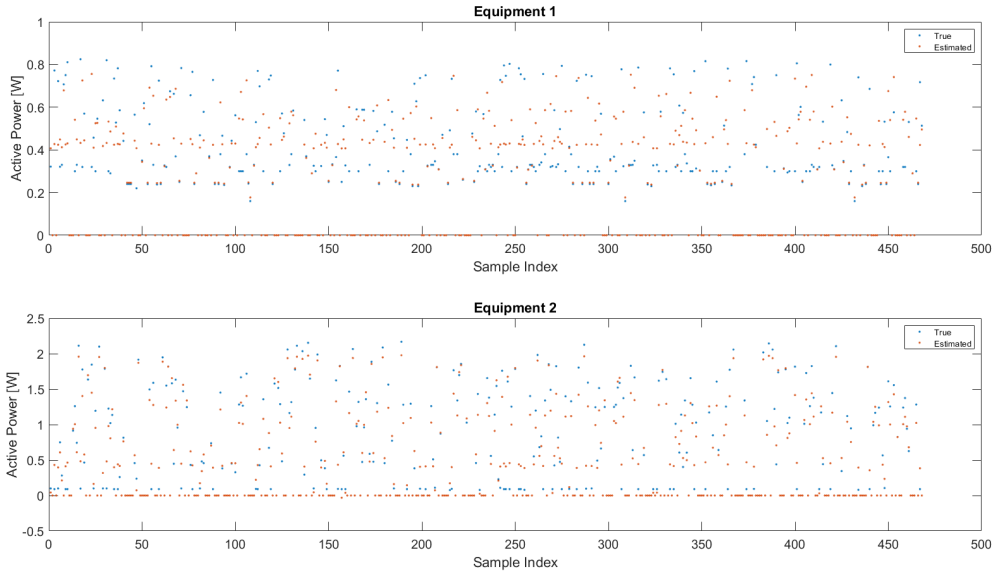

Figure B.16: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two and three.

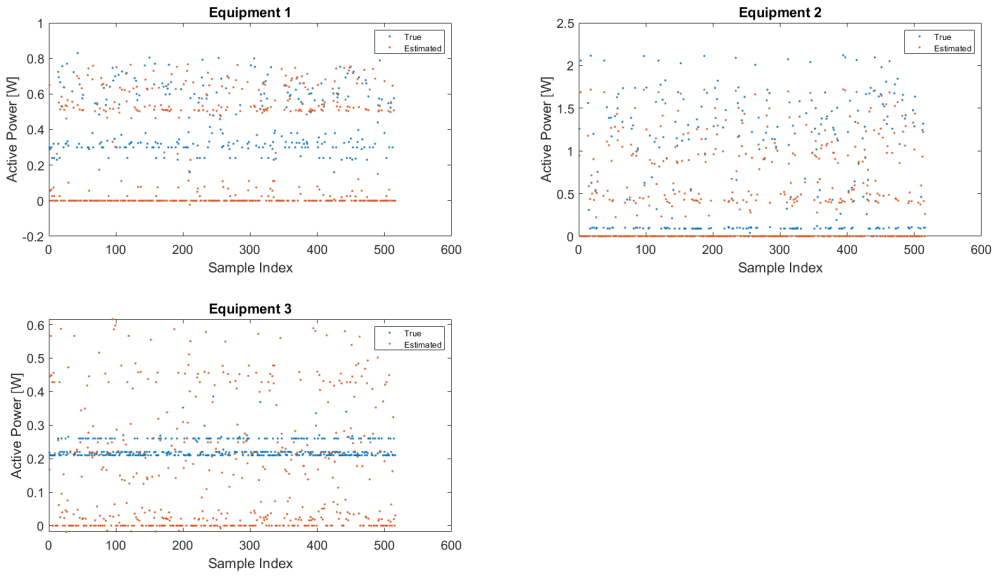

Figure B.17: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two through four.

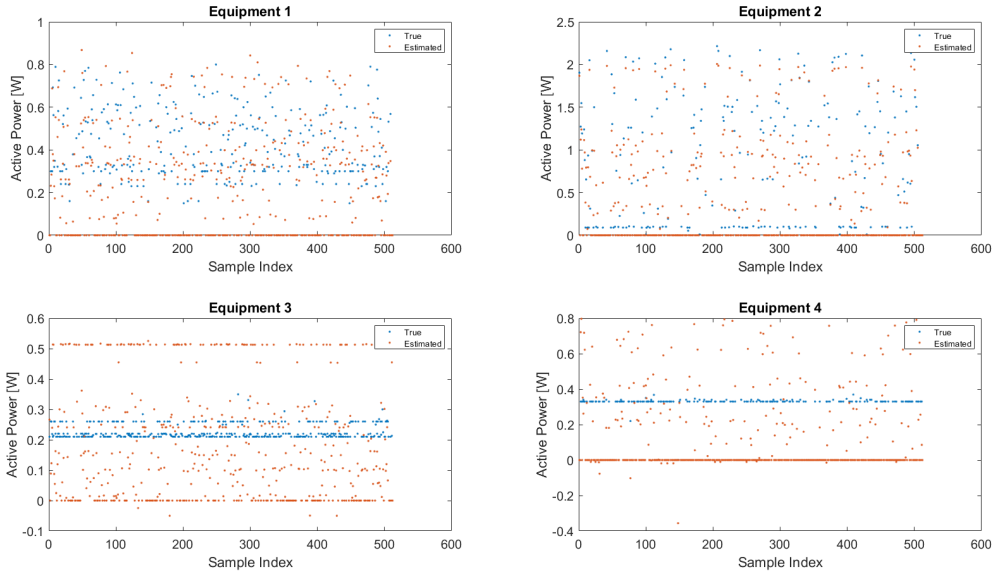

Figure B.18: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two through five.

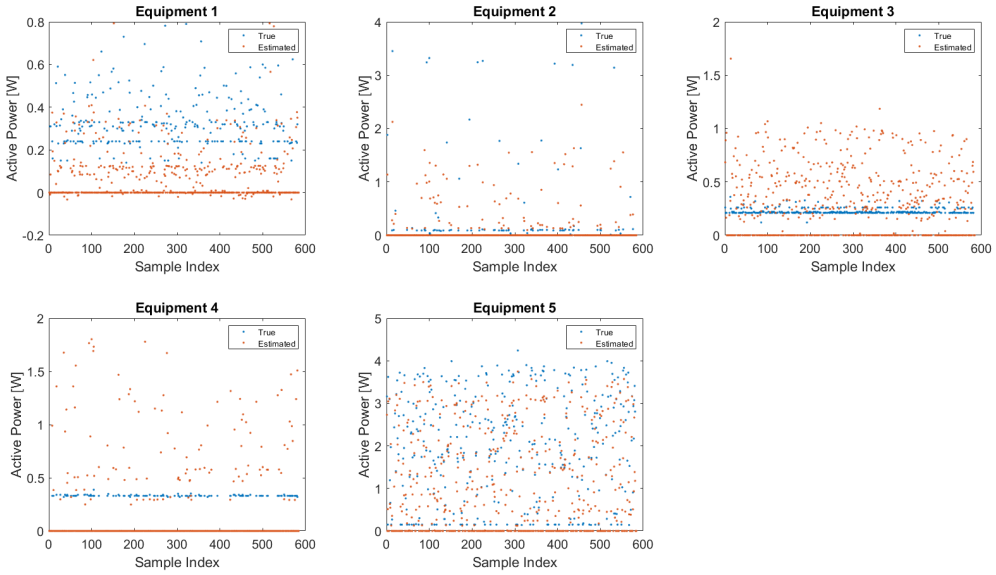

Figure B.19: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two through six.

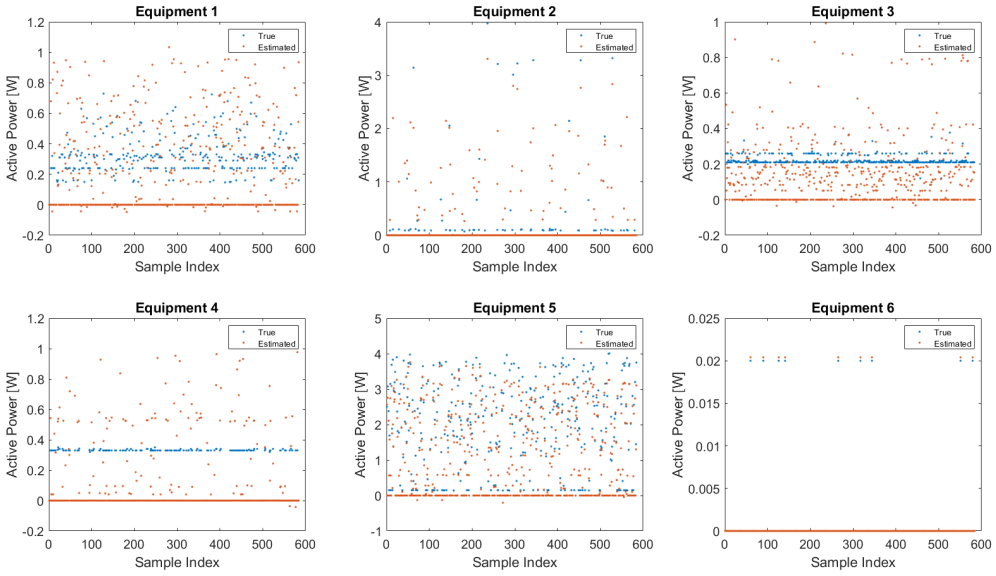

Figure B.20: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two through seven.

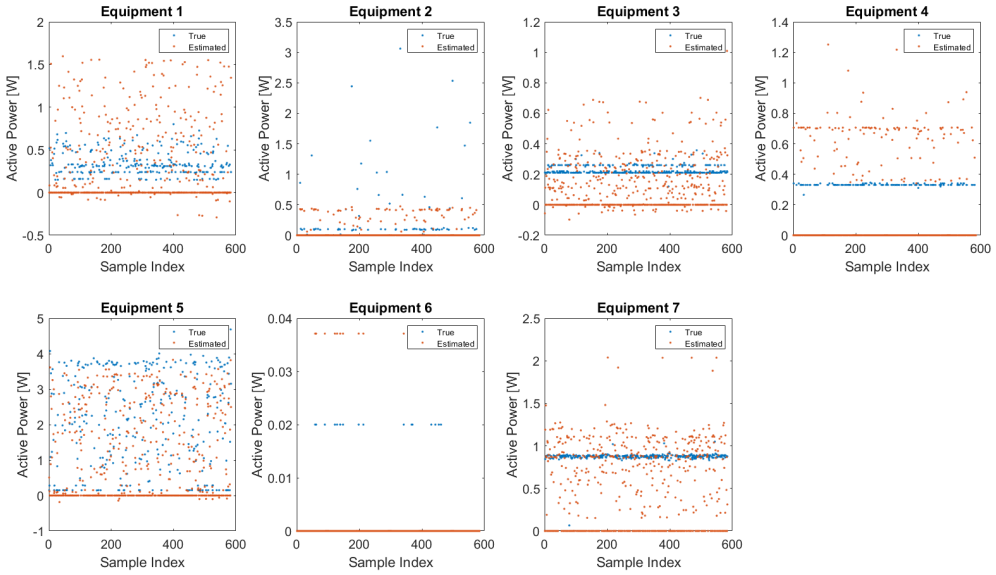

Figure B.21: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two through eight.

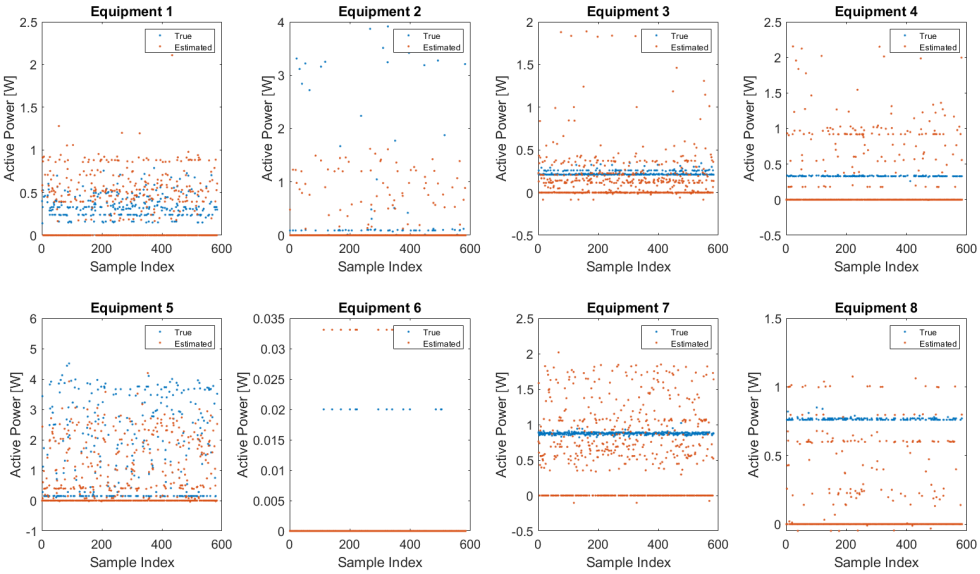

Figure B.22: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two through nine.

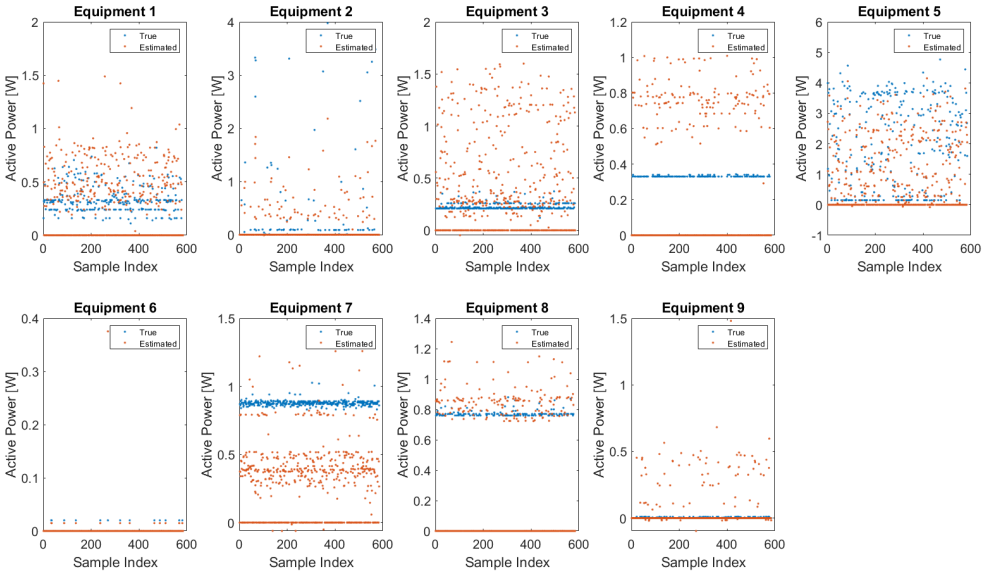

Figure B.23: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method. The aggregate was calculated as the sum of the equipment with indexes two through ten.

B.2.1.3 Estimations by the UNN Method with Fourier Mapping

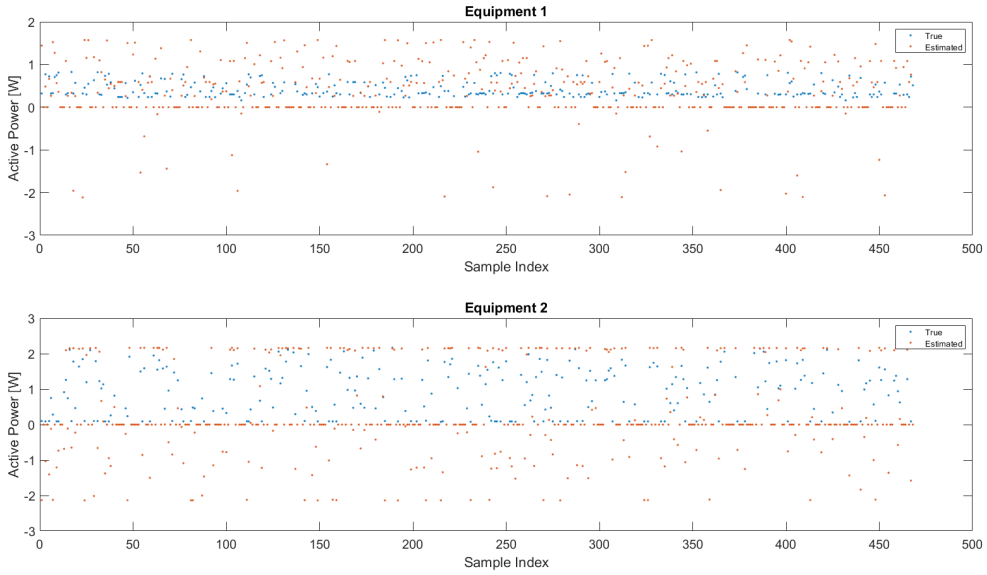

Figure B.24: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two and three.

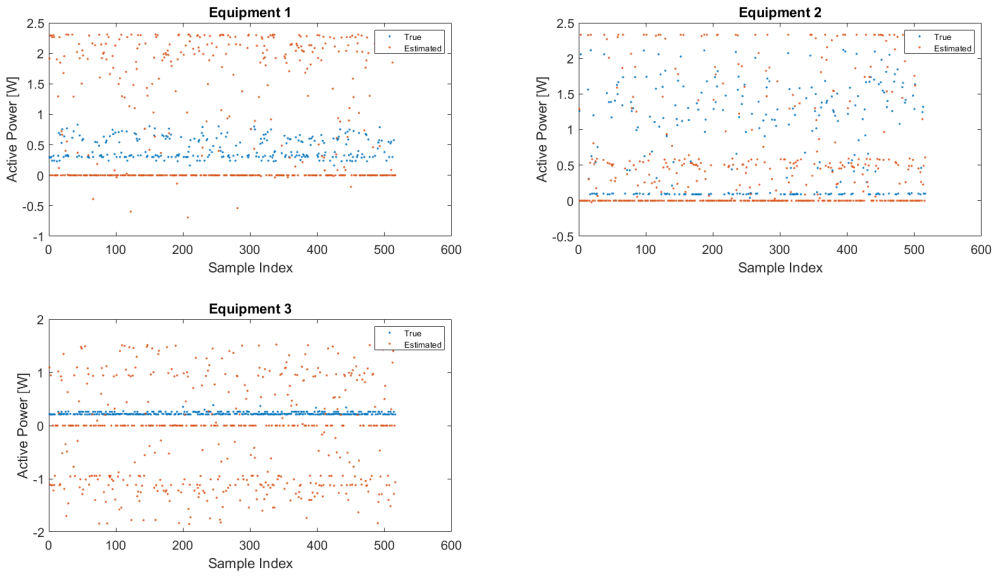

Figure B.25: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two through four.

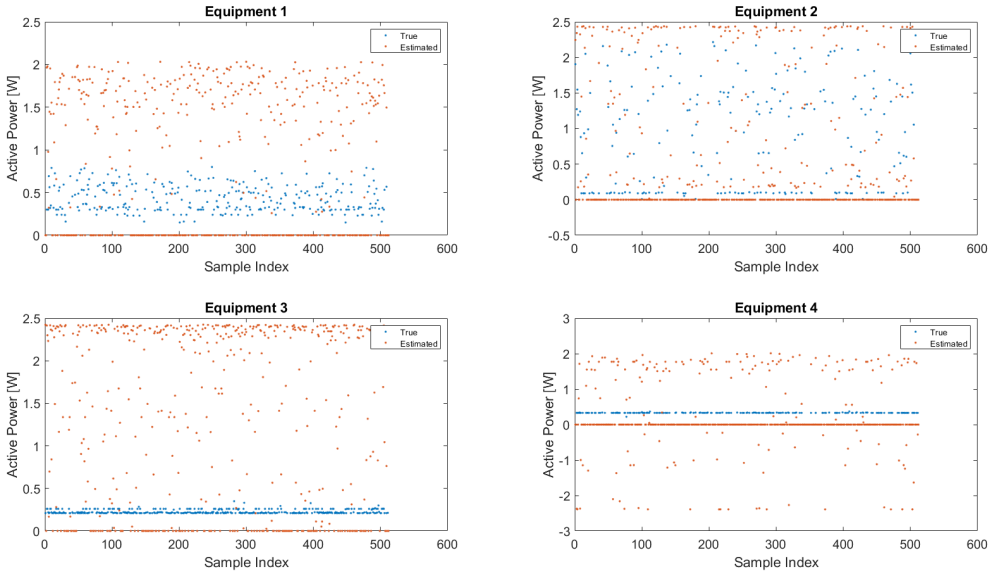

Figure B.26: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two through five.

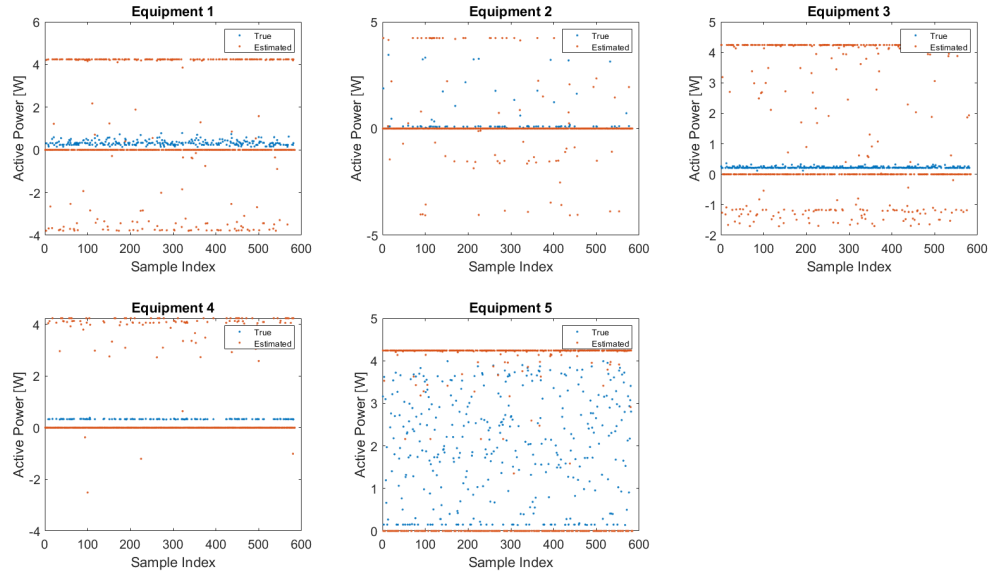

Figure B.27: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two through six.

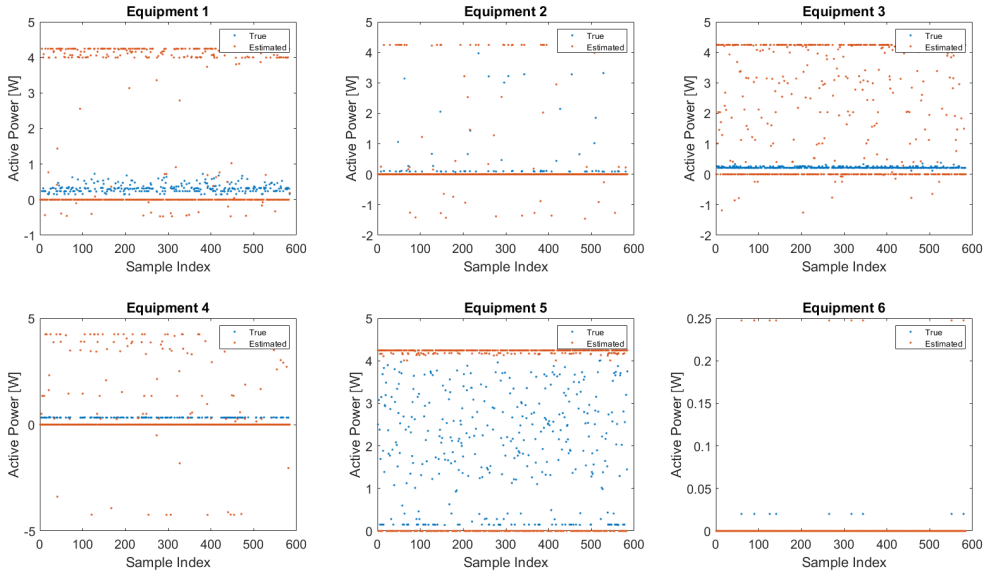

Figure B.28: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two through seven.

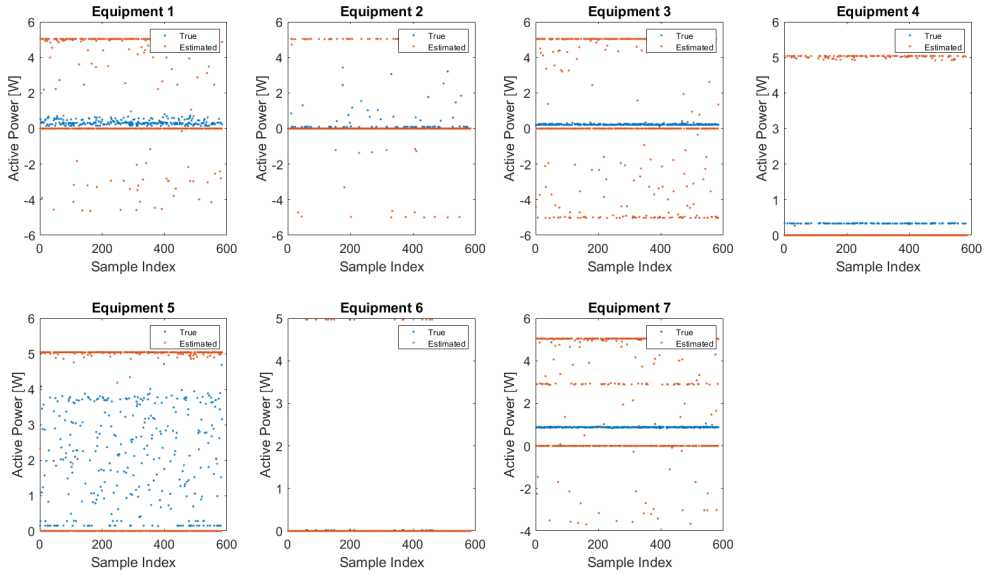

Figure B.29: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two through eight.

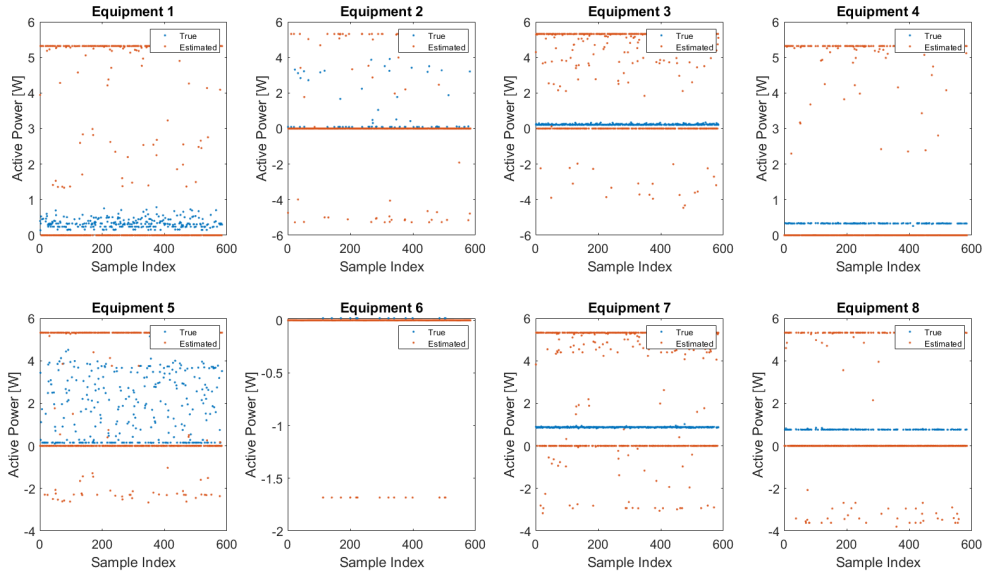

Figure B.30: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two through nine.

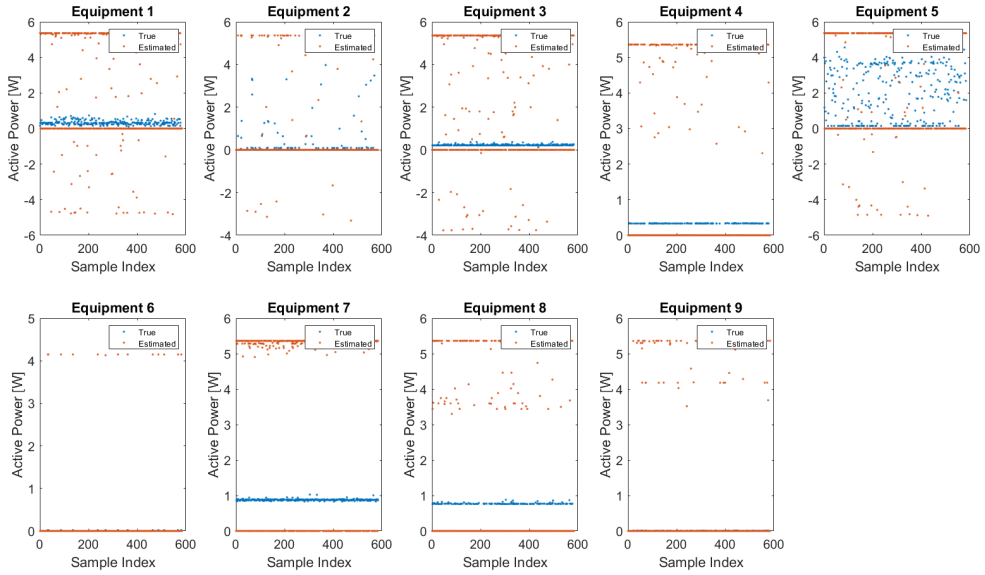

Figure B.31: Expected and estimated equipment active power samples for the HIPE dataset, estimated by the UNN method with Fourier mapping. The aggregate was calculated as the sum of the equipment with indexes two through ten.

B.2.2 Estimation for the IMDELD Dataset

B.2.2.1 Estimations by the EMUPF Method

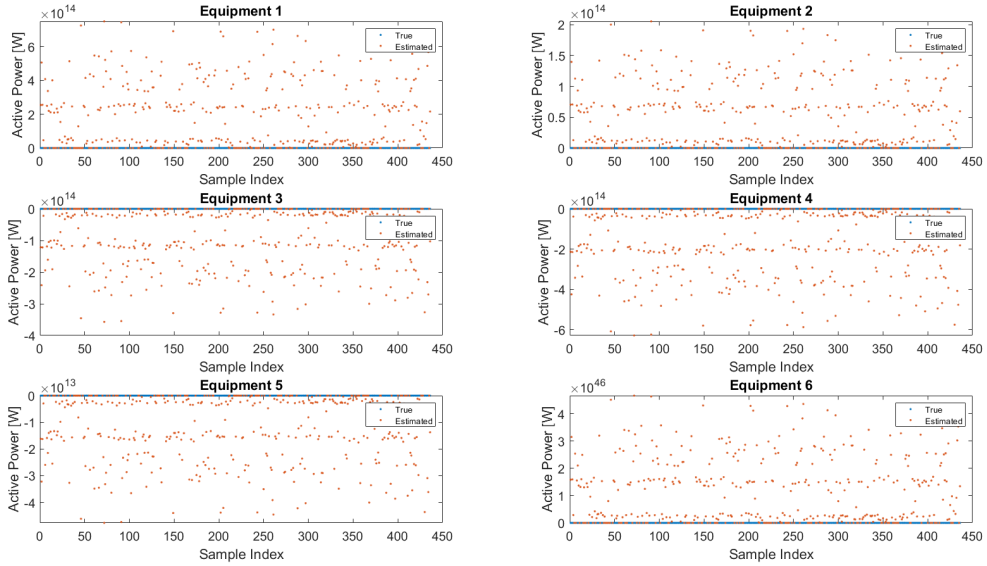

Figure B.32: Expected and estimated equipment active power samples, estimated by the EMUPF method, for the IMDELD dataset.

B.2.2.2 Estimations by the UNN Method

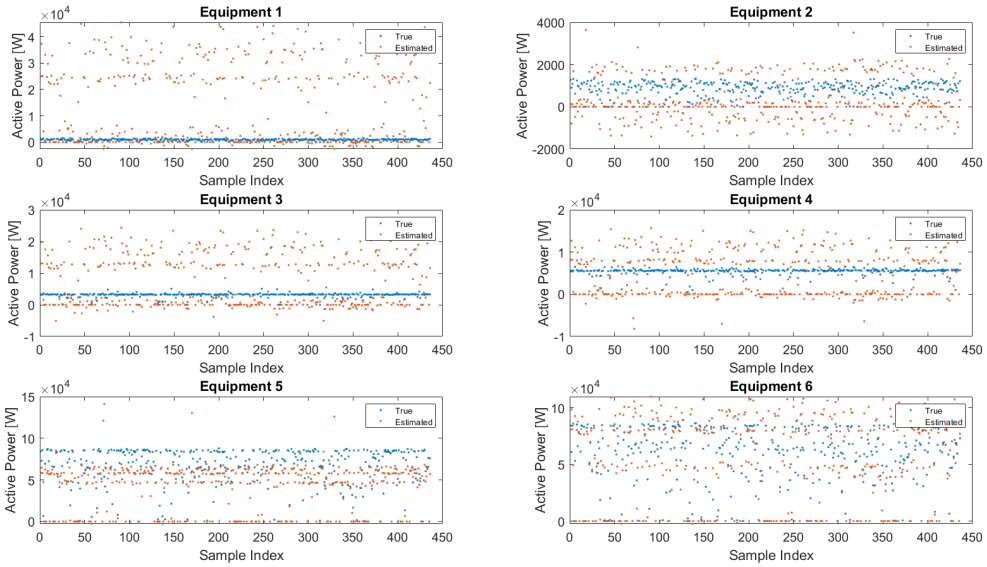

Figure B.33: Expected and estimated equipment active power samples, estimated by the UNN method, for the IMDELD dataset.

B.2.2.3 Estimations by the UNN Method with Fourier Mapping

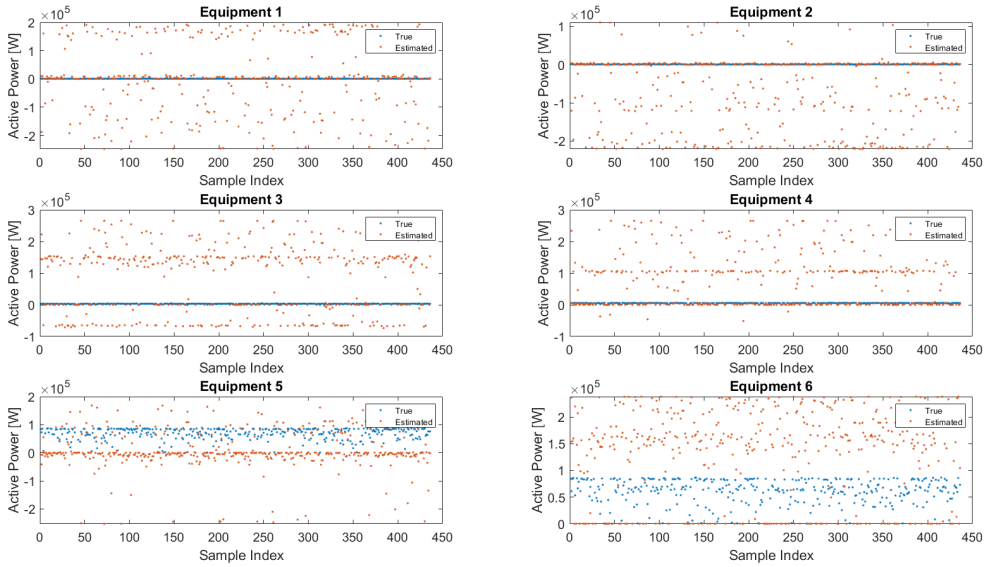

Figure B.34: Expected and estimated equipment active power samples, estimated by the UNN method, with Fourier mapping, for the IMDELD dataset.

B.3 MSE and RMSE Values for the HIPE Dataset

B.3.1 MSE and RMSE Values for the EMUPF Method

Table B.1: MSE for the equipment active power samples, estimated by the EMUPF method, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.1192	0.6264	-	-	-	-	-	-	-
Agg. 2	0.1647	0.4556	0.046	-	-	-	-	-	-
Agg. 3	0.1681	0.041	0.0136	0.0218	-	-	-	-	-
Agg. 4	3.2647	0.8883	0.1096	0.1287	3.6171	-	-	-	-
Agg. 5	3.1022	4.4528	0.333	0.0467	2.8758	0.0008	-	-	-
Agg. 6	0.7532	43.517	11.685	1.3279	2.857	0.0006	346.4	-	-
Agg. 7	0.0225	2.3615	18.645	0.1818	1.7586	0.0003	0.7848	9.9832e+05	-
Agg. 8	0.3615	0.2944	4.161	0.4112	9.0028	0.0057	1404.3	9.089	9.125e+07

Table B.2: RMSE for the equipment active power samples, estimated by the EMUPF method, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.3453	0.7915	-	-	-	-	-	-	-
Agg. 2	0.4058	0.675	0.2145	-	-	-	-	-	-
Agg. 3	0.41	0.2025	0.1166	0.1476	-	-	-	-	-
Agg. 4	1.8068	0.9425	0.3311	0.3587	1.9019	-	-	-	-
Agg. 5	1.7613	2.1102	0.5771	0.2161	1.6958	0.0283	-	-	-
Agg. 6	0.8679	6.5967	3.4184	1.1523	1.6903	0.0245	18.612	-	-
Agg. 7	0.15	1.5367	4.318	0.4264	1.3261	0.0173	0.8859	999.16	-
Agg. 8	0.6012	0.5426	2.0399	0.6412	3.0005	0.0755	37.474	3.0148	9552.5

B.3.2 MSE and RMSE Values for the UNN Method

Table B.3: MSE for the equipment active power samples, estimated by the UNN method, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.0275	0.0277	-	-	-	-	-	-	-
Agg. 2	0.0328	0.047	0.0247	-	-	-	-	-	-
Agg. 3	0.0345	0.0436	0.0257	0.0194	-	-	-	-	-
Agg. 4	0.0324	0.1145	0.0943	0.0805	0.2544	-	-	-	-
Agg. 5	0.061	0.1153	0.0207	0.0189	0.1481	0	-	-	-
Agg. 6	0.1626	0.0656	0.0191	0.0262	0.1155	0	0.0683	-	-
Agg. 7	0.0661	0.2291	0.0766	0.1406	0.3025	0	0.1274	0.0443	-
Agg. 8	0.0599	0.0662	0.2091	0.0518	0.3649	0.0002	0.1441	0.0043	0.0182

Table B.4: RMSE for the equipment active power samples, estimated by the UNN method, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.1658	0.1664	-	-	-	-	-	-	-
Agg. 2	0.1811	0.2168	0.1572	-	-	-	-	-	-
Agg. 3	0.1857	0.2088	0.1603	0.1393	-	-	-	-	-
Agg. 4	0.18	0.3384	0.3071	0.2837	0.5044	-	-	-	-
Agg. 5	0.247	0.3396	0.1439	0.1375	0.3848	0	-	-	-
Agg. 6	0.4032	0.2561	0.1382	0.1619	0.3399	0	0.2613	-	-
Agg. 7	0.2571	0.4786	0.2768	0.375	0.55	0	0.3569	0.2105	-
Agg. 8	0.2447	0.2573	0.4573	0.2276	0.6041	0.0141	0.3796	0.0656	0.1349

B.3.3 MSE and RMSE Values for the UNN Method with Fourier Mapping

Table B.5: MSE for the equipment active power samples, estimated by the UNN method with Fourier mapping, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.4197	0.6843	-	-	-	-	-	-	-
Agg. 2	1.1302	0.1319	1.0073	-	-	-	-	-	-
Agg. 3	0.8365	0.6386	2.3091	0.7737	-	-	-	-	-
Agg. 4	6.4784	1.6187	6.555	2.5704	3.8796	-	-	-	-
Agg. 5	6.0134	0.9614	7.0107	2.1179	3.9107	0.0008	-	-	-
Agg. 6	8.3724	2.7539	12.044	4.8597	5.3693	0.6705	8.6949	-	-
Agg. 7	10.205	3.8132	13.745	6.2497	5.939	0.0695	11.02	5.2141	-
Agg. 8	10.41	2.469	13.505	5.9649	5.7504	0.3789	12.485	4.8396	3.7414

Table B.6: RMSE for the equipment active power samples, estimated by the UNN method with Fourier mapping, for the testing data from the HIPE dataset.

	Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	0.6478	0.8272	-	-	-	-	-	-	-
Agg. 2	1.0631	0.3632	1.0036	-	-	-	-	-	-
Agg. 3	0.9146	0.7991	1.5196	0.8796	-	-	-	-	-
Agg. 4	2.5453	1.2723	2.5603	1.6032	1.9697	-	-	-	-
Agg. 5	2.4522	0.9805	2.6478	1.4553	1.9775	0.0283	-	-	-
Agg. 6	2.8935	1.6595	3.4705	2.2045	2.3172	0.8188	2.9487	-	-
Agg. 7	3.1945	1.9527	3.7075	2.4999	2.437	0.2636	3.3197	2.2834	-
Agg. 8	3.2264	1.5713	3.6749	2.4423	2.398	0.6155	3.5334	2.1999	1.9343

B.4 Descriptive Statistical Analysis

B.4.1 Analysis for the HIPE Dataset

B.4.1.1 Maximum, Minimum, Median and Sum Values for the EMUPF Method

Table B.7: Maximum active power values for each equipment, for the expected and estimated values calculated by the EMUPF method, for the HIPE dataset.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0.8248	2.1679	-	-	-	-	-	-	-
	Estimated	1.50915	0.159669	-	-	-	-	-	-	-
Agg. 2	Expected	0.83	2.12	0.3856	-	-	-	-	-	-
	Estimated	0.86412	0.243791	0	-	-	-	-	-	-
Agg. 3	Expected	0.7999	2.2154	0.3495	0.3694	-	-	-	-	-
	Estimated	2.05681	2.15237	0.518802	0.12601	-	-	-	-	-
Agg. 4	Expected	0.79	3.9703	0.36	0.3878	4.2412	-	-	-	-
	Estimated	7.12293	0.29629	1.26898	3.11844	0.0149771	-	-	-	-
Agg. 5	Expected	0.7288	3.9703	0.377	0.35	4.0081	0.02	-	-	-
	Estimated	6.54564	15.4126	0.0527945	1.52108	0.472843	0.252218	-	-	-
Agg. 6	Expected	0.8014	3.2136	0.3574	0.3454	4.7107	0.02	1.05	-	-
	Estimated	3.41681	46.3239	9.88307	0.146988	0.253395	0	24.9324	-	-
Agg. 7	Expected	0.7876	3.9134	0.347	0.3494	5.1471	0.02	1.0215	0.85	-
	Estimated	0.311008	0	14.0516	2.55827	1.08503	0.132737	3.90077	5344.14	-
Agg. 8	Expected	0.8206	3.9765	0.3744	0.3435	5.1028	0.02	1.0259	0.88	0.01
	Estimated	0.0129706	6.05432	6.96494	4.29919	0.00687302	0	50.4064	18.5576	66106.1

Table B.8: Minimum active power values for each equipment, for the expected and estimated values calculated by the EMUPF method, for the HIPE dataset.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0	0	-	-	-	-	-	-	-
	Estimated	0	0	-	-	-	-	-	-	-
Agg. 2	Expected	0	0	0	-	-	-	-	-	-
	Estimated	0	0	-0.0229015	-	-	-	-	-	-
Agg. 3	Expected	0	0	0	0	-	-	-	-	-
	Estimated	0	0	0	0	-	-	-	-	-
Agg. 4	Expected	0	0	0	0	0	-	-	-	-
	Estimated	0	-3.62909	-0.00639624	0	-0.193847	-	-	-	-
Agg. 5	Expected	0	0	0	0	0	0	-	-	-
	Estimated	0	0	-1.4192	-0.0446273	0	0	-	-	-
Agg. 6	Expected	0	0	0	0	0	0	0	-	-
	Estimated	0	0	0	-6.28221	-0.0230436	-0.124548	0	-	-
Agg. 7	Expected	0	0	0	0	0	0	0	0	-
	Estimated	0	-5.85447	0	0	0	0	0	0	-
Agg. 8	Expected	0	0	0	0	0	0	0	0	0
	Estimated	-1.64326	-0.0497186	0	0	-4.03606	-0.486072	0	0	-0.0933964

Table B.9: Median active power values for each equipment, for the expected and estimated values calculated by the EMUPF method, for the HIPE dataset.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0.3	0.09	-	-	-	-	-	-	-
	Estimated	0.12825	0.11792	-	-	-	-	-	-	-
Agg. 2	Expected	0.24	0.09	0.211	-	-	-	-	-	-
	Estimated	0.12843	0.16035	-0.01802	-	-	-	-	-	-
Agg. 3	Expected	0.25	0	0.21	0	-	-	-	-	-
	Estimated	0.14452	0	0.10383	0	-	-	-	-	-
Agg. 4	Expected	0	0	0.21	0	0.50465	-	-	-	-
	Estimated	0	0	0.019924	0	0	-	-	-	-
Agg. 5	Expected	0	0	0.21	0	0.40395	0	-	-	-
	Estimated	0	0	0	0	0.020006	0	-	-	-
Agg. 6	Expected	0	0	0.21	0	0.15	0	0.86205	-	-
	Estimated	0	0	0.54418	0	0.033342	0	23.862	-	-
Agg. 7	Expected	0	0	0.21	0	0.15	0	0.86985	0	-
	Estimated	0	0	0.63784	0	0.054635	0	0.10542	0	-
Agg. 8	Expected	0	0	0.21	0	0.14535	0	0.87	0	0
	Estimated	0	0	0.27832	0	-0.034273	0	47.408	0	0

Table B.10: Sum of the active power values for each equipment, for the expected and estimated values calculated by the EMUPF method, for the HIPE dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 2	Expected	125.08	249.07	-	-	-	-	-	-	-
	Estimated	121.39	38.11	-	-	-	-	-	-	-
Agg. 3	Expected	122.08	234.09	87.348	-	-	-	-	-	-
	Estimated	151.71	55.846	-7.3584	-	-	-	-	-	-
Agg. 4	Expected	126.32	198.99	84.838	58.34	-	-	-	-	-
	Estimated	160.33	209.8	69.445	14.305	-	-	-	-	-
Agg. 5	Expected	88.739	52.943	80.841	38.863	729.3	-	-	-	-
	Estimated	505.6	-54.665	132.63	61.218	-14.538	-	-	-	-
Agg. 6	Expected	91.099	44.555	82.972	44.145	727.96	0.18	-	-	-
	Estimated	524.41	324.11	-122.61	32.846	56.733	2.27	-	-	-
Agg. 7	Expected	90.068	33.5	75.573	42.741	661.82	0.32	318.44	-	-
	Estimated	280.78	933.52	1170	-171.76	52.898	-1.9928	8896.3	-	-
Agg. 8	Expected	93.94	67.378	80.764	52.377	547.9	0.28	331.22	121.18	-
	Estimated	55.608	-150.46	1446.5	128.13	122.25	1.8583	366.96	2.4392e+05	-
Agg. 9	Expected	90.777	53.91	77.235	50.051	582.88	0.26	322.97	121.27	0.6008
	Estimated	-111.43	117.21	705.5	158.08	-427.64	-6.3158	17693	861.16	1.5124e+06

B.4.1.2 Maximum, Minimum, Median and Sum Values for the UNN Method

Table B.11: Maximum active power values for each equipment, for the expected and estimated values calculated by the UNN method, for the HIPE dataset.

		Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	Expected	0.8248	2.1679	-	-	-	-	-	-	-
	Estimated	0.7558	1.9776	-	-	-	-	-	-	-
Agg. 2	Expected	0.83	2.12	0.3856	-	-	-	-	-	-
	Estimated	0.7673	1.7204	0.6161	-	-	-	-	-	-
Agg. 3	Expected	0.7999	2.2154	0.3495	0.3694	-	-	-	-	-
	Estimated	0.8679	2.0062	0.525	0.7967	-	-	-	-	-
Agg. 4	Expected	0.79	3.9703	0.36	0.3878	4.2412	-	-	-	-
	Estimated	0.793	2.4464	1.9142	1.8019	3.7458	-	-	-	-
Agg. 5	Expected	0.7288	3.9703	0.377	0.35	4.0081	0.02	-	-	-
	Estimated	1.0578	3.306	0.9916	1.1228	3.7582	0.0204	-	-	-
Agg. 6	Expected	0.8014	3.2136	0.3574	0.3454	4.7107	0.02	1.05	-	-
	Estimated	1.5964	0.4741	1.02	1.2514	3.8266	0.0371	2.4001	-	-
Agg. 7	Expected	0.7876	3.9134	0.347	0.3494	5.1471	0.02	1.0215	0.85	-
	Estimated	2.1605	1.62	1.8879	2.155	4.1953	0.0331	2.0178	1.3772	-
Agg. 8	Expected	0.8206	3.9765	0.3744	0.3435	5.1028	0.02	1.0259	0.88	0.01
	Estimated	1.858	2.1815	1.5995	1.0088	4.1476	0.3753	1.3129	1.2436	1.4803

Table B.12: Minimum active power values for each equipment, for the expected and estimated values calculated by the UNN method, for the HIPE dataset.

		Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	Eq. 10
Agg. 1	Expected	0	0	-	-	-	-	-	-	-
	Estimated	0	-0.0317	-	-	-	-	-	-	-
Agg. 2	Expected	0	0	0	-	-	-	-	-	-
	Estimated	-0.0226	0	-0.0189	-	-	-	-	-	-
Agg. 3	Expected	0	0	0	0	-	-	-	-	-
	Estimated	0	0	-0.05	-0.3561	-	-	-	-	-
Agg. 4	Expected	0	0	0	0	0	-	-	-	-
	Estimated	-0.0327	0	0	0	0	-	-	-	-
Agg. 5	Expected	0	0	0	0	0	0	-	-	-
	Estimated	-0.047	0	-0.0434	-0.0425	-0.2037	0	-	-	-
Agg. 6	Expected	0	0	0	0	0	0	0	-	-
	Estimated	-0.2906	0	-0.0976	0	-0.185	0	0	-	-
Agg. 7	Expected	0	0	0	0	0	0	0	0	-
	Estimated	0	0	-0.0836	-0.0046	-0.036	0	-0.1035	-0.0496	-
Agg. 8	Expected	0	0	0	0	0	0	0	0	0
	Estimated	0	-0.0092	-0.0472	0	-0.08	0	-0.0618	0	-0.0966

Table B.13: Median active power values for each equipment, for the expected and estimated values calculated by the UNN method, for the HIPE dataset.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0.3	0.09	-	-	-	-	-	-	-
	Estimated	0.3172	0.3888	-	-	-	-	-	-	-
Agg. 2	Expected	0.24	0.09	0.211	-	-	-	-	-	-
	Estimated	0.0257	0.3664	0.08345	-	-	-	-	-	-
Agg. 3	Expected	0.25	0	0.21	0	-	-	-	-	-
	Estimated	0.161	0	0.11955	0	-	-	-	-	-
Agg. 4	Expected	0	0	0.21	0	0.50465	-	-	-	-
	Estimated	0	0	0.2547	0	0.52675	-	-	-	-
Agg. 5	Expected	0	0	0.21	0	0.40395	0	-	-	-
	Estimated	0	0	0.0853	0	0.69215	0	-	-	-
Agg. 6	Expected	0	0	0.21	0	0.15	0	0.86205	-	-
	Estimated	0	0	0.0348	0	0.1014	0	0.57475	-	-
Agg. 7	Expected	0	0	0.21	0	0.15	0	0.86985	0	-
	Estimated	0	0	0.11135	0	0.10135	0	0.5912	0	-
Agg. 8	Expected	0	0	0.21	0	0.14535	0	0.87	0	0
	Estimated	0	0	0.2311	0	0	0	0.3253	0	0

Table B.14: Sum of the active power values for each equipment, for the expected and estimated values calculated by the UNN method, for the HIPE dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 2	Expected	125.08	249.07	-	-	-	-	-	-	-
	Estimated	130.02	242.89	-	-	-	-	-	-	-
Agg. 3	Expected	122.08	234.09	87.348	-	-	-	-	-	-
	Estimated	128.72	229.71	84.816	-	-	-	-	-	-
Agg. 4	Expected	126.32	198.99	84.838	58.34	-	-	-	-	-
	Estimated	111.34	206.18	92.685	59.328	-	-	-	-	-
Agg. 5	Expected	88.739	52.943	80.841	38.863	729.3	-	-	-	-
	Estimated	38.604	55.982	182.73	93.441	613.83	-	-	-	-
Agg. 6	Expected	91.099	44.555	82.972	44.145	727.96	0.18	-	-	-
	Estimated	119.21	88.975	78.045	53.624	658.38	0.1836	-	-	-
Agg. 7	Expected	90.068	33.5	75.573	42.741	661.82	0.32	318.44	-	-
	Estimated	162.23	25.637	71.248	81.82	579.24	0.5936	309.04	-	-
Agg. 8	Expected	93.94	67.378	80.764	52.377	547.9	0.28	331.22	121.18	-
	Estimated	161.79	65.947	98.466	138.06	419.4	0.4634	366.56	79.674	-
Agg. 9	Expected	90.777	53.91	77.235	50.051	582.88	0.26	322.97	121.27	0.6008
	Estimated	149.28	53.595	218.28	115.11	444.13	0.5517	163.13	134.28	23.443

B.4.1.3 Maximum, Minimum, Median and Sum Values for the UNN Method with Fourier Mapping

Table B.15: Maximum active power values for each equipment, for the expected and estimated values calculated by the UNN method with Fourier mapping, for the HIPE dataset.

		Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	Expected	0.8248	2.1679	-	-	-	-	-	-	-
	Estimated	1.5747	2.1654	-	-	-	-	-	-	-
Agg. 2	Expected	0.83	2.12	0.3856	-	-	-	-	-	-
	Estimated	2.3089	2.3328	1.5224	-	-	-	-	-	-
Agg. 3	Expected	0.7999	2.2154	0.3495	0.3694	-	-	-	-	-
	Estimated	2.0322	2.4377	2.4242	2.0109	-	-	-	-	-
Agg. 4	Expected	0.79	3.9703	0.36	0.3878	4.2412	-	-	-	-
	Estimated	4.2412	4.2412	4.2412	4.2412	4.2412	-	-	-	-
Agg. 5	Expected	0.7288	3.9703	0.377	0.35	4.0081	0.02	-	-	-
	Estimated	4.2447	4.2447	4.2447	4.2447	4.2447	0.2474	-	-	-
Agg. 6	Expected	0.8014	3.2136	0.3574	0.3454	4.7107	0.02	1.05	-	-
	Estimated	5.0407	5.0407	5.0407	5.0407	5.0407	4.967	5.0407	-	-
Agg. 7	Expected	0.7876	3.9134	0.347	0.3494	5.1471	0.02	1.0215	0.85	-
	Estimated	5.3196	5.3196	5.3196	5.3196	5.3196	0	5.3196	5.3196	-
Agg. 8	Expected	0.8206	3.9765	0.3744	0.3435	5.1028	0.02	1.0259	0.88	0.01
	Estimated	5.3625	5.3625	5.3625	5.3625	5.3625	4.1468	5.3625	5.3625	5.3625

Table B.16: Minimum active power values for each equipment, for the expected and estimated values calculated by the UNN method with Fourier mapping, for the HIPE dataset.

		Eq. 2	**Eq. 3**	**Eq. 4**	**Eq. 5**	**Eq. 6**	**Eq. 7**	**Eq. 8**	**Eq. 9**	**Eq. 10**
Agg. 1	Expected	0	0	-	-	-	-	-	-	-
	Estimated	-2.115	-2.1368	-	-	-	-	-	-	-
Agg. 2	Expected	0	0	0	-	-	-	-	-	-
	Estimated	-0.6938	-0.0205	-1.8519	-	-	-	-	-	-
Agg. 3	Expected	0	0	0	0	-	-	-	-	-
	Estimated	0	-0.0028	0	-2.3909	-	-	-	-	-
Agg. 4	Expected	0	0	0	0	0	-	-	-	-
	Estimated	-3.7902	-4.0649	-1.7046	-2.5114	0	-	-	-	-
Agg. 5	Expected	0	0	0	0	0	0	-	-	-
	Estimated	-0.4699	-1.4571	-1.2564	-4.2428	0	0	-	-	-
Agg. 6	Expected	0	0	0	0	0	0	0	-	-
	Estimated	-4.6296	-5.0006	-5.0274	0	0	0	-3.6912	-	-
Agg. 7	Expected	0	0	0	0	0	0	0	0	-
	Estimated	0	-5.2696	-4.4569	0	-2.6592	-1.6827	-3.1645	-3.7987	-
Agg. 8	Expected	0	0	0	0	0	0	0	0	0
	Estimated	-4.8136	-3.3127	-3.768	0	-4.8992	0	0	0	0

Table B.17: Median active power values for each equipment, for the expected and estimated values calculated by the UNN method with Fourier mapping, for the HIPE dataset.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 1	Expected	0.3	0.09	-	-	-	-	-	-	-
	Estimated	0.274	0	-	-	-	-	-	-	-
Agg. 2	Expected	0.24	0.09	0.211	-	-	-	-	-	-
	Estimated	0.0864	0.1362	0	-	-	-	-	-	-
Agg. 3	Expected	0.25	0	0.21	0	-	-	-	-	-
	Estimated	1.0812	0	1.6254	0	-	-	-	-	-
Agg. 4	Expected	0	0	0.21	0	0.50465	-	-	-	-
	Estimated	0	0	0	0	4.2329	-	-	-	-
Agg. 5	Expected	0	0	0.21	0	0.40395	0	-	-	-
	Estimated	0	0	2.0369	0	4.1815	0	-	-	-
Agg. 6	Expected	0	0	0.21	0	0.15	0	0.86205	-	-
	Estimated	0	0	0	0	4.9474	0	2.9127	-	-
Agg. 7	Expected	0	0	0.21	0	0.15	0	0.86985	0	-
	Estimated	0	0	4.0765	0	0	0	4.5736	0	-
Agg. 8	Expected	0	0	0.21	0	0.14535	0	0.87	0	0
	Estimated	0	0	3.5702	0	0	0	5.2938	0	0

Table B.18: Sum of the active power values for each equipment, for the expected and estimated values calculated by the UNN method with Fourier mapping, for the HIPE dataset. The highlighted yellow cells correspond to the equipment with the highest active power consumption values within the aggregate.

		Eq. 2	Eq. 3	Eq. 4	Eq. 5	Eq. 6	Eq. 7	Eq. 8	Eq. 9	Eq. 10
Agg. 2	Expected	125.08	249.07	-	-	-	-	-	-	-
	Estimated	170.29	139.65	-	-	-	-	-	-	-
Agg. 3	Expected	122.08	234.09	87.348	-	-	-	-	-	-
	Estimated	452.86	272.72	-99.008	-	-	-	-	-	-
Agg. 4	Expected	126.32	198.99	84.838	58.34	-	-	-	-	-
	Estimated	451.99	300.35	690.72	100.28	-	-	-	-	-
Agg. 5	Expected	88.739	52.943	80.841	38.863	729.3	-	-	-	-
	Estimated	425.08	62.205	846.46	436.35	1496.3	-	-	-	-
Agg. 6	Expected	91.099	44.555	82.972	44.145	727.96	0.18	-	-	-
	Estimated	999.3	175.57	1197.8	285.32	1529.7	2.2266	-	-	-
Agg. 7	Expected	90.068	33.5	75.573	42.741	661.82	0.32	318.44	-	-
	Estimated	921.55	225.26	587.39	647.8	1567.8	79.472	1429.6	-	-
Agg. 8	Expected	93.94	67.378	80.764	52.377	547.9	0.28	331.22	121.18	-
	Estimated	1336	94.172	1595.3	805.46	1292.4	-23.558	1623.9	494.46	-
Agg. 9	Expected	90.777	53.91	77.235	50.051	582.88	0.26	322.97	121.27	0.6008
	Estimated	1080	328.49	1536.7	768.21	1319.3	53.891	1960.7	777.73	426.96

B.4.2 Analysis for the IMDELD Dataset

B.4.2.1 Maximum, Minimum, Median and Sum Values for the EMUPF Method

Table B.19: Maximum, minimum, mean and median active power values for each equipment, for the expected and estimated values calculated by the EMUPF method, for the IMDELD dataset.

		Eq. 1	**Eq. 2**	**Eq. 3**	**Eq. 4**	**Eq. 7**	**Eq. 8**
max	Expected	1421.7	1331.9	5077	6525	87954	86525
	Estimated	7.4793e+14	2.0534e+14	0.018491	0	0	4.6627e+46
min	Expected	0	0	0	0	0	0
	Estimated	-0.0019108	0	-3.5712e+14	-6.2907e+14	-4.7713e+13	-0.0025909
median	Expected	844.59	835.01	3227.4	5510	66336	60229
	Estimated	8.3303e+13	1.6742e+13	-3.9775e+13	-7.0065e+13	-5.314e+12	5.1932e+45

B.4.2.2 Maximum, Minimum, Median and Sum Values for the UNN Method

Table B.20: Maximum, minimum and median active power values for each equipment, for the expected and estimated values calculated by the UNN method, for the IMDELD dataset.

		Eq. 1	**Eq. 2**	**Eq. 3**	**Eq. 4**	**Eq. 7**	**Eq. 8**
max	Expected	1421.7	1331.9	5077	6525	87954	86525
	Estimated	45486	3639.3	24337	15646	1.4122e+05	1.0982e+05
min	Expected	0	0	0	0	0	0
	Estimated	-2608.9	-1407.8	-5116.5	-8180.4	-2793.3	-2502.1
median	Expected	844.59	835.01	3227.4	5510	66336	60229
	Estimated	6692.6	43.14	4841.7	4566.9	53199	58485

B.4.2.3 Maximum, Minimum, Median and Sum Values for the UNN Method with Fourier Mapping

Table B.21: Maximum, minimum and median active power values for each equipment, for the expected and estimated values calculated by the UNN method with Fourier mapping, for the IMDELD dataset.

		Eq. 1	**Eq. 2**	**Eq. 3**	**Eq. 4**	**Eq. 7**	**Eq. 8**
max	Expected	1.9265e+05	1.107e+05	2.653e+05	2.653e+05	1.6841e+05	2.3824e+05
	Estimated	45486	3639.3	24337	15646	1.4122e+05	1.0982e+05
min	Expected	0	0	0	0	0	0
	Estimated	-2.4798e+05	-2.2089e+05	-73718	-51212	-2.5312e+05	-349.63
median	Expected	844.59	835.01	3227.4	5510	66336	60229
	Estimated	0	0	1.2761e+05	45470	0	1.5627e+05